Lexicography: Overview

P Hanks, Brandeis University, Waltham, MA; USA and Berlin-Brandenburg Academy of Sciences, Berlin, Germany

© 2006 Elsevier Ltd. All rights reserved.

Introduction

Lexicography is the art or craft of writing dictionaries. In contrast to lexicology (*see* **Lexicology**), which is the theoretical study of words, their use, and their meaning, lexicography is a practical activity that involves compiling an inventory of the words in a language (or some particular subset of them) and saying something about each of them. Lexicographers were described by Trench (1858) as "the inventory clerks of language." Their primary duty is to collect an inventory of all the words in the language (or a selected subset), and to say something about each of them. There are many different kinds of dictionaries, giving many different kinds of information about words. The main ones are: (1) scholarly dictionaries of record; (2) practical dictionaries for everyday use; (3) pedagogical dictionaries; (4) dictionaries of linguistic phenomena such as slang or idioms; and (5) special-subject dictionaries. All the foregoing kinds are monolingual. To these we must add: (6) bilingual dictionaries; (7) onomasiological dictionaries (thesauruses, synonym dictionaries); and (8) term banks. There are also hybrid dictionaries, for example, monolingual dictionaries for second-language learners with marginal glosses in a relevant foreign language.

Four issues of general principle that must be considered for all serious types of dictionaries in any language are: (1) breadth, not depth; (2) consistency; (3) descriptive versus prescriptive approach to the language; and (4) historical versus synchronic approach. The third of these, in particular, is controversial: there are many circumstances in which a dictionary, rightly or wrongly, is expected to play a prescriptive role in a language.

General procedures for many kinds of lexicography include collecting, 'tokenizing,' and 'lemmatizing' words. Having established a draft word list, the lexicographer must decide what to say about each word. This includes decisions about (1) orthography of both the headword and inflected forms; (2) guidance on pronunciations; (3) word-class classification (part of speech); (4) definition writing; (5) examples; (6) phraseology; (7) disputed points of usage; and (8) etymology and word histories.

Brief History of Lexicography

Lexicography had its beginnings in the ancient civilizations of the Middle East (*see* **Bilingual Lexicography**) and ancient Chinese civilization (*see* **Chinese Lexicography**). In ancient Greece there were already dictionaries in the fifth century B.C. These were compiled to explain words used by Homer and other writers from the Archaic period, which had become rare or obsolete by the Classical period. Classical Greek lexicography was exclusively monolingual until quite a late period (*see* **Greek Lexicography, Classical**). It was left to the Romans to introduce bilingual lexicography (Latin → Greek, Greek → Latin), from the first century B.C. onward.

In many of the languages of Europe, the origins of lexicography can be traced back to interlinear glosses in medieval manuscripts (*see*, for example, **English Lexicography, German Lexicography**). Monks noted the vernacular equivalents of unfamiliar Latin words on the manuscript as an aid to understanding the text. In some monasteries, the glosses were subsequently collected and arranged roughly in alphabetical order in a separate manuscript. The development of European dictionaries as we understand them today, however, had to await the invention of printing technology in Germany by Johannes Gutenberg in 1450. Printed dictionaries became practical tools that could be mass-produced by the hundreds and thousands, enabling identical information about words to be disseminated extremely rapidly throughout a community. A major milestone in the development of European lexicography was the publication in Paris in 1531 of the *Dictionarium, seu Latinae linguae thesaurus* by the great Renaissance scholar and printer Robert Estienne (Robertus Stephanus). In 1538 he followed this with a Latin–French dictionary and in 1539 a French–Latin dictionary. These magnificent works remained unchallenged as authorities for over two hundred years and were the source of innumerable derivatives and revisions. Bilingual dictionaries played a major part in the revival of learning (Latin and Greek) during the Renaissance and an equally important part, first in the spread of Italian and French culture into northern Europe, and subsequently in the interplay of languages that was characteristic of Europe from the 15th century to the late 20th century, at which time, once again, an international lingua franca (English this time, not Latin) was to emerge. European dictionaries of the 15th and 16th centuries were of two kinds: those that played a role in the learning of Latin and those that were

