THE ART OF LEXICOGRAPHY

Niladri Sekhar Dash  
Linguistic Research Unit, Indian Statistical Institute, Kolkata, India

Keywords: Lexicology, linguistics, grammar, encyclopedia, normative, reference, history, etymology, learner’s dictionary, electronic dictionary, planning, data collection, lexical extraction, lexical item, lexical selection, typology, headword, spelling, pronunciation, etymology, morphology, meaning, illustration, example, citation

Contents

1. Introduction  
2. Definition  
3. The History of Lexicography  
4. Lexicography and Allied Fields  
4.1. Lexicology and Lexicography  
4.2. Linguistics and Lexicography  
4.3. Grammar and Lexicography  
4.4. Encyclopedia and lexicography  
5. Typological Classification of Dictionary  
5.1. General Dictionary  
5.2. Normative Dictionary  
5.3. Referential or Descriptive Dictionary  
5.4. Historical Dictionary  
5.5. Etymological Dictionary  
5.6. Dictionary of Loanwords  
5.7. Encyclopedic Dictionary  
5.8. Learner's Dictionary  
5.9. Monolingual Dictionary  
5.10. Special Dictionaries  
6. Electronic Dictionary  
7. Tasks for Dictionary Making  
7.1. Panning  
7.2. Data Collection  
7.3. Extraction of lexical items  
7.4. Selection of Lexical Items  
7.5. Mode of Lexical Selection  
8.1. Headwords  
8.2. Spelling  
8.3. Pronunciation  
8.4. Etymology  
8.5. Morphology and Grammar  
8.6. Meaning  
8.7. Illustrative Examples and Citations  
9. Conclusion  
Acknowledgements
null
[Pierre Bercheur, a 15th century French lexicographer] in his dictionary describeth it”. Latin dictio (also the source of English diction (15th century) was a derivative of the verb dicere ‘say’. Its original meaning was ‘point out’ rather than ‘utter’, as demonstrated by its derivative indicare (source of English indicate) and words in other languages, such as Greek deiknunai ‘show’, Sanskrit dic- ‘show’ (later ‘say’), and German zeihen “accuse”, which come from the same source (Ayto 1990: 170).

According to the Concise Oxford Dictionary for Current English (1998), a dictionary is a “book that deals with the individual words of a language (or certain specified class of them) so as to set forth their orthography, pronunciation, signification and use, their synonyms, derivation and history, or at least some of these facts; for convenience of reference the words are arranged in some stated order, now in most languages, alphabetical, and in larger dictionaries the information given in illustrated by quotations from literature”. Similarly, Berg gives the definition of a dictionary in the following way: “A dictionary is a systematically arranged list of socialized linguistic forms compiled from the speech-habits of a given speech community and commented on by the author in such a way that the qualified reader understands the meaning … of each separate form, and is informed of the relevant facts concerning the function of that form in its community” (Zgusta 1971: 197).

Although a dictionary serves various practical needs, it is usually consulted for the following reasons:

(a) A dictionary is used as a useful reference book for different types of linguistic information of various lexical items. The information is related to pronunciation, spelling, etymology, grammar, meaning, usage, etc. From this perspective, a dictionary performs the function of a ‘store house’ for a language or a variety.

(b) It is consulted as a guidebook for distinguishing good usages of words from bad usages. From this perspective, a dictionary performs the function of ‘court house’ (legislative) for a language or variety.

Besides the functions stated above, a dictionary is also a source of information on life, language, and society. In order to perform this function satisfactorily, a dictionary needs to obtain systematically both linguistic and extralinguistic information from as many sources as possible, since the information provided in a dictionary is always ‘tagged’ with invisible stickers of authenticity and reliability.

3. The History of Lexicography

Lexicography, the art and craft of dictionary making, has an important place in the history of language study. We find that dictionaries and glossaries of various types (monolingual and bilingual, general and technical etc.) were compiled and used extensively from the early age of civilization in China, India, Middle East, Greece, and Rome. The earliest known prototypes of dictionary were the West Asian bilingual word lists, dating from the Second Millennium BC. The word lists contained the Sumerian and Akkadian words inscribed in parallel columns on clay tablets in cuneiform writing. The contents were organized thematically, as thesauruses, for easy and quick reference.
After the invention of alphabet later in the same millennium, many centuries passed before an alphabetic ordering system became a common tool for organizing information. The lists came into existence because the Akkadians (Babylonians) had inherited through conquest the culture and traditions of the Sumerians and used the sets of signs as a means by which their scribes could learn what was, in effect, the classical language of writing. Over two thousand years later, in medieval Europe, the same principle was used when scribes who spoke vernacular languages learned to read and write in Latin.

