

Lightly edited version of a chapter first published in Keith Allan (ed., 2013), *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press; pp. 503-536.

## Chapter 22: Lexicography from Earliest Times to the Present<sup>1</sup>

Patrick Hanks

### 22.1 What is a dictionary?

A dictionary, as Trench (1858) observed, is an inventory of the words of a language (with explanations of meaning and other information). All the world's major literary languages, as well as some less common ones, have evidently felt the need for such an inventory, and the trend has spread to rare and endangered languages. According to this view, lexicographers are, first and foremost, linguistic inventory clerks, but, as we shall see, there are other motives, too, for compiling a dictionary.

At first glance, the humble occupation of collecting words, defining them, and arranging them in some sort of order—usually, alphabetical—would not seem to call for any profound theoretical insight. However, when the activity begins to be undertaken in earnest, theoretical and practical linguistic questions begin to crowd in.

- What is the relationship between words and phrases? In traditional dictionaries of English and other European languages, a false dichotomy is presented between words and idioms (otherwise known, rather misleadingly, as ‘fixed phrases’), as if there were some sharp dividing line between the two categories. During the past two or three millennia, dictionaries have tended to reinforce naïve theoretical reductionism as regards questions about where meaning resides. Recent advances, both in construction grammar (e.g. Goldberg, 1995, 2006) and in corpus linguistics (e.g. Sinclair 1991, 1998; Hanks, 2004, 2013), suggest that meaning resides not only in lexical items but also in phraseology.
- How far should a dictionary go in recording the millions of attested nominal phrases

---

<sup>1</sup> I would like to thank Xia Lixin, Rocky Miranda, Mohammad Resa Aslani, Elisabetta Jezek, Paz Battaner, Irene Renau Araque, and Araceli Alonso Campo for their comments on earlier drafts of this chapter. Any errors that remain are of course entirely my responsibility.

such as *fire escape* and *forest fire*, each of which has at least one unique meaning, which is very often not derivable from the sum of its parts? English dictionaries ignore most such combinations. Different languages deal with such compound concepts in different ways.

- How strictly should a dictionary confine its inventory to recorded usage? Should it be allowed to speculate about possible words and meanings in addition to recorded words and meanings? For example, is it a rule of English that adjectives have an adverb derivative ending in *-ly*. If this is right, should possible adverbs such as \**saintlily* be recorded, even though there is little or no evidence that they have ever been used? The adjective *lame* has two senses: ‘suffering from an injury to the leg’ (e.g. *a lame horse*) and ‘inadequate’ (e.g. *a lame excuse*). Should two parallel senses be recorded for the adverb *lamely*, even if the available evidence suggests that the ‘injured’ sense is rare or nonexistent for the adverb?
- What can be said by way of definition of a word’s meaning(s)? Can meanings be defined strictly and formally according to genus and differentia, as desired by Wilkins (1668) and Leibniz (see Couturat, 1901)—or can they only be typified and hinted at?
- Can a spelling form be shared by more than one word? For example, should *record*, verb, and *record*, noun, be recorded as a single lexical entry (despite difference in pronunciations and meaning), or should there be two separate entries? What about *keep*, verb, and *keep*, noun? Should all lemmas that are used as two (or more) parts of speech have separate entries (homographs) for each part of speech? Different dictionaries adopt different policies with respect to such questions.
- What can be said about the origins and history of a word? This was a central topic in 19th-century European linguistics, following the gradual uncovering of the family of Indo-European languages, and it is still the main focus of the great historical dictionaries of European languages, such as OED and Grimm.
- How much attention should be paid to etymology in a synchronic dictionary? For example, philologists tell us that *weave* as an intransitive verb, meaning ‘to move rapidly in and out’, has a different etymology from *weave* as a transitive verb, meaning ‘to create cloth or a garment (by moving threads rapidly in and out on a loom)’. Is this a good reason for making them separate entries, in defiance of the common-sense perception that the two uses are somehow cognitively related and

(whatever their history may be) can now be regarded as ‘the same’ word?

Questions such as these can have a profound effect on linguistic theory—our understanding of the way language works. Many such questions have been addressed by linguists, but others have been neglected, especially in the English-speaking world. In this chapter, we take a brief look at how dictionaries have developed in different cultures and related to linguistic theory.

## 22.2 Roots: lexicography before printing

### 22.2.1 China

The earliest dictionaries in the world were compiled in China during the Han Dynasty. The history of Chinese lexicography has been summarized in English by Li Ming (2006), more fully by Yong and Peng (2008). For various reasons, not least the fact that Chinese writing is ideographic rather than alphabetical, there is room for debate about what counts as the earliest Chinese dictionary. Two works in particular may be mentioned here. It is generally agreed among Chinese scholars that the *Erya* (爾雅 ‘Near Correctness’), dating from the 2nd or 3rd century BC, must be classified as a work of encyclopedic lexicography: it contains explanations of the meanings of words, phrases, and other passages in classic Chinese texts. According to Karen Chung (personal communication), the *Erya* falls somewhere between a thesaurus and a topically organized lexicon.

During the Eastern Han Dynasty (1st-2nd centuries AD), Xu Shen compiled the *Shouwen jiezi* (說文解字 ‘Origin of Chinese Characters’). This remarkable work of scholarship is the foundation of all subsequent Chinese lexicography and linguistics. It was presented to the Emperor in AD 121, though actually completed many years earlier. It is recognizably a dictionary, even to Western readers. It contains entries for approximately 10,000 Chinese characters, with information about their origins, meanings, and pronunciation. It is organized in 540 sections according to the ‘radicals’ of each word. According to Boltz (1993), Xu’s motives in compiling the *Shuowen* were pragmatic and political, rather than communicative, and sprang from the Confucianist belief that using the correct terms for things was essential for proper government. It is, therefore, a normative (prescriptive) dictionary rather than a descriptive one.

### 22.2.2 India and Persia

Sanskrit dictionaries and thesauruses were compiled over two thousand years ago, and these were the start of a long tradition of native lexicography in Indian languages. Three terms are particularly relevant to the Indian lexicographic tradition: *nighantu*, *kosha*, and *nirukta*.

*Nighantu* simply means ‘lexicon’. The earliest known *nighantu* gives explanations of obscure words found in Vedic texts (sacred literature). In the second or third century BC, a scholar called Yaska, about whom nothing else is known, wrote an etymological commentary (*nirukta* ‘explanation’) on words found in a lexicon (*nighantu*).

A *kosha* is literally a storehouse or treasury. Unusually for lexicography, the earliest *kosha* was written in verse. It contains entries for nouns and indeclinable forms, but not verbs, and was intended for use by poets. The best-known such work is the *Amarakosha* by Amarasinha, a Buddhist scholar and poet who probably lived in the sixth century AD.

Throughout the medieval period there was much cultural interchange between India and neighbouring Persia (modern Iran), which intensified after the emergence of Islam, up to 1947, when Pakistan was established as an independent Islamic state between India and Iran. There is allusive evidence that Persian dictionaries existed during the Sassanid dynasty (3rd–7th centuries AD), but these have not survived. The most important surviving dictionary of Persian before modern times is undoubtedly the *Loghat-e-Fors* (Lexicon of Persian), compiled by the epic poet Abu Mansur Ali ibn Ahmad Asadi Tusi (died 1072). Asadi’s declared aim was not only to record and explain words found in Persian poetry—words that might be unfamiliar to his contemporaries—but also to foster the continuation of Persian traditional literature. The entries are illustrated with citations from poetry; they are arranged according to the last letter of each entry word, in order to help poets find suitable rhymes—a practice also followed in medieval Arabic and Hebrew lexicography.

Another Persian work that must be mentioned is a lexicon compiled by Faxr-e-Qavas Qaznavi in India in 1291. This is the *Farhang-e-Panj Baxši* ‘culture in five sections’, so called because the entries are arranged hierarchically on semantic principles in five sections, on principles somewhat similar to those of Roget’s Thesaurus (1852). For example, the fourth section of the book contains words for animals. It is subdivided into five ‘varieties’, of which the fifth concerns words for human beings, and this in turn is divided into two parts. The first part concerns human organs, and the second part is about humans and their environment. Each entry is presented with a verse from poetry as well as a definition. This

type of dictionary is known in French as a *dictionnaire analogique*. The best-known example is the *Dictionnaire analogique de la langue française: répertoire complet des mots par les idées et des idées par les mots*, by Jean-Baptiste Prudence Boissière, published by Larousse in 1862.

The Persian lexicographical tradition continued mainly in India, where many Persian writers resided. Between the 9th and the 19th centuries, about 130 Persian dictionaries were compiled. The first Persian dictionary with explanations in Farsi was compiled in the 15th century, after the Mongol invasion. It contained about 2300 lexemes and was written by Hendu Shah Nakhjavani, known as Shams-e-Monshi. A much larger work, *Borhan-e-Ghate*, with 20,000 entries, published by Mohammad Hossein Tabrizi in 1724, also belongs to this period of Persian and Indian history.

### **22.2.3 Classical Greece and Rome**

The history of Greek lexicography has been summarized by Stathi (2006), who gives both an account both of the efforts of Greek lexicographers in classical times and of the lexicography of ancient Greek since the Renaissance.

From the fifth century BC onwards, it was customary for Greek scribes to insert glosses into manuscript copies of the works of Homer and other earlier writers, explaining obsolete and unusual words. From the third century BC these glosses were compiled into separate glossaries by scholars at the library in Alexandria. All except a few fragments of these compilations have since been lost.

In the second century AD, a different kind of Greek lexicography developed, as a result of the puristic linguistic movement known as Atticism. The form of Greek that was used as a lingua franca throughout the eastern Mediterranean during the period of the Roman Empire was regarded by literati as ‘incorrect’ or ‘impure’. A dictionary was therefore needed that would present and define ‘correct’ words and terms, i.e. those that had been used by the great writers of Athens in the fourth to second centuries BC. Such a work was the *Eklogē* (Ἐκλογή ‘selection’) of Attic words and phrases compiled in Byzantium in the second century by Phrynichos of Bithynia. This is a collection of Byzantine Greek words to which Phrynichos and his ilk had objections, the ‘pure’ or ‘correct’ Attic equivalents being given alongside the ‘impure’ colloquial forms. Ironically, the chief interest of this work to modern scholars is the light that it sheds on colloquial Byzantine Greek.

Lexicography in ancient Rome can be approached through the proceedings (Ferri, ed., 2011) of the 2008 conference Pisa in December 2008, held under the title ‘The Latin of Roman Lexicography, from Verrius to the *Corpus glossariorum*’, and through the Festus project at University College London (<http://www.ucl.ac.uk/history2/research/festus/index.htm>). Ancient Latin linguists shared the general Roman predilection for classifying and imposing order on everything, and this is no doubt one reason why grammarians such as Priscian and Aelius Donatus had a profound influence on subsequent traditions in linguistics throughout Europe up to the mid-twentieth century, while Latin lexicographers are comparatively little known; works of classical Latin lexicography have been partly or wholly lost. It is known, too, that the Romans created bilingual Greek-Latin word lists, but these have likewise not survived.