designed to explain the words of one vernacular in terms of another. Many of these early printed dictionaries displayed great energy, learning, and sophistication, although an examination of them also provides a reminder of the important part played by the linguistic philosophy of the Enlightenment (17th and 18th centuries) in developing lexicographical standards of accuracy and consistency. Sixteenth-century dictionaries were full of information, but much of it was anecdotal and discursive. Balance and consistency were not among the lexicographical virtues of Renaissance dictionaries. Multilingual dictionaries were also popular during the Renaissance; valiant attempts were made to represent the meaning of words in one language by equivalents in several others. Only gradually did people come to realize how difficult it is to represent one language in terms of another. Languages do not map easily onto one another, and trying to map three or more languages onto each other is an impossibility, except in the case of superficial equivalents such as terms for artifacts in a technology that is shared by many linguistic cultures.

During the 18th century, large dictionaries were compiled in all the languages of Europe that were not the tongue of an oppressed minority. Academies were established with the aim of preserving the purity of vernacular languages (*see* **Academies: Dictionaries and Standards**). Great advances were made in data collection. (The aim was to collect and define all the words of a language, not just the 'hard words.') Scientific standards of definition writing were pursued. Some 18th-century lexicographers made strenuous efforts to implement the 'substitution principle,' a notion that can be traced back to Leibniz's dictum that two things are the same if the one can be substituted for the other without affecting the truth value (*salva veritate*). The idea was – and is – that a definition should be substitutable in any context for the word being defined (the *definiendum*). The substitution principle, coupled with unthinking reductionism (according to which every sense of every word is defined as an independent linguistic entity, rather than as part of a phraseological whole) is responsible for some of the convolutions still to be found in present-day dictionaries, for example in the definitions of English reciprocal verbs:

> **meet** (sense 2): to come into or be in conjunction or contact with (something or someone else or each other).

This definition may be convoluted, but at least it is an attempt to capture the reciprocal syntactic potential of the verb, as in examples 1–5 below. Some other English dictionaries present these as separate senses, while others ignore the syntax completely or indicate only 'transitive' and 'intransitive.'

1. John met Sally.
2. Sally met John.
3. John and Sally met.
4. John and Sally met each other.
5. John and Sally met with each other.

Eighteenth-century dictionaries saw their role not only as recorders of the words of a language, but also as reinforcers of 'correct' usage, to warn against barbarisms. Thus, Johnson (1755) included citations from "the best authorities" to show how English words are used, and included comments on words that he disapproved of (e.g., "**clever**: a low word"). During the 19th century, a shift from such prescriptive attitudes to scientific descriptive principles occurred in many important European lexicography projects.

The proposal by Sir William Jones in 1786 that Sanskrit, Greek, Latin, and other languages were "sprung from some common source which, perhaps, no longer exists" had a profound effect on scholarly lexicography. Although the American lexicographer Noah Webster (*see* **American Lexicography**) rejected the Indo-European hypothesis, preferring to believe that English words were derived from 'Chaldee,' scholarly lexicography in Europe throughout the 19th century devoted much energy to discovering the Indo-European roots of everyday words and establishing cognate relationships among words in different languages. Etymology, which in the 18th century had been a matter of wild guesswork, became a painstaking scientific discipline and came to dominate lexicography (*see* **Etymology**). The task of the scholarly lexicographer in all the leading languages of Europe was seen partially, if not primarily, as a historical one: not only to define the meaning of words but also to trace their origins. In 1864 Webster's successor Noah Porter and his etymologist C. A. F. Mahn quietly dropped 'Chaldee' and brought the leading American English dictionary into line with the rest of Western scholarship.

The establishment of lexicography as a scholarly discipline on historical principles led to the establishment of scholarly dictionaries of record for many languages. Also in the 19th and 20th centuries, large scholarly dictionaries of ancient and medieval languages began to be compiled, presenting a detailed account of the recorded vocabulary of major languages at earlier historical periods. Some of these great scholarly projects continued for many decades. The culmination of historical and in particular Indo-European lexicographical scholarship is still being worked out. In 1949 Carl Darling Buck published *A dictionary of selected synonyms in the principal Indo-European languages*. At the University of Leiden, a major project is currently underway with the goal of creating an Indo-European etymological database on the Internet and, eventually, compiling a new Indo-European etymological dictionary, designed to supersede Julius Pokorny's *Indogermanisches etymologisches Wörterbuch* (1959).