The Chinese tradition of dictionary making is very old. The first known lexicographic work in China is *Shizhou*, which dates back to the 9th century BC. Unfortunately, this work did not survive. After a gap of centuries, during the Han Dynasty (206 BC - 220 AD), the art of lexicography was revived as a part of resurgence in literature. At that time, dictionaries such as *Shuōwén Jiēzì* and *Erya* were valuable reference works for understanding the ancient classics. In the later period, particularly during the Tang (618 AD-907 AD) and the Song (960 AD-1279 AD) dynasties, a few more dictionaries like *Yupian*, *Qieyun*, and *Guangyun* were compiled (Shiqi 1982).

An altogether different style of dictionary preparation flourished in ancient India. It started with the collection of obscure words as exemplified in the *Nighantus*, continued with formation of one of the bases of *Nirukta* and the *Padapatha* during the first millennium BC, and ended with a large group of *Kosas* composed during the past fifteen to seventeen hundred years after the Christian era began. Thus, a continuous activity of dictionary making in ancient India provided models for later lexicographical works in the Indo-Aryan and Dravidian languages in India (Katre 1980).

Both *Nighantu* and *Nirukta* are monumental examples of early Indian efforts in dictionary making. Since then a large number of dictionaries and glossaries are prepared in Sanskrit as well as in Pali, Prakrit, and Apabhramsha. Some of them are general purpose dictionaries like *Amarakosa*, while some others are partial ones like *Deshi Namamala*, and still others like *Dhatupatha* and *Ganapatha* fall under the category of technical glossaries. These dictionaries and glossaries follow different criteria and bases - semantic as well as alphabetic - in their structure and text representation.

In *Nighantu* all the designations used for a particular word in Vedic literature were included in the list. The ordering was known as *Paryay*, which was an essential part in the practice of dictionary making in ancient India. There were another two phases known as *Nanartha* that contained various meaning of words, and *Linga* that contained the gender information of words. Each *paryay* had dictionary of various types. For instance, the *Kalpadrumakos* and the *Abhidhana Chintamani* belonged to *Paryay* group, *Nanartha Shabdaratna*, *Nanartha Manjari*, *Anekartha Cintamani*, and *Nanartha Shabda* belonged to *Nanartha* group, while *Lingadi Sangraha* fell in *Linga* group.

Although there were well-defined schemes of classification of words according to their gender, meaning, and usage as well as the classified dictionaries, which would include words according to the predefined criteria, there were some dictionaries which included all three types of information of words within a single volume. The best examples of this kind is *Amarakosha* that included all thee three phases of words within three parts conglomerated within a single volume. Due to this fact, this dictionary is known as
Trikanda (there parts). According to the historical evidences, it was compiled between 400 AD and 600 AD.

The most notable aspect of Sanskrit dictionaries is that these were composed in poetic form and words were not arranged in alphabetical order. Among some of the working principles that had been adopted in the construction of lexicons in verse forms may be mentioned the grouping of words according to the number of syllables, and in a few cases, on the basis of the final syllable. Words are most often divided into various classes or groups based on certain features that forced the words to be included in particular groups. This norm of compiling dictionaries in Sanskrit as well as in other vernaculars was practiced not only in ancient India but also in medieval India for a long period of time. The Sanskrit model also provided the basis of similar works in some of the modern Indian languages, particularly for some Dravidian languages (Katré 1980).

Although there was a rich Indian tradition of lexicographical works, it was hardly ever adopted for modern Indian languages. The opening of the ‘Western windows’ (i.e. the advent of European scholars - missionaries as well as non-missionaries) helped the Indian scholars to adapt an altogether new method for dictionary compilation in modern India. In fact, the production of several bilingual and trilingual dictionaries in various Indian languages with direct involvement of the Western missionaries had a lasting impact on Indian scholars who applied the westernized lexicographical tradition to the Indian languages. The format of presentation, ordering of lexical entries, and the composition of dictionaries were more or less modeled after the dictionaries produced in the West.