An ambitious monolingual dictionary called *De Verborum Significatu* ‘on the meaning of words’ was compiled by the philologist and educationist Marcus Verrius Flaccus (c. 55 BC – 20 AD), tutor to the grandsons of the Emperor Augustus. By all accounts it was a huge work (letter A alone took up four books) and was concerned with etymology and cultural history as well as word meaning. Entries were supported by citations from literature. In the 2nd century AD Sextus Pompeius Festus edited a revised version of this work, part of which (from letter M onwards) has survived in a single seriously damaged manuscript. It is also known that Festus omitted obsolete and archaic words from his version of the dictionary and wrote them up in a separate work called *Prisorum verborum cum exemplis*, which, sadly, has been lost: it would have been a priceless source of information about early Latin. In the eighth century, the historian Paulus Diaconus created an abridged version of Flaccus’s dictionary, and this has survived. It is our chief source of information about ancient Roman lexicography. The entries have a much greater emphasis on cultural practices and beliefs than modern readers of a dictionary would expect, and little or nothing is said about function words and grammar. Thus, it may be classified as a cultural rather than a linguistic compendium.

#### **22.2.4 Arabic and Hebrew**

Between the 7th and the 13th centuries AD, a number of Arabic dictionaries were compiled, with a variety of different purposes, including regulation of ‘correct’ language, the facilitation of poetry, and deepening understanding of the words of the Qur’an. These developments are described by Haywood (1960), who comments, “In the compilation of dictionaries and other lexicographical works, the Arabs ... were second to none until the Renaissance, with the possible exception of the Chinese.” From Baghdad and Basra in the east to Granada and

Cordoba in the west, the first five centuries of Islam witnessed an extraordinary flowering of literature. It is noteworthy that lexicographical activities of Muslim Spain, to the far west of the Mediterranean, are included here. As explained by Roth (1994), both Arabic and Hebrew lexicography flourished, along with many other modes of scholarship, under the enlightened regimes of the Muslim rulers in medieval Spain, before they were all swept away by the Christian 'reconquest', culminating in the 15th century with the intolerant religious fundamentalism of Queen Isabella the Catholic and the Spanish Inquisition.

Latin and Greek enjoyed a universally accepted conventional alphabetical order from time immemorial, but the history of the conventional order of letters in the Arabic writing system is more complex. The complexity is compounded by the fact that in written Arabic normally only consonants are represented; readers are left to supply the vowels for themselves. For further details of written Arabic and indeed other writing systems, see Daniels and Bright (1996). Early Arab lexicographers experimented with various arrangements for ordering words: for example, both Al-Jawhariyy's *As-sihah* 'the Strong' (10th century) and Ibn Manzur's monumental *Lisan Al-'Arab* 'Language of the Arabs' (13th century) order words according to the last consonant and arrange the consonants in an order that is determined to some extent by the mode of articulation. Modern Arabic dictionaries follow a different alphabetical order.

The most important work of medieval Arabic lexicography is the *Kitab Al-'Ayn* (literally, Book of the 'ayn), compiled by Al-Khalil ibn Ahmad in the 8th century. An 'ayn is a written symbol representing the Arabic voiced pharyngeal fricative consonant /ʕ/, which has no equivalent in Latin or Greek. Al-Khalil used the name of this symbol as the title of his work, which in actuality is nothing less than a comprehensive dictionary of the Arabic language. It is claimed by some that this was the first systematic attempt anywhere to compile a comprehensive lexicon of any language.

The earliest known works of Hebrew lexicography were compiled in the Middle East in the 10th century AD (see Drory 2000; Cohen and Choueka 2006). Hebrew had already become rare or extinct during the Roman Empire, so these were, in effect, dictionaries of a dead language. For two millennia Hebrew survived mainly (or only) as a liturgical and literary language, being preserved and cherished as a symbol of the ethnic and religious identity of Jews during the diaspora. From at least the 2nd century AD it was no longer a medium of everyday communication. For that purpose, it was supplanted first by Aramaic and subsequently by Arabic.

*Sefer ha-Egron* (902) is a lexicon of approximately 1000 Hebrew words for poetic purposes, compiled in Egypt by Sa'adiah ben Josef. Not all of it has survived. The words are presented in two arrangements: first, they are listed alphabetically, not for ease of reference (as we might expect), but to help poets compile acrostic verses. The second list is of words according to their final consonant, in order to facilitate rhyming. Some years later Sa'adiah issued a version of his work with glosses in Arabic, to facilitate understanding of the meaning of Hebrew words. Sa'adiah was a scholar, philosopher, and theologian as well as a lexicographer; in 928, after many years' residence in Jerusalem, he was appointed gaon (spiritual leader) of the Jewish community in Sura, south of Baghdad; Gaon is often represented as his surname.

*Kitab Jami al-Alfaz* (c. 945) is a Hebrew-Arabic dictionary of words in the Bible, compiled in Fez, Morocco, by David ben-Abraham El-Fasi.

According to Cohen and Choueka (2006), “probably the finest achievement of medieval Hebrew lexicography is ‘the Book of Roots’ by Yonah ibn-Janah.” Ibn-Janah, otherwise known as Abu al-Walīd Marwān ibn-Janāh, was a Jewish linguist and grammarian of the eleventh century who lived in Cordoba, Spain. His lexicon, *Kitab al-'usul*, is the second part of a work known as *Kitab al-Anqih* ‘the Book of Exact Investigation’. The first part is a Hebrew grammar with the exotic title *Kitab al-Luma* ‘Book of the Multicoloured Flowerbeds’. The arrangement of the lexicon is based on the three-letter root system that is now recognized as universal for Semitic languages. The explanations or glosses are in Arabic. In the thirteenth century, a revised grammar and lexicon based on the work of Ibn-Janah was compiled in Narbonne by the Biblical scholar and philosopher Rabbi David Qimhi (1160–1235). This work, *Sefer Hashorashim*, was to be an influential source of Christian Old Testament scholarship.

### **22.2.5 Medieval Europe**

In medieval Europe, lexicography originated (or rather, re-invented itself) in the form of interlinear vernacular glosses on words in medieval Latin manuscripts. These glosses came to be gathered up into separate works, sorted roughly into alphabetical order (see Murray 1900, Castro 1991, Hanks 2006, Kramer 2006). During the Middle Ages, a variety of manuscript lexicons of Latin with glosses in various vernaculars were compiled for use by novices in monasteries as they studied Latin—the universal European lingua franca of the Christian



religion, of philosophy, and of scholarship. The earliest known example is the 8th-century *Glosses de Reichenau* (so called because the manuscript formed part of the library of the Benedictine Abbey on Reichenau Island in Lake Constance). This consists of over 5,000 words of the Latin Vulgate with glosses in Gallo-Roman medieval Latin, a precursor of Old French.

Probably the best-known example in England of such a glossary is the *Promptorium Parvulorum* ('Young People's Storeroom'), compiled in about 1440 by Galfridus Anglicus (alias Galfridus Grammaticus 'Geoffrey the Grammarian'), a Dominican friar living in Norfolk. Its 10,000 entries (words and phrases) were laboriously copied out by hand many times—the only means of dissemination possible before the invention of printing—and the copies transported to the libraries other monasteries. Then, in 1499, something revolutionary occurred. The *Promptorium* was set in type and printed. This meant that identical copies of the work could be created and made available more or less immediately to anyone who wanted one and was able to pay for it. The role of the medieval scribe was at an end and immense new opportunities for ambitious lexicography opened up.

### **22.3 The Renaissance: the impact of printing on lexicography**

Dictionaries are not only vast, systematic inventories of minutiae concerning lexical items; they are also vehicles that disseminate such information, thereby encouraging the growth and preservation of cohesive cultural and linguistic conventions in a language community. This disseminative role only began to realize its full potential with the invention of printing, so that identical copies of a work, however large, could be printed off and distributed within a very few days. Before the invention of printing in the mid fifteenth century, each copy of a work had to be laboriously written out by hand. Thus, the invention of printing is of the greatest importance in the history of lexicography.

An equally important and related development was the revolution in typographical design and metal type-founding in Venice the 1470s, which rapidly spread northwards, reaching Paris in the 1490s and England some decades later. The central figure here is Nicolas Jenson (1420-80), an expert in metals who had been master of the French Royal Mint in Paris before moving to Venice and setting up a printing business there. Jenson was a type-founder who introduced new standards of elegance and legibility, including skilful use of space on the page, with minute attention to the tiniest details of letter spacing, kerning, etc., and systematic distinctions between capital letters (based closely on Roman monumental inscriptions) and

lower-case letters (based on Carolingian minuscules). Another important Venetian was the great Renaissance scholar Aldus Manutius, whose typographer Francesco Griffo faithfully observed Jenson's principles, adding further options such as the distinction between roman and italic type. The Venetian typographers abandoned the heavy, hard-to-read, space-consuming black-letter type of Gutenberg and other early printers and set new standards for all future typography of printed books and journals in the Roman alphabet outside Germany. It is no exaggeration to say that the typographic achievements of Jenson, Griffo, and their immediate successors (notably Johann Froben in Basel and Claude Garamond in Paris) were a crucial factor, not only in the flood of classical texts that were rescued and printed during the Renaissance, but also in making modern lexicography possible, enabling lexicographers to cram vast quantities of information elegantly and legibly onto each page and to disseminate large numbers of identical copies of completed dictionaries quickly and efficiently. These developments are discussed in more detail in Hanks (2010). If we compare the black-letter type of *Promptorium Parvulorum* (1499), the earliest printed dictionary in England, whose black-letter type aimed at nothing more ambitious than replication of the letters used by scribes in monasteries, with Robert Estienne's masterly, elegant, and huge Latin *Dictionarium* (1531), we see a quantum shift in presentation, affecting both the quality and the quantity of information. Estienne's work would simply not have been possible in black-letter type. It would have been an unmanageable and unnavigable monster.

Estienne was a master printer in Paris, as well as a lexicographer and one of the leading intellectuals of his day. Even a casual inspection will show that his great work has most of the characteristics that present-day readers have come to expect of scholarly monolingual dictionaries, including:

- Comprehensive list of words (lemmas);
- Morphology: selected inflected forms for lemmas ('principal parts'), giving guidance on conjugations and declensions;
- Clearly distinguished definitions, capturing an appropriate level of generalization for each meaning of each word;
- Citations from literature (much of which was printed for the first time by Estienne himself), supporting each definition.

There is also a feature that has only very recently been revived by modern lexicographers:

- Extensive selection of idiomatic phraseology.

Estienne's *Dictionarium* is a monolingual Latin dictionary: its definitions are in Latin, accompanied by occasional vernacular glosses in French for 'hard words'. It very clearly aims to be an inventory of the classical Latin language, or at least of the vocabulary of the classical Latin texts that were printed and published in Estienne's day. His motivation, according to Considine (2008), was to contribute to the preservation of the heritage of classical literature, and the same is true of the equally ambitious and equally monumental *Thesaurus Graecae Linguae*, published by his son Henri Estienne in 1572.