In the 20th century another school of lexicography grew up, the purpose of which was to explain the meaning and use of words in the contemporary language, relegating etymology and obsolete senses to a subsidiary role. Historical principles were superseded in these dictionaries by synchronic principles, more suitable for practical everyday works. The first work to make this distinction explicitly was Funk and Wagnall's *Standard dictionary of the English language* (1893). Some but not all American dictionaries followed Funk and Wagnall's lead, for example the *American College dictionary* (1949), its successor the *Random House dictionary* (1964), and the *American Heritage dictionary* (1969). Others, including the Merriam-Webster dictionaries, clung to the more traditional historical principles. Among British and Australian one-volume dictionaries, it is now standard to follow synchronic principles rather than historical principles. One-volume dictionaries of the English language (both historical and synchronic) characteristically focus on explaining meaning, with comparatively few examples. In other languages, for example German and Greek, the focus of one-volume dictionaries is often very much more on phraseology and usage, with many more examples of usage than are found in one-volume English dictionaries and comparatively less space devoted to definitions. In Russia, dictionary compilation was associated with important developments in linguistic theory (*see* **Russian Lexicography**), for example the meaning = text theory of Mel'čuk (*see* **Mel'čuk, Igor Aleksandrovic (b.1932)**).

Beginning in the late 1980s, lexicography began to respond to the opportunities offered by computer technology, in particular the use of computers to collect and process evidence of word use on a very large scale (*see* **Corpus Lexicography**).

Typology of Dictionaries

There are innumerable works calling themselves 'dictionaries' in almost every language in the world that has a literature. Indeed, modern lexicography also plays a major part in recording the vocabulary of and establishing written norms for languages that do **not** have a literature. These works are extremely heterogeneous: some of them have nothing in common other than the word 'dictionary' in the title. Others are clearly members of the same family of reference works. A typology of dictionaries may be attempted, identifying the following types: scholarly dictionaries of record, practical dictionaries for everyday use, pedagogical dictionaries, dictionaries of slang and/or idioms, special-subject dictionaries – all of which are monolingual and explanatory in nature – along with bilingual dictionaries, onomasiological dictionaries, and term banks. In addition, there are various less frequent classes, for example hybrid works such as monolingual dictionaries for second-language learners with marginal glosses in a relevant foreign language.

Scholarly Dictionaries of Record

The 19th and 20th centuries saw an explosion of scholarly lexicographical activity. Every language in the world that has an established literary tradition now has (or will soon have) a major dictionary of record, recording the vocabulary of the language and usually but not always organized on historical principles, with plentiful citations illustrating usage and supporting the definitions. Classic examples of great historical dictionaries are the *Oxford English dictionary* (OED) (discussed in **English Lexicography**), the *Trésor de la langue française* (*see* **French Lexicography**), and the *Deutsches Wörterbuch* of the brothers Grimm (*see* **German Lexicography**).

The present encyclopedia includes articles on the history and current state of the art of lexicography in all the world's major languages. There are also articles on major areas of historical lexicographical scholarship (*see* **Old English Dictionaries, Middle English Dictionaries**, and **Old French Dictionaries**). In other cases, historical dictionaries are surveyed in the relevant national articles (*see*, for example, **Dutch and Flemish Lexicography, German Lexicography**).

As a subgroup of the category of scholarly dictionaries, dictionaries of etymology and word history should also be mentioned. These range from popular works such as Chantrell's very readable *Oxford dictionary of word histories* (2002) and Picoche's *Dictionnaire étymologique du français* (1992) to larger works of scholarship such as Kluge's *Etymologisches Wörterbuch der deutschen Sprache* (1st edn., 1883; 24th edn., 2002), the Dutch *Etymologisch*