The lexicographic tradition in Iran dates back to the pre-Islamic period. The oldest dictionaries are the Qim and Pahlavic dictionaries, both of which were compiled during the Sassanian Dynasty between the first and seventh centuries. The first dictionaries of modern Persian (or Farsi) were compiled in the 9th century due to the growth and spread of Persian literature (Nafisi 1999:180). The history of Persian lexicography is divided into three periods. In the first, prior to the 14th century, Iran and Central Asia were the main centers of dictionary making, and Persian dictionaries were compiled in Farsi-speaking regions. With the spread of Persian to neighboring countries and its acceptance as the language of the Royal court and literature in the subcontinent and Turkey, India became the major center of lexicographic activity. This second period lasted from the 14th century to the 19th century. Since the second half of the 19th century, there has been continuous growth of lexicographic work and research in Iran. Different kinds of dictionaries have been compiled and published, their main characteristic being a strong encyclopedic orientation. These efforts have resulted in the formation of hundreds of monolingual and bilingual dictionaries (Taherian 1998: 212-216).

In Europe, “the earliest ‘list of words’ that might be said to constitute the beginnings of English lexicography were the glossaries of Anglo-Saxon priests and schoolmen, compiled to enable those whose competence in Latin was lacking to read Latin manuscripts” (Jackson 1988: 111). These glossaries were essentially lists of Latin words with English glosses. The Promptorium Parvulorum (‘Storehouse [of words] for children’), published around 1440, was an English-Latin dictionary that contained around twelve thousand entries in alphabetical order with verbs and nouns listed
The first printed English-Latin lexicon is known as John Withals’ *Shorte Dickearie for Young Beginyners* (1553) which had a thematic arrangement of words. The term ‘dictionary’ in the title of this book was a sixteenth-century borrowing from the Latin *dictionarium* (‘the collection of words’). It is in these English-Latin dictionaries of the Renaissance period that we should perhaps recognize the beginnings of the lexicography of English (Jackson 1988: 112).

Sir Thomas Elyot compiled the first Latin-English Dictionary entitled *Bibliotheca Eliotae* in 1538. It was followed by a French-English dictionary, *Thesaurus Linguae Romanae et Britannicae* which was published in 1556 by Thomas Cooper. Other dictionaries followed, such as Richard Mulcaster’s *Elementarie* (1582), John Florio’s *A Worlde of Wordes* (1598), Randle Cotgrave’s *A Dictionarie of the French and English Tongues* (1611), and John Kinshieu’s *Ductor in Lingus… The Guide into the Tongues* (1617).

The first monolingual English dictionary is attributed to Robert Cawdrey, the author of *A Table Alphabeticall* (1604). It contained nearly 3,000 lexical items with short definitions. Other monolingual dictionaries followed, including John Bulloker’s *An English Expositor* (1616), Henry Cockerman’s *The English Dictionary* (1623), Thomas Blount’s *Glossographia* (1656), Edward Phillips’s *The New World of English Words* (1658), John Kersey’s *A New English Dictionary* (1702), and Nathan Bailey’s *Dictionarium Britannicum* (1730).

The credit for the first most comprehensive, reliable and exemplary dictionary in English always goes to Samuel Johnson’s *The Dictionary of The English Language* (1755). It was followed by a series of dictionaries published in English during the 18th and the 19th century. Some of the most notable examples were James Buchanan’s *Linguae Britannicae* (1757), William Johnston’s *Pronouncing and Spelling Dictionary* (1764), John Entick’s *Spelling Dictionary* (1764), and Thomas Sheridan’s *A General Dictionary of the English Language* (1780). This history of English lexicography took a completely new turn with the publication of the *Oxford English Dictionary* (1882) under the competent editorship of John Murray. Detailed information on the history and the diversity of English lexicography are available in McDavid and Duckert (1973), Hartmann (1983), Ilson (1986), Hausmann et al. (1989-91), Béjoint (1994), Béjoint (2000), Jackson (2002), Hartmann (2003), and others.

4. Lexicography and Allied Fields

4.1. Lexicology and Lexicography

Lexicography is closely related to lexicology. Both words are derived from Greek *lexikos*, an adjective derived from noun *lexis* meaning “speech” or “word”. While *lexicology* means ‘science of words’ (< lexicos ‘of word’ + logos ‘science’), *lexicography* means ‘writing of words’ (< lexicos ‘of words’ + graphè ‘writing’). Although both the fields are closely related due to their common concerns for words or lexical units, lexicography depends heavily on lexicology in many ways.