Two later dictionaries published by Robert Estienne show a different side of this great lexicographer. His main concern in 1531 was to cater to the needs of scholars and literati. But he was also sensitive to the needs of less erudite students and language learners. The *Dictionnaire francoislatin* of 1539 is a practical work explicitly aimed at students learning to express themselves in Latin. A noticeable feature is the large number of idiomatic French phrases for which Latin equivalents are offered. To take just one rather striking example, the phrase *l'ordre et collocation des mots* is glossed as 'verborum constructio'. Here we have two words—*collocation* and *construction*—that are buzzwords in today's linguistics in the English-speaking world, as it slowly, painfully, and belatedly turns to the empirical and theoretical analysis of lexis and phraseology. Robert Estienne placed considerable emphasis on phraseology and context: it seems likely that he would have been sympathetic to and even excited by modern theories of collocation and construction grammar. A complementary and equally practical work is his *Dictionarium Latino-Gallicum* (1552). This is not a revised version of his 1531 work. Instead, it is a practical guide whose aim is to help students decode the meanings of Latin words and Latin texts into their native French—an early example of a bilingual dictionary.

One of the first monolingual European dictionaries devoted to a vernacular language (i.e. not Latin) was the *Tesoro de la lengua castellano o española* (Madrid, 1611) by Sebastian de Covarrubias, a sophisticated linguist and cultured humanist who included not only definitions and Latin etymons for words but also place-names and a number of subjective comments on lexical issues. This is a substantial work of over 1400 beautifully typeset pages, in the best tradition of Robert Estienne. An on-line facsimile can be seen at <http://fondosdigitales.us.es/fondos/libros/765/1275/tesoro-de-la-lengua-castellana-o-espanola/>

## 22.4 Polyglot dictionaries and the emergence of bilingual lexicography

In terms of literature, the European Renaissance was primarily a revival of the literature and learning of ancient Greece and Rome and, as suggested above, printed dictionaries of classical Latin and Greek played a substantial role. Linguistically, however, the Renaissance marked the beginning of the long, slow decline of Latin as an international lingua franca and the flourishing of vernacular languages as media for communication and culture throughout Europe. In these circumstances, one might have expected an exuberant growth of vernacular bilingual dictionaries, for example offering translations of words and phrases from Italian into French, Italian into English, French into Hungarian, or Spanish into German.

Surprisingly, however, up to the end of the sixteenth century bilingual dictionaries of vernacular languages were few and far between. Instead, scholars, translators, travellers, and diplomats alike were constrained either to speak Latin or to rely on vernacular glosses appended to a rather inferior Latin dictionary. This was the *Dictionarium* of Ambrogio Calepino, an Augustinian friar living in Bergamo. Calepino's original edition (1502) was a Latin vocabulary, with glosses in Latin supported by citations, together with encyclopedic entries for the figures of classical mythology. In a second edition, glosses in Italian and French were added. By a process of accretion, the vocabularies of other languages, starting with Greek and Hebrew, were gradually added by others to successive editions of Calepino's original. In the words of Freed (2007), "it evolved into the first polyglot dictionary." By 1580, a dozen different editions, containing glosses in up to eleven different languages, all attributed to Calepino, were in print, published in locations as far apart as Reggio nell'Emilia, Venice, Paris, Strasbourg, Hagenau, Lyon, and Rome. In Paris alone, five competing editions appeared between 1524 and 1541. The 1573 edition printed and published in Venice includes the following comment in its front matter, quoted and translated by Freed:

*In hac postrema editione, ut hoc dictionarium commodius exteris nationibus inservire possit, singulis vocibus latinis italicas, gallicas, & hispanicas interpretationes inseri curavimus.*

In this latest edition, in order that this dictionary might more fully serve foreign nations, we have taken care to insert Italian, French, and Spanish definitions among the lone Latin entries.

By this time Ambrogio Calepino himself (1450–1510) was long dead and his book had become common property. Indeed, his very name had become common property. The *OED* has an entry for the obsolete English word *calepin*, supported by sixteenth and seventeenth century citations and glossed as:

A dictionary (sometimes 'a polyglot'); *fig.* one's book of authority or reference; one's notebook or

memorandum-book.

The *OED* also notes the French phrases “je consulterai là-dessus mon calepin”, “cela n’est pas dans son calepin”, “mettez cela sur votre calepin” (i.e., make a note of that to serve as a lesson), and the obsolete English expression “to bring someone to his calepin”, i.e. to the utmost limits of his information. Evidently, Renaissance readers expected a dictionary to be a comprehensive inventory.

The complex bibliographical history of Calepino’s dictionary and its derivatives have been traced by Labarre (1975). This shows that multilingual editions really began to take off in the 1550s; by the 1580s they had come to include lexical items in up to 11 languages—not only Latin, Greek, Hebrew, Italian, French, and Spanish, but also outlandish tongues such as German, English, Polish, and Hungarian. By the end of the century, a Latin-Portuguese-Japanese ‘Calepino’ had appeared, supporting the missionary work of the Portuguese Jesuits who were at that time seeking to Christianize Japan. It has often been said that Calepino’s original work is deficient in scholarly precision, while these polyglot derivatives are great, cumbersome things, not suitable for carrying around and not particularly user-friendly. Nevertheless, these were the principal works that served for interpretation among vernacular languages in the sixteenth century.

The great English linguist and lexicographer John Palsgrave was French tutor to Mary Tudor, sister of Henry VIII, who was destined to marry the King of France. Palsgrave compiled a magnificent bilingual French-English dictionary and phrase book (in many cases with amusing and diverting illustrative phrases) as the major part of his general account of the French language, *Lesclaircissement de la langue francoyse* (1530). However, strangely, few scholars at the time followed Palsgrave’s lead. The idea of bypassing Latin with bilingual dictionaries did not really catch on until the very end of the sixteenth century.

Among the first monolingual European dictionaries devoted to a vernacular language (i.e. not Latin) was the *Tesoro de la lengua castellano o española* (1611), published in Madrid. This is a substantial work of over 1400 beautifully typeset pages, compiled by Sebastian de Covarrubias, a sophisticated linguist and cultured humanist who included not only definitions and Latin etymons for words but also place-names and a number of subjective comments on lexical issues. An on-line facsimile of this fine dictionary can be seen at <http://fondosdigitales.us.es/fondos/libros/765/1275/tesoro-de-la-lengua-castellana-o-espanola/>

## 22.5 The expectation that etymology guarantees meaning

A hundred years later, during the European Enlightenment, in the case of living languages, a need was increasingly felt to prescribe standards. These standards were based on the ill-defended assumptions that earlier forms of a language are somehow more ‘correct’ than contemporary forms and that etymology guarantees meaning. A moment’s thought will convince us that this assumption is incorrect. For example, none of the meanings of the modern word *subject* or its cognates in other modern languages have anything to do with the Latin etymology, which literally means ‘something thrown under’, and the same is true of thousands of other words that have changed their meaning, in some cases many times, in the course of recorded history (not to mention reconstructed lexical prehistory). Nevertheless, the notion that etymology guarantees meaning was prevalent in Europe as the Renaissance developed into the Enlightenment, and indeed it was responsible for some remarkably fine scholarly lexicography. The notion was effectively refuted by Johnson (1755) in the preface to his dictionary, though it persists in a wistfully hankered-after form in some of the more conservative academies of Europe.

In 1612 (after over 20 years of work) the Accademia della Crusca published a *Vocabolario* for the Italian language, the aim of which was explicitly prescriptive, conservative, and indeed retrogressive, i.e. to establish the already old-fashioned Florentine dialect of the 14th century (as written in particular by Dante, Boccaccio, and Petrarch) as a gold standard for Italian. This was followed in 1640 for French by the first edition of the *Dictionnaire* of the Académie Française, whose aim was equally prescriptive and conservative: “to give definite rules to our language and to render it pure”.

The Real Academia Española was founded with similar aims in 1713, and still proudly announces that its mission is to regulate the Spanish language—“to fix the voices and vocabularies of the Castilian language with propriety, elegance, and purity”. The first edition of its dictionary, published under the title *Diccionario de autoridades* (‘Dictionary of Authorities’) in 1726. It is called a “dictionary of authorities” because its definitions are supported by citations from literature. Unfortunately, in the mid 18th century, it was decided that including citations is a waste of space in a dictionary whose role is to regulate the language. Since then, the *Diccionario de la lengua española de la Real Academia Española* (DRAE), the 23rd edition of which is available on line, has tended to be extremely conservative. It was slow to admit neologisms and, at least up to 2006, had an inadequate system for labelling register, i.e. for distinguishing racist, sexist, and other offensive word

uses from normal usage<sup>2</sup>.

## 22.6 Samuel Johnson

In the early eighteenth century several English lexicography projects were proposed on the model of the French and Italian academy dictionaries, with the aim not only of inventorizing and defining all the words in English, but also of ‘fixing’ the language in its then supposed state of excellence. This aim eventually bore fruit in Samuel Johnson’s *Dictionary* (1755), but with some interesting modifications of purpose, arising from Johnson’s profound understanding of the nature of language.

Johnson was not only a lexicographer but also a major intellect: essayist, poet, biographer, critic, editor, and conversationalist. He set out with the aim, suggested to him by a consortium of booksellers, of ‘fixing’ the language, but in the course of the work, he came to recognize that a living language cannot be ‘fixed’: language change is inevitable. The lexicographer must therefore set out to observe and describe, rather than to pontificate and prescribe.

Those who have been persuaded to think well of my design require that it should fix our language, and put a stop to those alterations which time and chance have hitherto been suffered to make in it without opposition. With this consequence I will confess that I flattered myself for a while; but now begin to fear that I have indulged expectation which neither reason nor experience can justify. When we see men grow old and die at a certain time one after another, from century to century, we laugh at the elixir that promises to prolong life to a thousand years; and with equal justice may the lexicographer be derided, who being able to produce no example of a nation that has preserved their words and phrases from mutability, shall imagine that his dictionary can embalm his language, and secure it from corruption and decay. (*Preface to the Dictionary*, Johnson 1755, §84)

The *Preface*, in particular, deals with many of the issues that concern modern lexicologists, as explained in Hanks (2005): issues that were not revisited until the work of twentieth century scholars—philosophers of language such as I. A. Richards, Ludwig Wittgenstein, and Hilary Putnam, anthropologists such as Bronislaw Malinowski and Eleanor Rosch, and

---

<sup>2</sup> Since writing this paragraph, I have discovered the *Redes* ‘Network’ dictionary (2004) of Ignacio Bosque, a member of the Real Academia Española. *Redes* is a detailed lexicographic study of collocations and meanings in present-day Spanish. In its innovative approach, it is far in advance, both in principle and in execution, of lexicographical works in Spanish and many other European languages.

linguists such as J. R. Firth and J. M. Sinclair. Johnson's recognition that language change is inevitable spared the English language the impertinence of an academy of learned men (and, later, women) impotently debating the acceptability or otherwise of behavioural phenomena (patterns of word meaning and word use) which in reality they have no power to alter.

Among Johnson's many merits and influences as a lexicographer are the following.

- Extensive use of illustrative citations from literature—not only to prove the existence of a particular sense of a word, but also to illustrate elegant usage and to delight and educate the reader.
- Arrangement of senses in a rational order, so that each dictionary entry stands as a coherent discourse, reflecting meaning development, influenced by but not governed by etymology and not just a list of senses in historical order.
- Extensive use of Aristotelian-Leibnizian principles of definition—stating first what kind of thing in general a word denotes and then adding carefully selected differentia.
- Respect for the vagaries of a living language—he recorded word meanings as he found them, not necessarily as he may have wished them to be. He confined his value judgments to a few acerbic comments (e.g. *clever*, “a low word”) and he observed, for example, that, although previous English lexicographers had found it convenient to define *ardent* as meaning ‘burning’, this etymological sense of the word never made the transfer from Latin to English.
- Effective treatment of phrasal verbs.

Johnson's was the standard dictionary of English until the end of the 19th century, when it was superseded by the Philological Society's *New English Dictionary on Historical Principles* (*NED*, 1884–1928).

## 22.7 Historical principles

The *NED* was published by Oxford University Press and in the 1930s it was re-christened *The Oxford English Dictionary* (*OED*). It was followed by a shortened version, the *Shorter Oxford English Dictionary* (*SOED*), in two large volumes—whose title is sometimes wrongly thought to be some kind of joke, since it is so very much bigger than most other English dictionaries—and by a plethora of regional works on similar principles, including the *Dictionary of American English* (*DAE*), the *Dictionary of the Older Scottish Tongue* (*DOST*), the *Scottish National Dictionary* (*SND*), the *Australian National Dictionary* (*AND*), and the



*Dictionary of South African English (DSAE)*, among others. It was also the supreme example of a general 19<sup>th</sup>-century European movement to compile historical dictionaries of national languages, which included the *Deutsches Wörterbuch* of the brothers Grimm, the *Trésor de la langue française*, the *Woordenboek der Nederlandsche Taal*, and many others.

A dictionary on historical principles places the etymology at the start of each entry and traces the semantic development of the word by arranging senses in historical order. Thus, in the *Oxford English Dictionary*—a dictionary on historical principles—the entry for *camera* starts by explaining that the word is from classical Latin *camera*. The first sense is “the department of the papal Curia dealing with finance; the papal treasury”. Sense 2 is “an arched or vaulted roof, chamber, or building (also more generally: any room or chamber)”. It is not until senses 4b and 4c respectively in the third edition that we get the familiar modern senses, “a device for taking photographs” and “a device for capturing moving pictures or video signals” (*OED* 3rd edition; entry revised and updated in 2010). In the first edition this sense was not present; this is not surprising, because cameras in the modern sense had only just been invented. Sense 4a is a cross-reference to the entry for *camera obscura*. All this faithfully reflects the chronological development of the word in English though it is no doubt somewhat confusing to a naïve user who wants to know what *camera* means in modern English. Dictionaries on historical principles are of great value to literary scholars, social historians, historians of science, historical linguists, and others. However, they are not intended for language learners, translators, computational linguists, or casual inquirers into word meaning. Failure to make this simple typological distinction has resulted in considerable confusion and even misuse of great works of scholarship. It also resulted in mindless application of historical principles of lexicography to smaller one-volume dictionaries intended for general use.

Merriam Webster’s *Third New International Dictionary (MWIII 1961)* is a large American dictionary on historical principles, with impressive coverage of technical terminology in fields ranging from Agriculture to Zoology. Its definitions for everyday words are sometimes less than satisfactory, as a glance at entries such as *door*, *hotel*, *sugar*, and *mimosa* will show. The root of *MWIII*’s problematic definitions lies in a failure to distinguish word meaning from concept meaning, compounded by the editor’s instruction to his staff that all explanations should be couched in terms of a single one-phrase definition.

**sugar:** a sweet crystallizable substance that consists entirely or essentially of sucrose that is colorless or white when pure and usu. yellowish to brown otherwise, that occurs naturally in the most readily

available amounts in sugarcane, sugar beet, sugar maple, sorghum and sugar palms, that is obtained commercially principally by processing the juice expressed from sugarcane or the aqueous extract of sliced sugar beets and refining so that the final product is the same regardless of the source, and that forms an important article of human food and is used chiefly as a condiment and preservative for other foods and for drugs and in the chemical industry as an intermediate. (MWIII 1961)

This starts well enough, defining the meaning of the noun *sugar* by stating a genus term (“a crystallizable substance”) and adding differentiae (“sweet”, “consists of sucrose”). However (setting aside any doubts we may have about whether sugar can be sugar before it is crystallized), we can see that the definition begins to go haywire after “sucrose”. It is a rule of English grammar that a restrictive relative clause governed by *that* modifies the meaning of the preceding noun, but the natural interpretation of the relative clause in question would be incorrect, for it is not intended to distinguish one kind of *sucrose*, namely the colorless or white kind, from other kinds of sucrose; instead, it is a further differentia of “crystallizable substance”. From here on the syntax, structure, and wording of the definition become increasingly bizarre until a mystical point of incomprehensibility is reached, culminating in a final homage to jargon (and insult to comprehensibility) with the use of the word *intermediate* in a sense that is highly specific to chemistry.

What went wrong? We may identify at least three principal problems, which (in less extreme forms) are pervasive in modern lexicography: 1) confusion of essential properties (e.g. “sweet”, “consists of sucrose”), which may reasonably be expected to contribute to a definition, with accidental properties (e.g. “used as a condiment and preservative for other foods and for drugs”), which are incidental or indeed irrelevant to definition; 2) excessive reverence for scientific correctness coupled with indifference to making the text understandable by ordinary readers; 3) theoretical ignorance, in particular of the fuzzy and variable nature of word meaning. The latter problem is hardly surprising, as this dictionary was compiled in the 1950s, whereas the importance of prototype theory was not fully recognized until the 1970s. What is more surprising is that forty years later many 21st-century lexicographers continue to display profound ignorance of prototype theory.

*MWIII* was savaged in America by journalists and pedants alike when it was first published, mainly because it was perceived as being insufficiently prescriptive (see Sledd and Ebbitt 1962, Morton 1994). However, the weaknesses of definition, lack of an apparatus for describing register, and hard-to-read typography are more serious faults, though not the main subject of the general outcry.

The Merriam dictionaries trace their history back to the *American Dictionary of the English Language* compiled by the polemical lexicographer Noah Webster in 1828. It contains no less than 70,000 entries. Webster (1758–1843) was an indefatigable collector of words with a rare gift for definition writing. Only some of his definitions were taken directly from Johnson’s dictionary, and he introduced some sensible spelling reforms (*color*, *center*) into American English, although unfortunately some of them (e.g. *tung* for *tongue*) did not achieve acceptance by the American public. At the same time, he added and defined Americanisms such as *caucus* and *wigwam*. A fuller account of this extraordinary man, his achievement, and his legacy will be found in Micklethwait (2000). Unfortunately, his etymologies were influenced by his belief that modern languages, including English, are derived from something called Chaldaean, which he believed was the language used by Adam and God for their conversations in the Garden of Eden and the immediate precursor of Hebrew. After his death, his successors—including his son-in-law, Chauncey H. Goodrich, and the redoubtable Noah Porter, president of Yale College—quietly abandoned the Chaldaean hypothesis and brought the etymologies into line with the findings of Germanic and Indo-European scholarship.

In continental Europe, the academies did not maintain a monopoly on dictionaries on historical principles. For French, *Le trésor de la langue française* is a massive dictionary, with nearly half a million citations from literature, of the French language as it developed from 1789 (the Revolution) to 1960. A modern Italian dictionary on historical principles equivalent to OED is Salvatore Battaglia’s *Grande dizionario della lingua italiana* (1960).

## 22.8 The Russian tradition

In the English speaking world since the nineteenth century, dictionary-making and linguistics developed in such a way that the two camps ended up having little common ground and being more or less incapable of having a sensible conversation with each other about matters that might be supposed to be of mutual interest. In the English-speaking world (America in particular), ignorance, arrogance, hostility, and suspicion have been prevalent on both sides, mixed with not a little mutual contempt. ’Twas not ever thus. In the Russian tradition, there has been a long and harmonious relationship between lexicography and linguistic theory. Vladimir Ivanovich Dal (1801–1872) was a comparative linguist who did primary research on at least four of the Turkic languages of the Russian Empire. Between 1862 and 1866 he

published a massive four-volume *Explanatory Dictionary of the Living Great Russian Language*, of which new, expanded editions were regularly published after his death. The latest revision appeared in 1955 and has been reprinted many times since. It has also been used, with minor revisions, as the basic text for recent publications such as the *Illustrated Explanatory Dictionary of the Living Great Russian Language* (2007). Dal's dictionary achieved for the Russian language what the brothers Grimm were attempting to do for German at around the same time, but with a difference. Dal's interest was not only in language and its workings but also in culture, literature, and folklore expressed through language. He also published a collection of over 30,000 *Sayings and Bywords of the Russian people*. In keeping with the fashion of his time, he favoured coinages based on native Russian morphemes to express novel concepts, rather than foreign borrowings from Greek, Latin, or any other language.

Two successive revised editions of Dal's dictionary were prepared by Jan Niecisław Ignacy Baudouin de Courtenay (1845–1929), a Polish Slavicist of French extraction, also known in Russian under the name Ivan Aleksandrovič Boduen de Kurtene. Baudouin de Courtenay was a first-rate theoretical linguist, a founder member of the Prague Linguistic Circle. Among his many achievements was the development of the theory of the phoneme, later perfected by Roman Jakobson. Like his Swiss contemporary Ferdinand de Saussure, Baudouin de Courtenay was a champion of synchronic linguistics at a time when only historical Indo-Europeanist studies was regarded as academically respectable. Unlike Saussure, he was involved in practical lexicography, and he was able to bring a number of improvements to Dal's dictionary, systematically revising the methodology as well as the coverage.

The next Russian lexicographer-linguist who must be mentioned is Lev V. Ščerba (1880–1944). In contrast to Saussure, Ščerba identified three rather than two objects of study: speech activity, language system, and language material. In his theoretical work, he emphasized human linguistic creativity: the capacity of speakers to produce sentences never previously heard. He also emphasized the importance of experimentation in linguistics, particularly experiments yielding negative results—utterances that a language as a system does not allow—which of course can never be recognized through analysis of any corpus, however large.

Ščerba's pupil Sergei I. Ožegov (1900–1964) was to become the editor of another standard Soviet dictionary, the *Dictionary and Culture of Russian Speech*, the editorship of which he

inherited from Dmitri N. Ušakov (1873–1942). After Ožegov’s death the dictionary was regularly updated by Academician Natalia J. Švedova (1916–2009), while in 2007 a competing revised edition by L.I. Skvortsov appeared. Despite reputedly being described by the novelist Vladimir Nabokov as “moronic”, Ožegov’s is still the most widely used Russian dictionary today. It is not clear whether Nabokov’s objection is to dictionaries in general or Ožegov’s work in particular. It is noteworthy that, in the best Russian tradition, Švedova not only maintained a major dictionary but also wrote a grammar of the Russian language.

Two of the most important recent Russian contributors to linguistic theory have been lexicographers, namely Jurij D. Apresjan (born 1930) and Igor A. Mel’čuk (born 1932). Apresjan was a bilingual lexicographer at the Russian Academy of Sciences in Moscow, who, among other things, compiled an English-Russian Synonym Dictionary (1979). His observations of regular semantic patterns in language led to his theory of regular polysemy (Apresjan, 1973) and his book *Systematic Lexicography* (2000). Apresjan argues that lexicographers have a duty to represent the particular world view that is encoded in the lexicon of a particular language. This leads to an interaction between words (which represent beliefs) and idiomatic phraseology. Apresjan argues that a command of lexical synonyms and their subtle differences plays a vital role in enabling a speaker to express his or her thoughts in any language or culture. He says, for example:

Each of the adjectives *healthy*, *healthful*, *wholesome*, *salubrious*, and *salutary* has the sense ‘fostering the improvement or maintenance of health’. Thus, if we say *a salubrious diet*, *salubrious food*, or *a salubrious way of life*, we are making no semantic error: in principle the synonym selected is capable of expressing the required idea and we may be assured that we will be correctly understood. Nevertheless, none of the above collocations is fully correct (the best choices will be: *a healthy diet*; *wholesome food*, *a healthy way of life*). Each of them violates a co-occurrence constraint, which, though not binding, is observed in pedantic and literary discourse, and requires that *salubrious*, unlike all its synonyms, be used chiefly with the nouns *air* and *climate*.

Here we see Apresjan, who did not have the advantage of corpus evidence and corpus tools, struggling, by using his intuitions to account for the phenomenon of collocational preferences, which no modern lexicographer can afford to ignore. The principle of collocational preference is correctly understood, but the details are sometimes wrong, because Apresjan did not have sufficient evidence at his disposal.

Mel'čuk, with co-workers who included Apresjan, compiled a fragment for an *Explanatory Combinatorial Dictionary* of Russian. A year after being forced to flee from the Soviet Union in 1976 for his support of political dissidents, Mel'čuk accepted a research and teaching post at the University of Montreal, Canada, where he set up an *Explanatory Combinatorial Dictionary* of modern French (*Dictionnaire explicatif et combinatoire du français contemporain*; 'DEC'; 1984, 1987, 1993). Despite its three volumes, this work does not offer anything like full coverage of the lexicon of French; instead it elaborates a theory: Meaning -Text Theory. According to this, a natural language is conceived as "a specific set of correspondences between an infinite set of meanings and an infinite set of texts". Mel'čuk's aim is to show that there is a wide range of lexical relations in text, which are governed by a finite set of lexical functions. For example, the lexical function *Magn*, which denotes the ways in which a lexical unit can be intensified, is realized by different words in different contexts: thus, the noun *maladie* 'illness' is intensified with the adjectives *serieuse*, *grave*, etc., while the verb *remercier* 'to thank' is intensified with the adverbs *vivement*, *chaleureusement*, and *de tout coeur*. Part of Mel'čuk's importance as a linguistic theorist is simply that he assigned a central role to the lexicon in understanding the nature of language at a time when others were focused obsessively on syntax:

Most current linguistic theories view a linguistic description of a language as a grammar; a lexicon is taken to be an indispensable but somehow less interesting annex to this grammar, where all the idiosyncrasies and irregularities that cannot be successfully covered by the grammar are stored. By contrast, Meaning-Text Theory considers the lexicon as the central, pivotal component of a linguistic description; the grammar is no more than a set of generalizations over the lexicon, secondary to it. — Mel'čuk (2006)

This all-too-brief summary section has given some indication of the relationship between lexicography and linguistic theory in the Russian tradition. It is now time to turn to the practical concerns of lexicography in the English-speaking world.

## 22.9 Synchronic principles

### 22.9.1 *The American tradition*

In a dictionary on synchronic principles, the aim is to describe the current conventions of usage and meaning. The usual modern meaning of the word is placed first, followed by other, less frequent senses in some sort of logical order, and the etymology comes at the end. Thus, in the (*New*) *Oxford Dictionary of English* (1998; (N)ODE<sup>3</sup>), *camera* is defined first as “a device for recording visual images in the form of photographs or video signals”. The sense “a chamber or round building” is recorded as a separate homograph—that is, it is regarded by the dictionary as a different word that just happens to have the same spelling.

The first dictionary to issue an explicit challenge to historical principles was Funk and Wagnall’s *Standard Dictionary of the English Language* (1894–97; F&W). F&W was conceived as a popular dictionary, albeit on a grand scale, and its editors therefore made little attempt to justify their innovations in scholarly terms or to draw attention to the difficulty of what they were doing. F&W recognized that most ordinary dictionary users are more likely to want to know what a word means in the contemporary language than to ask questions about its etymology and archaic or historical usages.

The *American College Dictionary* (*ACD* 1947), edited by Clarence Barnhart, was a dictionary that set out quite explicitly to place the current meaning of each word first, following (without acknowledgement to F&W) the commonsensical principles of organization first adumbrated by Isaac Funk half a century earlier. *ACD* represented the best practices of American synchronic lexicography in the twentieth century, and it was to become the ancestor of a worldwide family of derivative dictionaries, including the *Random House Dictionary of the English Language* (*RHD* 1966, 1987, American English); the *Hamlyn Encyclopedic World Dictionary* (1971, British English); and the *Macquarie Dictionary* (1981, Australian and New Zealand English). Lexicography is typically accretive—each new dictionary building on foundations laid by its predecessors.

In his preface to *ACD*, Barnhart explained his descriptive synchronic principles thus:

This dictionary records the usage of the speakers and writers of our language; no dictionary founded on the methods of modern scholarship can prescribe as to usage; it can only inform on the basis of the facts of usage. A good dictionary is a guide to usage much as a good map tells you the nature of the terrain

---

<sup>3</sup> The word “New” was dropped from the title of the second edition (2005).

over which you may want to travel. It is not the function of the dictionary-maker to tell you how to speak, any more than it is the function of the mapmaker to move rivers or rearrange mountains or fill in lakes. ... To select the words and meanings needed by the general user, we utilized the Lorge-Thorndike Semantic Count which measures the occurrences of various meanings in the general vocabulary. By using this count, which is based upon a reading of modern standard literature, we have been able to select the important meanings needed by the reader of today and to have some statistical assurance of the occurrence of the meanings. This count has also been of considerable importance in the arrangement of meanings, since it has enabled us to determine with some certainty which are the common meanings and to put them first.

Modern corpus-driven lexicographers may be forgiven a wry smile at Barnhart's glib assurances, for even with sophisticated computational techniques and corpora many times larger than that of Lorge and Thorndike, it is still difficult, for some words, to establish which meaning is the most frequent one. For example, what is the most frequent modern meaning of *admit*? Is it "to say reluctantly" or is it "to allow to enter"? It is difficult to answer such questions with confidence, even with corpus evidence. Without it, we are merely guessing. Moreover, there are no generally agreed criteria for deciding where one meaning of a word ends and another begins, nor even for what counts as a meaning. Indeed, some lexicographers (see Kilgarriff 1997) go so far as to deny the very existence of word meanings. Hanks (1994) agrees that, strictly speaking, words do not have meanings, but goes on to argue that what dictionaries offer are statements of 'meaning potentials'—the potential of a word to make a given meaning when used in a particular context. Should launching a boat be a separate sense of the verb *launch* from launching a newly built ship? *ACD* has them as separate senses of the verb *launch*, but many people would say that they are one and the same. *ACD* (1947) does not record *launching a missile or rocket*: that sense developed later. Many people nowadays would regard this as the most literal sense, which should come first. The language has changed in this respect, and synchronic lexicographers must respond accordingly. Even when one meaning of a word has been successfully distinguished from another, it is by no means clear which one should be placed first. For example, *ACD* gives as definition 1 of the verb *launch*, "to set (a boat) afloat; lower into the water". Sense 4 is "to set going: *to launch a scheme*." Corpus analysis shows that sense 4 is much more common than sense 1, and the same was almost certainly true in 1947. Yet Barnhart's decision with regard to the arrangement of the senses of this word is defensible. The idea that launching is something that you do primarily to boats (or missiles) rather than schemes is cognitively salient for English speakers. For that reason it deserves first place, even in a synchronic dictionary.



Senses involving ‘imageable’ concrete objects and events have cognitive preference over abstract notions. Thus, “launching a boat or ship” can be seen as activating the most literal sense of this verb, while “launching a scheme (or a new product)” can be interpreted as a metaphor exploiting the boat or missile sense. If the most frequent sense of a word is perceived as being a linguistic metaphor exploiting another, more literal sense, it takes second place in Barnhart’s dictionary, regardless of frequency.

The leading present-day dictionary in America on synchronic principles is the *American Heritage Dictionary* (1969; 4th edition 2006), which may be regarded as carrying on the tradition of F&W and ACD, even though there is no formal relationship among these works.

### **22.9.2 The British tradition**

In Britain, synchronic principles were introduced from America, first by the *Hamlyn Encyclopedic World Dictionary* (1971) and subsequently by *Collins English Dictionary* (CED; 1979), which greatly extended the lexicographic coverage of scientific and technical words compared with other dictionaries of the time. The one-volume (*New*) *Oxford Dictionary of English* (*NODE*, 1998; 2nd edition, *ODE* 2005; 3rd edition 2010) is a one-volume dictionary on synchronic principles, more similar in design and structure to *RHD*, *AHD*, and Collins than to the great historical dictionary (*OED*) published by the same publishing house. It is based on an unrivalled body of citation evidence, for it is the only dictionary of English aimed at general users to use analysis of corpus evidence as an organizing principle for arranging and refining the definitions of complex words, as well as a source of citations of actual usage. For unusual words and senses, it draws on citations collected by the *OED*’s traditional reading programme. Among other things, (*N*)*ODE* adopts a more sophisticated approach to word grammar than most monolingual dictionaries aimed at the home market. It attempts, not always successfully, to identify “core meanings” and group subsenses under a core meaning.

Some readers may wonder why America’s favourite dictionary (if sales are anything to go by) has not been mentioned in this brief survey of synchronic dictionaries. The dictionary in question is *Merriam Webster’s Collegiate*. The reason for this omission is simple: the *Merriam Webster Collegiate* is a dictionary based on historical principles; it is not a synchronic dictionary. It is based on Merriam’s vast *Third New International Dictionary* of 1961 and its two predecessors. The current edition of the *Collegiate* is the 11th edition (2001). According to the publisher, a 12th edition is due to appear in 2014. It will be

interesting to see whether this new edition will adhere to the long-standing Merriam preference for historical principles.

### **22.9.3 Synchronic lexicography in other countries**

In this section so far, examples of synchronic dictionaries have been taken from the English-speaking world. Only brief and selective mention can be made of modern synchronic dictionaries in other languages, sufficient perhaps to illustrate the variety of different social linguistic functions that a synchronic dictionary is expected to perform in different cultures.

**German-speaking lands:** In 1880 the schoolteacher Konrad Duden published a spelling dictionary, *Die deutsche Rechtschreibung*, which became accepted as the de facto standard reference for German spelling in German-speaking lands. This was the first in a series of reference books on different aspects of the German language, including a synonym dictionary, a guide to usage, a pronunciation dictionary, a historical and etymological dictionary, a children's dictionary, and other volumes. The publisher's flagship is the *Deutsches Universalwörterbuch*, the third edition of which was published in 2011 and is available online. It is compiled on modern descriptive principles.

During the Communist era, state funding of lexicographical research was normal in central and Eastern Europe. One of the finest works of synchronic lexicography created during this era was the *Wörterbuch der deutschen Gegenwartssprache* (WDG; 1964-77), edited by Ruth Klappenbach and Wolfgang Steinitz. If we can bring ourselves to ignore the occasional 'politically correct' entries for terms like *Kapitalismus* and *Sozialismus* and the extensive coverage of the terminology of the Volkspolizei, we find a very fine dictionary based on modern descriptive principles, making a real effort to account for the phraseology in which each word is used. Unfortunately, in the 1960s and 70s, the editors did not have access to large electronic corpora, which had not yet been invented, so although their principles repay close examination, the details of the implementation is sometimes deficient. They had the same problem as Apresjan: insufficient evidence for the description of normal phraseology.

**Greece:** Throughout the 19th century, since Greece won independence from Turkey, attitudes to the modern Greek language have been divided between pragmatists, willing to accept the language as it is, and purists, with a desire to purify the language and in particular to expunge all the Turkish words which had come in since the Turkish occupation of the 16th-19th centuries. In modern Greek, the triumph of Dimotiki, the everyday form of the language, over Katharevousa, the archaizing 'purified' form of the language, is now pretty

well total, thanks in no small measure to the success of the dictionary of Giorgios Babiniotis, which follows rigorously descriptive principles.

**Czech Republic and Slovakia:** In Czech, something very different has happened. Modern literary Czech is an artificially constructed language, harking back to the language of the 16th century, when Bohemia was a European great power. The Dictionary of the Standard Czech Language (*Slovník spisovného jazyka českého*, ‘SSJČ’; 1960-71) therefore has a prescriptive function, legislating about correct vs. incorrect usage of words and occasionally inventing a word or a meaning to fill a lexical gap. A department of the Institute for the Czech Language has a similar prescriptive or advisory function. SSJČ is overdue for revision or replacement, and work started some years ago on a project that now seems to have been discontinued. There does not seem to be any general agreement on the principles on which a new Czech dictionary should be based.

Neighbouring Slovakia, by contrast, is richly furnished with the up-to-date and ongoing products of a fine lexicographic tradition, including not only a great historical dictionary and a dialect dictionary, but also a fully descriptive Dictionary of Contemporary Slovak (*Slovník súčasného slovenského jazyka*), of which volumes 1 and 2 (A-G; H-L) have been published.

**Spain:** A fine dictionary of Castilian Spanish compiled on synchronic principles is the *Diccionario del español actual* (1999; 2nd edition 2011), compiled by Manuel Seco, Olimpia Andrés, and Gabino Ramos, which gives a detailed and explicit account of contemporary Spanish, using evidence of contemporary usage culled from the Internet and other sources. Maria Moliner’s *Diccionario de uso del español* (1966-67; 3rd edition 2007) is important because it was originally aimed specifically at repairing the deficiencies of the Royal Academy’s dictionary (*DRAE*) in accounting for word usage in contemporary Spanish and this gave a new impetus to lexicography in Spanish.

**Catalonia:** The Catalan language plays a central role in the Catalan sense of national identity, so it is not surprising that the lexicon of Catalan and closely related languages has been painstakingly inventorized and defined in the *Diccionari català-valencià-balear: inventari lexicogràfic i etimològic de la llengua catalana* by Antoni M. Alcover and Francesc de Borja (Moll, 1988).

**Italy:** Two important Italian dictionaries on synchronic principles are Tullio de Mauro’s *Grande dizionario italiano dell’uso* (‘GRADIT’: 1999-2000) and the *Dizionario Italiano Sabatini Coletti* (‘DISC’: 1997).

**France:** French monolingual dictionaries tend to be more strongly influenced by historical principles than a modern British or Australian reader would expect, and a careful distinction is not made between literary French and practical modern French. An exception is *Le Petit Robert* (1967), which describes itself as “une traitement moderne et soucieux de la réalité social du français”. Bilingual dictionaries in France generally adopt a thoroughly pragmatic synchronic approach to language description.

This rapid and superficial survey of synchronic dictionaries in various cultures is a less than adequate treatment of the subject, but it should be sufficient to illustrate the growing emphasis on empirical description of the contemporary language that is characteristic of lexicography in many but by no means all languages and cultures in the modern world. Many of these dictionaries are aimed at helping foreign learners of a language, a subject to which we now turn.

## 22.10 Dictionaries for language learning

During the 1930s a major development in English lexicography took place in Japan, a development that was eventually to have an effect on lexicography in other languages too. The linguist Harold Palmer, founder of the Institute for Research in English Teaching, the English teacher A. S. Hornby, and some other teachers of English in Japan observed that the then-current dictionaries of English were not suitable for foreign learners of English and decided to do something about it. The result was the *Idiomatic and Syntactic Dictionary (ISED)*, developed and tested in Japanese classrooms and published by Kaitakusha just after the outbreak of World War II. This work was designed primarily as a dictionary for encoding purposes, that is, to help learners with their writing and speaking skills. It contains a deliberately limited selection of vocabulary—words that were in active use and that learners might be expected to know and to be able to use correctly and idiomatically. The apparatus gave a great deal of information about the syntactic structures associated with each word. Hornby’s verb patterns in particular were in use among English language teachers for almost half a century before eventually being superseded by corpus-based research. *ISED* was republished unaltered in 1948 by Oxford University Press as *A Learner’s Dictionary of Current English*, subsequently re-titled the *Oxford Advance Learner’s Dictionary of Current English (OALDCE)*. A massive influx of additional vocabulary items was added to the 2nd edition, which diminished rather than enhanced its original intention as an encoding tool for learners. The editors had ceased to ask themselves the unanswerable question, “Does a

learner need to know how to use this word idiomatically?” For, of course, different learners need different words for different circumstances. The 6th edition, edited by Sally Wehmeier (2000), was extensively revised using evidence from the British National Corpus, while adhering to the principle that vocabulary selection, definitions, and examples of usage must be driven by classroom needs rather than corpus evidence. It is therefore unabashed about using invented examples alongside or instead of text-derived examples of usage. 2011 saw its eighth edition.

In 1978, the supremacy of *OALDCE* in the marketplace for EFL (English as a foreign language) was challenged by the *Longman Dictionary of Contemporary English (LDOCE)*; (<http://www.ldoceonline.com/>). This is the dictionary of choice for many researchers in computational linguistics. Like *OALDCE*, it is driven by perceived classroom needs, but was extensively revised in the 1990s using evidence from the British National Corpus. It devotes considerable attention to spoken English.

In 1987, with the publication of the COBUILD dictionary (an acronym for ‘Collins Birmingham University International Language Database’, 1987, 1995), a radical new kind of lexicography emerged: the corpus-driven dictionary. COBUILD’s innovations included examples selected from actual usage for naturalness, rather than invented by the lexicographer or teacher, while its unique defining style expresses links between meaning and use by encoding the target word in its most typical phraseology (e.g. “when a horse *gallops*, it runs very fast so that all four legs are off the ground at the same time”). The editor-in-chief of COBUILD, John Sinclair, briefed his editorial team: “Every distinction in meaning is associated with a distinction in form.” This was more a signpost for the future than a practical guideline for interpreting the then-available evidence. A great deal more research is required to determine exactly what counts as a distinction in meaning, what counts as a distinction in form, and what is the nature of the association. COBUILD was the first ever large-scale corpus-based dictionary research project. Its principles were set out in an associated book of essays (Sinclair, ed., 1987). Unfortunately, a few years later the Cobuild research programme was cut short by News International, which had bought Collins, the publisher funding the work.

Another addition to the stock of corpus-based dictionaries for learners of English was the *Cambridge International Dictionary of English (CIDE 1995)*. Subsequent editions (2003, 2005, 2008) were published as the *Cambridge Advanced Learner’s Dictionary* (<http://dictionary.cambridge.org>). This work has a number of associated data modules, such as lists

of verb complementation patterns, semantic classifications of nouns, and semantic domain categories. The second edition and subsequent editions were re-titled *Cambridge Advanced Learners Dictionary (CALD)*.

The most recent addition to the stock of such dictionaries published in Britain is the *Macmillan English Dictionary for Advanced Learners (MEDAL 2002)*. This dictionary is corpus-based but not corpus-driven. It makes eclectic use of some of the principles developed for other major lexicographical projects, and pays special attention to two things in particular: conventional metaphors and collocations. For the latter, it uses the Sketch Engine, a computer program that identifies statistically significant collocations of each target word, which the lexicographers were in many cases able to associate with specific senses of the target word.

In 2008 Merriam-Webster brought out *Merriam-Webster's Advanced Learner's English Dictionary*. This is a practical American work, with a sensible selection of currently used words and meanings in American English. It owes more to the definitions in rival British EFL dictionaries than to the Merriam tradition of historical lexicography and it pays little or no attention to primary research in phraseology, cognitive linguistics, or corpus linguistics.

In his 1987 paper, entitled 'The nature of the evidence', Sinclair stresses the importance of distinguishing significant collocations from random co-occurrences. The first attempt to undertake statistical analysis of collocations in a corpus for lexicographical and other purposes was by Church and Hanks (1990), but it was not until Kilgarriff, Rychlý, and their colleagues developed the Word Sketch Engine (Kilgarriff *et al.* 2004) that a user-friendly tool was made widely available for people to see at a glance how the meanings of a semantically complex word are associated with and indeed activated by its collocates.

## **22.11 The impact of computer technology on modern lexicography**

Section 22.3 mentioned the impact of the invention of printing on Renaissance lexicography. A comparable impact has been had on modern lexicography by computational text processing. There are three aspects to this impact: compilation, evidence, and use.

### **22.11.1 Computers and dictionary compilation**

In the 1960s and 70s some adventurous lexicographers found that they could be freed by the computer from the tyranny of alphabetical order and proceed instead in a logical order, dictated by content rather than the vagaries of the alphabet. So, for example, the editor writing medical entries would work systematically through the field, starting, say, with

definitions of terms denoting bones and organs of the body, before moving on to physiology, pathology, diseases, clinical psychology, and so on. Simultaneously, specialists in the arts could make their contributions by defining terms of, say, music, ballet, opera, and theatre, while others contributed the terminology of poetics, printing, and publishing. Meanwhile, a phonetician would write phonological transcriptions, while a team of etymologists summarized what is known about the origin and history of each word. It was no longer necessary for the special-subject editors to be polymaths with competence in grammar, phonology, and etymology. These various contributions were then slotted by computer into a framework of general definitions compiled by a team of general editors. A further group of editors would read through the text of each entry, correcting errors and inconsistencies, eliminating duplications, plugging gaps, and generally polishing up the work for publication. This, in very broad outline, is how the large one-volume synchronic dictionaries of the second half of the twentieth century (notably the *Random House Dictionary of the English Language*, *American Heritage Dictionary*, *Collins English Dictionary*, and *New Oxford Dictionary of English*) were compiled, with consequent improvements in quality and a dramatic reduction in elapsed time between start-up and publication. Because the text is compiled in a database or structured text file and because each dictionary entry has a basic uniformity of structure, the dictionary text can be run through a typesetting program and output page proofs in a matter of hours rather than months.

This aspect of lexicographical technology encouraged the editors of such dictionaries to ride roughshod over the traditional distinction between a dictionary and an encyclopedia, and to take the view instead that a dictionary is a sort of collective cultural index, which must summarize, for the practical benefit of users, all the most salient cognitive and social features associated with the meaning of every word and name that is in common currency.

### **22.11.2 Lexical evidence**

An even more important development, from the point of view of studying words and how they go together in idiomatic language use, was the emergence in the 1980s and 1990s of corpus linguistics. Up to that time, lexicographers had insufficient evidence to represent accurately the conventions of word meaning and word use. Corpus evidence changed the nature of lexicography. It demonstrated clearly that definitions in pre-corpus dictionaries had a tendency to be biased in favour of unusual rather than central and typical uses of words, and that introspection is not a good source of evidence. These developments have been fully described elsewhere, for example by Hanks (2009), and there is no need to repeat them here.

Just one example will suffice to illustrate the radical impact that corpus technology has begun to have on lexicography. This concerns the meaning of the conventional metaphor *gleam*. Conventional metaphors are secondary senses of words and as such are (or ought to be) recorded in dictionaries. There is no disputing that the primary meaning of *gleam* is “a faint or brief light”, but what is its secondary meaning, applied to an emotion appearing briefly in someone’s eyes? Consulting their intuitions, cognitive linguists have invented examples such as “Amusement gleamed in his eyes” as a supposed realization of the conceptual metaphor HAPPINESS IS LIGHT. This hypothesis appears to be supported by OED’s sense 2b of the noun *gleam*, “a bright or joyous look.” But, as Deignan (2005) points out, corpus evidence shows that in the 20th century (at any rate) a gleam in someone’s eyes does not normally signal happiness, but rather cynical amusement, mischief, or even malice. And even OED, a historical, pre-corpus dictionary par excellence, supports its definition with a citation from 1852 that might set alarm bells ringing in the head of an alert reader:

1852 H. B. Stowe *Uncle Tom's Cabin* vi. His black visage lighted up with a curious, mischievous gleam.

The following examples, selected from the British National Corpus (BNC), are typical of 20th-century usage of this word in its secondary, metaphorical sense.

Rosita looks at me indignantly, with a furious *gleam* in her eyes, a look of hatred.

He had a zealot's *gleam* in his dishwater eyes.

... the sardonic *gleam* in his eyes.

... a rather nasty *gleam* in his blue eyes.

[She] didn't understand the wicked *gleam* in his eye

His eyes *gleamed* malevolently.

These are only six of sixty or more examples in BNC that could have been selected to illustrate this point. They are not matched by other examples designating happiness. These examples also illustrate another important contribution of corpus linguistics to lexicography, namely the identification of collocations. Collocations are recurrent co-occurrences of words in different texts. The word *gleam* collocates significantly with *eye*, but also with the adjectives *sardonic*, *mischievous*, *unsettling*, *predatory*, *manic*, *visionary*, *wry*, *wicked*, *amused*, *cynical*, *fierce*, and *mad*. And a *gleam* is (in descending order of statistical significance) a gleam of *amusement*, *malice*, *triumph*, or *humour*. It seems safe to predict that dictionaries of the future, in the age of the Internet and large corpora, will pay far more



careful attention than previously to collocation and phraseology, using various measures of statistical significance to identify salient collocations, and that this new trend, marching arm in arm with other developments such as construction grammar, will continue to bring about a change that has already begun in perceptions among linguists of the relationship between words and meaning.

### 22.11.2 On-line dictionaries

Ironically, the revolution that has brought exciting new potential for lexical description has at the same time destroyed the business model that traditionally would have funded such work. Compiling a new dictionary is a huge, expensive, labour-intensive task, but if every well-educated member of a community feels the need to own a dictionary the financial incentives are likewise substantial. In the heyday of synchronic lexicography in the 20th century, there were half a dozen dictionary publishers competing for market share. With the advent of the Internet, all that has changed. The market for printed dictionaries on paper has sharply declined. Dictionaries are typically used for rapid and uncritical look-up, for which the Internet is ideally suited, but the Internet offers a free-for-all, in which some very inferior and indeed inaccurate products jostle for position with some very sophisticated accounts of words and their meanings. The OED on-line must be singled out for mention as an example of the best that on-line lexicography can offer. The content of the dictionary is based on 19th-century principles (this fact alone is a tribute to the robustness of James Murray's linguistic insight and lexicographic skill), while the techniques of information retrieval and presentation are at the cutting edge of modern lexicographical technology. It remains to be seen whether new business models (or funding models) will emerge that will enable new lexicographical projects to undertake large-scale, detailed (and possibly cross-linguistic) investigations of words, their collocations, their phraseological patterns and their meaning.

## 22.12 Thesauruses and ontologies

Almost all the dictionaries mentioned so far are semasiological – that is, they start with a word or phrase and ask how it spelled, how it is pronounced, what it means, etc. Before concluding, brief mention must be made of an alternative approach to the lexicon, namely onomasiology, which starts with a concept and asks, is there a word or phrase to express it?

During the European Enlightenment, starting in the 17th century, attempts were made to

arrange all human knowledge in conceptual hierarchies. Since concepts can only be represented by words, this is necessarily a quasi-lexicographical undertaking.

The most important of these 17th-century conceptual and lexical models of the universe forms part of John Wilkins' *Essay towards a Real Character and a Philosophical Language* (1668), a vast and astonishing work (the term "Essay" in the title is misleading), which contains among other things an attempt to summarize and organize all conceptual knowledge. The starting point is that it seems obvious that a *dog* is a kind of *animal* and an *animal* is a kind of *physical object* and a *physical object* is a kind of *entity*. Wilkins assumed that all concepts could be arranged in hierarchies of this sort, applicable this hierarchical schema to all words and all concepts in a way that would be universal to all languages. This central part of Wilkins' *Essay* is a forerunner of Roget's famous *Thesaurus* (1852), as Peter Mark Roget himself acknowledged. It is also a direct predecessor of WordNet (Miller 1985, Fellbaum 1998). In the words of Eco (1995), Wilkins' *Essay* was "the most complete project for a universal and artificial philosophical language that the 17th century was ever to produce." As a preliminary step, Wilkins undertook a review of all knowledge, "to establish what the notions held in common by all rational beings really were". The philosopher and logician Leibniz attempted an emulation (in Latin) of Wilkins' work, including a "table of definitions", but abandoned it after compiling only a few entries. The difficulty, in a world before Linnaeus, of building a satisfactory conceptual hierarchy of this sort can be illustrated with the word *dog*. Wilkins starts this part of his ontology by remarking that "Beasts" "may be distinguished by their several shapes, properties, uses, foods, their tameness or wildness, etc." He lumps dogs together with cats as being "rapacious" but not "cloven-footed". He distinguishes dogs from wolves because wolves howl but dogs bark, bay, or yelp. Here, as Eco remarked, Wilkins seems to be reaching for the modern concept of hypertext. "Rapacious beasts of the dog-kind" include not only *dogs* and *wolves*, but also *foxes* and *badgers* and "amphibious beasts of the dog-kind", namely *seals*. (Seals bark, don't they?)

Only a person of overweening intellectual self-confidence and demented energy could have even dreamed of such an undertaking. One of many questions begged by it is, can a Wilkinsian hierarchy of concepts be equated with or represented satisfactorily by a lexical hierarchy? It must be admitted that badgers and seals have quite a lot in common with dogs, wolves, and foxes. However, any schoolchild nowadays will tell you that neither seals nor badgers are really "of the dog-kind". The place of dog in a post-Linnaean hierarchy such as WordNet is rather different from the place assigned to it by Wilkins.

The full conceptual hierarchy for *badger* in WordNet is:

badger > musteline mammal > carnivore > placental mammal > mammal > vertebrate > chordate > animal > organism > living thing > whole > object > physical entity > entity

The conceptual hierarchy for *dog* begins with canine > carnivore. Thus, badgers are included in a set, not only with dogs, but also with all other carnivores, including felines (cats, lions etc. – “rapacious beasts”, in Wilkins’s terminology) but also bears, “procyonids” (whatever they may be—raccoons, apparently), and “fissiped carnivorous mammals”. It will be readily seen that this conceptual hierarchy has little to do with everyday usage of language and very much to do with the organization of scientific concepts.

The great philosophers of the 17th century, including not only Wilkins and Leibniz but also Hobbes, Comenius, and others, took it for granted that the obvious vagueness and fuzziness of word meaning were defects of natural language, which ought to be rectified. It was not until the 20th century, with the work of Wittgenstein, Putnam, Rosch, and others, that an alternative view began to emerge, namely that vagueness and fuzziness might be essential properties—design features, we might say—of natural language. The natural human yearning for conceptual precision can easily be satisfied by creating stipulative definitions (e.g. “I hereby assert that an *idea* is a kind of *concept* and not vice versa”), but we should not imagine that such definitions can be equated with the meaning of terms in a natural language. The task of the present-day lexicographer is to account for the vague and variable conventions of word meaning in natural usage, not to build conceptual hierarchies.

Confusion between scientific concept meaning and natural-language word meaning continues to bedevil the study of meaning in language and the false assumptions that it has generated must bear at least part of the responsibility for some of the failures of linguistics in Natural Language Processing. The attempt to make language precise was based on false assumptions about the relationship between scientific concepts and the everyday meaning of words and these are with us to this day. Lexicography has, so far, been slow to respond to the challenges and insights of 20th-century linguistic philosophy and anthropology.

### **22.13 Conclusion and future prospects**

In this brief survey of lexicography throughout the world from earliest times, I have tried to show how dictionaries have played a central role, not only in linguistics (the study and understanding of language), but also in the many and various conceptualizations of human

cultures. A dictionary is an inventory of words, and an inventory of words is an inventory of basic beliefs. Such beliefs may or may not be well founded, and a dictionary may or may not do a good job of encapsulating them, but if the lexicographer does not take a stand and fashion a view of the beliefs of the culture that he or she is describing, then dictionary definitions cannot be written at all. A consequence of this is that definitions in monolingual dictionaries are necessarily circular: all words are defined in terms of other words. Logicians sometimes complain about this so-called circularity of dictionary definitions, and philosophical linguists such as Wierzbicka and Goddard (2002), following the lead of the Port-Royal grammarians (Arnauld and Nicole 1662) have attempted to break the vicious circle by selecting a small number of basic words as indefinable logical primitives, which are universals in terms of which the meanings of all other words in all language can be defined.

It would be impossible to define every word. For in order to define a word it is necessary to use other words designating the idea we want to connect to the idea being defined. And if we wished to define the words used to explain that word, we would need still others and so on to infinity. Consequently, we necessarily have to stop at primitive terms which are undefined. – Arnauld and Nicole, 1662 [tr. Burker 1996].

The logic of this is impeccable, but it has nothing to do with either language or beliefs in the everyday world. In reality, it is certainly true that some words are broader in semantic scope than others—*say* is broader in scope than *whisper*, for example—but the steps from broad to narrow are more of a tangled hierarchy than an orderly progression. In practice, it is perfectly possible to compose a usable and true statement about any of the terms (such as *say*) identified by Wierzbicka and Goddard as ‘semantic primitives’, though only at the cost of circularity. The best that a practical lexicographer can hope to do is to accept the circularity but avoid direct reciprocity. If a dictionary defines a *helix* as a spiral and a *spiral* as a helix, it is vicious: something more must be said at one entry or the other, for example “winding in a continuous curve ...”.

Thus, dictionary definitions teeter uneasily on the sharp edge between the blindingly obvious and the philosophically profound. And then dictionaries nowadays are expected to give other information about words: most importantly about their orthography and morphology (inflections), but also about pronunciation, grammatical word class, and etymology or word history.

As a general rule, lexicography is accretive; one dictionary builds on another. Radical

innovations do occur (*WDG*, for example, and *COBUILD*), but they are few and far between. We have seen that there are many motivations for compiling a dictionary. In recent centuries, the main motive has been to compile an inventory of the words of a language, with summary information about conventions of usage and belief associated with each word. In the past, the function of lexicography was perceived more strongly as being to control and regulate the language. Sometimes a dictionary may have an influence on social attitudes to language. For example, the highly practical Modern Greek dictionary of Babinyiotis was one more nail in the coffin of the movement to create a ‘purified’ language (*katharevousa*), expunging words of Turkish and other non-classical origin.

In the Arabic, Hebrew, Persian, and Indian traditions, a motivation for lexicography was facilitation of poetry, and this motivation is also found in the modern English-speaking world in a modest form with the publication of rhyming dictionaries. During the Renaissance in Europe, the principal motivation for some of the greatest dictionaries ever compiled was the preservation and understanding of culture and heritage, in particular the heritage of ancient Latin and Greek literature. Surprisingly, bilingual lexicography was a slow starter. In Europe from classical times up to the 17th century, it was expected that all educated and civilized people would be able to talk to one another in Latin, so vernacular words were merely appended to monolingual Latin dictionaries, in particular the series of dictionaries known as ‘calpines’ after Ambrogio Calepino. It was not until the Enlightenment in Europe that the compilation of bilingual dictionaries became standard practice, although there were a few important precursors, notably Palsgrave (1530). In the 20th century, bilingual lexicography led the field in terms of understanding the importance of phraseology for language understanding.

More commercial motives arose in the 20th century with the advent of dictionaries for second-language learners, a practical tradition founded by A.S. Hornby in the 1940s and now a multi-million dollar business worldwide.

In the 21st century the business model for dictionary publication is rapidly switching from paper to on-line formats, and the user community has expanded to include machines: lexicography has to include provision for the needs of computational linguists and programs for natural language processing. Lexicography is currently in a state of transition. The Internet affords unrivalled opportunities for new lexicographical research, but at the same time the traditional business model of funding new developments in lexicography from prospective sales has collapsed. The book trade itself, too, is in crisis, and many booksellers

are going out of business. Hardback reference books no longer sell. The public has come to expect reference information to be free via the Internet, but unfortunately, this has opened the floodgates to a mass of free but inadequate, misleading, and even incorrect lexicographic information. Developments in electronic lexicography are surveyed from a variety of different viewpoints by a variety of authors in Granger and Paquot (2012).

## References

- Apresjan, Jurij D. 1973. 'Regular polysemy'. In *Linguistics* 142: 5-32.
- Apresjan, Jurij D. 1979. *Anglo-russkij sinonimičeskii slovar' [Dictionary of English and Russian Synonyms]*. Moscow: Russkij Jazyk.
- Apresjan, Juri D. 2000. *Systematic Lexicography*, translated by Kevin Windle. Oxford: Oxford University Press.
- Arnauld, Antoine, and Pierre Nicole. 1662. *La logique, ou l'art de pensée* [known as 'The Port-Royale Logic'.] English translation 1996 by Jill Vance Burker: *Logic or the Art of Thinking*. Cambridge: Cambridge University Press.
- Baker, C. F., C. J. Fillmore, and B. Cronin. 2003. 'The Structure of the FrameNet Database'. In: *International Journal of Lexicography*, 16.3.
- Boltz, William G. 1993. 'Shuo wen chieh tzu' in Michael Loewe (ed.), *Early Chinese Texts: a Bibliographical Guide*. Berkeley: Society for the Study of Early China, pp. 429-442.
- Castro, Americo, 1991. *Glosarios latino-españoles de la Edad Media*. Madrid: Consejo Superior de Investigaciones Cientificas.
- Cohen, Chaim, and Ya'akov Choueka. 2006. 'Hebrew Lexicography' in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, second edition; vol.5, pp.168-272. Elsevier.
- Considine, John. 2008. *Dictionaries in Early Modern Europe: Lexicography and the Making of Heritage*. Cambridge: Cambridge University Press.
- Couturat, Louis. 1901. *La Logique de Leibniz*. Paris: Felix Alcan.
- Daniels, Peter T., and William Bright (ed.). 1996. *The World's Writing Systems*. Oxford: Oxford University Press.
- Deignan, Alice. 2005. *Metaphor and Corpus Linguistics*. Amsterdam: John Benjamins.

- Drory, Rina. 1995. 'Bilingualism and cultural images: the Hebrew and the Arabic introductions to Saadia Gaon's *Sefer ha-Egron*' In Shlomo Isre'el and Rina Drory (eds), *Language and culture in the Near East*. Leiden: E. J. Brill for Israel Oriental Studies.
- Drory, Rina. 2000. *Models and contacts: Arabic literature and its impact on medieval Jewish culture*. Leiden: E. J. Brill.
- Eco, Umberto. 1995. *The Search for the Perfect Language*. Translated by James Fentress. Oxford: Blackwell.
- Fellbaum, Christiane (ed.). 1998. *WordNet: an Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Ferri, Rolando (ed.). 2011. *The Latin of Roman Lexicography*. Pisa: Fabrizio Serra Editori.
- Fried, Daniel. 2007. 'Defining Courtesy: Spenser, Calepine, and Renaissance Lexicography'. In *Review of English Studies*, 58: pp. 229-244.
- Goldberg, Adele. 1995. *Constructions: a Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Granger, Sylviane, and Magali Paquot. 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Hanks, Patrick. 2004. Corpus Pattern Analysis. In: G. Williams and S. Vessier (eds.), *Euralex Proceedings*. Vol. 1, Lorient, France, Université de Bretagne-Sud.
- Hanks, Patrick. 2006. 'English Lexicography' in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, Second Edition. Elsevier.
- Hanks, Patrick (ed.). 2008. *Lexicology: Critical Concepts in Linguistics*. 6 volumes. Routledge.
- Hanks, Patrick. 2009. 'The impact of corpora on dictionaries'. In: Paul Baker (ed.), *Contemporary Approaches to Corpus Linguistics*. Continuum.
- Hanks, Patrick. 2010. 'Lexicography, Printing Technology, and the Spread of Renaissance Culture.' In A. Dykstra et al. (eds), *Euralex Proceedings*. Leeuwarden: Frisian Institute.
- Hanks, Patrick. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.

- Haywood, John A. 1965. *Arabic lexicography: its history, and its place in the general history of lexicography*. Leiden: Brill.
- Kilgarriff, Adam. 1997. 'I don't believe in word senses'. In *Computers and the Humanities* 31: 91-113
- Kramer, Undine. 2006, 'German Lexicography' in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, Second Edition. Elsevier.
- Labarre, Albert. 1975. *Bibliographie du Dictionarium d'Ambrogio Calepino*. Koerner, Baden-Baden.
- Leibniz, Gottfried Wilhelm. See Couturat 1901.
- Li Ming. 2006. 'Chinese Lexicography' in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, Second Edition. Elsevier.
- Mel'čuk, Igor A. 2006. 'The explanatory combinatorial dictionary'. In Giandomenico Sica (ed.), *Open Problems in Linguistics and Lexicography*. Milan: Polimetrica.
- Mel'čuk, Igor A., André Clas, and Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Duculot.
- Micklethwait, David. 2000. *Noah Webster and the American Dictionary*. Macfarland.
- Miller, George. 1985. 'WordNet: a dictionary browser.' In: *Proceedings of the First International Conference on Information in Data*, University of Waterloo.
- Miranda, Rocky. 2006. 'Indian Lexicography' in K. Brown (ed.), *Encyclopedia of Language and Linguistics*, Second Edition. Elsevier.
- Morton, Herbert C. 1994. *The Story of Webster's Third: Philip Gove's Controversial Dictionary and its Critics*. Cambridge: Cambridge University Press.
- Murray, James A. H. 1900. *The Evolution of English Lexicography*. The Romanes Lecture. Oxford: Clarendon Press.
- Roth, Norman. 1994. *Jews, Visigoths, and Muslims in Medieval Spain: Cooperation and Conflict*. Leiden: E. J. Brill.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. 1998. 'The Lexical Item' in E. Weigand (ed.), *Contrastive Lexical Semantics*. Benjamins.



- Stathi, Ekaterini. 2006. 'Greek Lexicography, Classical' in K. Brown (ed.), *Encyclopedia of Language and Linguistics, Second Edition*. Elsevier.
- Sledd, James and Wilma R. Ebbitt, Wilma R. (eds.). 1962. *Dictionaries and That Dictionary*. Scott Foresman.
- Trench, Richard Chenevix. 1858. 'On some deficiencies in our English dictionaries'. In: *Proceedings of the Philological Society*.
- WDG = *Wörterbuch der deutschen Gegenwartssprache*, 5 vols., 1961-77, ed. R. Klappenbach and W. Steinitz. Akademie Verlag, Berlin. Available on line at <http://www.dwds.de/>
- Wierzbicka, Anna, and Cliff Goddard. 2002. *Meaning and Universal Grammar: Theory and Empirical Findings*. Benjamins.
- Wilkins, John. 1668. *Essay Towards a Real Character and a Philosophical Language, part II: Universal Philosophy*. London: the Royal Society. Excerpts in P. Hanks (ed.). 2008. *Lexicology: Critical Concepts in Linguistics*, vol. 1. Routledge.
- Halliday, M. A. K. 1966. 'Lexis as a linguistic level'. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.): *In Memory of J. R. Firth*. Longman.
- Hanks, P., and J. Pustejovsky. 2005. 'A pattern dictionary for natural language processing'. In *Revue Française de Linguistique Appliquée* 10:2.
- Micklethwait, David. 2000. *Noah Webster and the American Dictionary*. Jefferson, NC: Macfarland.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky, J., P. Hanks, and A. Rumshisky. 2004. 'Automated Induction of Sense in Context'. In *Proceedings of COLING 2004*, Geneva, Switzerland.
- Sinclair, J. M. 1966. 'Beginning the study of lexis'. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robins (eds.), *In Memory of J. R. Firth*. Longman.
- Sinclair, J. M. 2004. *Trust the Text: Language, Corpus and Discourse*. Routledge.
- Trench, Richard Chenevix. 1858. 'On some deficiencies in our English dictionaries'. London:

the Philological Society.

Yong Heming and Peng Jing. 2008. *Chinese Lexicography: a History from 1046 BC to AD 1911*. Oxford: Oxford University Press.

**Web sites**

CPA: Corpus Pattern Analysis: <http://nlp.fi.muni.cz/projects/cpa/>

**PDEV:** [Pattern Dictionary of English Verbs](http://pdev.org.uk): *pdev.org.uk*

FrameNet: <http://framenet.icsi.berkeley.edu/>