Words and their features are studied by both lexicology and lexicography. The sum total
of all the words available in a language forms the vocabulary or the lexical stock of that language. Although each word is an independent linguistic entity, it is indirectly related to other lexical items both paradigmatically and syntagmatically. While paradigmatic relation is based on the interdependence of words within the lexical system, syntagmatic relations show words in their patterns of arrangement. The vocabulary of a language is not an arbitrary frame of diversified phenomena; it is a well-defined system that consists of elements which, although independent, are interrelated in some ways or other.

As a phonological, grammatical and semantic unit, a word is made of by a particular group of sounds, and it has grammatical and semantic functions. Lexicology studies a word in all these aspects exploring the patterns of its phonological, morphological and contextual behavior as well as its semantic relationships. A word often undergoes changes in its form and meaning with respect to its origin resulting from its development and current usage. Since a word does not occur in isolation, its combinatorial possibilities are also studied in lexicology including its phrasal, idiomatic and proverbial functions.

The domain of lexicology is both diachronic (i.e. historical) and synchronic (i.e. descriptive). From the diachronic viewpoint it deals with the origin and development of the form and meaning of lexical units in a particular language across the time scale. From the synchronic perspective it studies various aspects of the vocabulary of a language at a particular point of time. This implies that in lexicology words cannot be studied in isolation without close reference to other fields.

From another point of view, lexicological study of words can be general and special. While general lexicology is concerned with general features of words common to all languages, special lexicology studies words with reference to a particular language. Furthermore, lexicological studies can be comparative and contrastive, based on the lexical systems of any two languages. Functionally, lexicology fulfills the needs of different branches of applied linguistics such as lexicography, stylistics, language teaching, etc.

Lexicography also studies lexicon but from a different angle. While lexicology concentrates on the general properties and features that can be viewed as systematic, lexicography typically deals with the individuality of each lexical unit (Zgusta 1973: 14). Lexicography is thus defined as the art of writing a dictionary or the science of compiling a dictionary. While lexicology studies words as elements of a system, lexicography approaches words as individual units with respect to their meaning and usage. We use a dictionary in order to learn about words in the process of language learning, comprehending a text in a better way or checking correct spellings and pronunciations of words, etc.

A word may have varied sets of characteristic feature, all of which may not be needed to a dictionary maker, since his work is mostly guided by the purpose of dictionary and the type of users. Words are presented in a dictionary in such a way that they can be accessed in real life situations. For instance, whatever theoretical basis for enumerating different meanings of polysemous words is accepted, meanings are arranged and presented keeping in mind the practical utility of dictionary of different users. While a
lexicologist presents materials according to his view of the study of vocabulary, a lexicographer is guided by the principle of convenience in the retrieval of data.

In principle, lexicology provides a theoretical basis to lexicography. A dictionary maker may know all the semantic details of a lexical unit, but he has to decide which details to include in definition. Lexicological study of words is governed by theories of semantics and word formation. Therefore, there is no scope for individual aberrations. In lexicography, on the other hand, definitions are often subjective and are not free from the bias of a dictionary maker (cf. meaning of oats in Johnson's Dictionary).

Lexicology is not language specific, since it deals with universal features of words. Lexicography is more or less language specific in spite of its universal theoretical background. Lexicography has no other relevance except for its practical applicability. Lexicology is more theory oriented, lexicography is more concrete in application of theories. In a certain sense “lexicography may be considered a superior discipline to lexicology, for results are more important than intentions and the value of theoretical principles must be estimated according to results” (Doroszewski 1973: 36).

Lexicology usually covers a wide range of interests and approaches to lexical study. It includes reconstruction of meaning and semantic change of words, lexical variation and change across time scale, evolution of vocabulary over centuries, neologism and word-loss within languages, lexical borrowing and derivation over time, structural and etymological analysis of lexical items, etc. with close interface between semantics, syntax, and pragmatics (Coleman and Kay 2000, Vera 2002). Lexicography, on the other hand, serves simply as a store house of information. In order perform this task adequately, it collects information from different sources and presents them within the scope of the dictionary users.

Bibliography

Arnold, I.V. (1973) The English Word. Moscow: Vysaya Skola. [It deals with form, nature, usage and meaning of English words]


Press. [A good book that presents interesting studies on the history of traditional lexicographic works along with special reference to the modern innovative ways adopted for dictionary compilation]

Bloomfield, L. (1933) Language. New York: Holt, Rinehart. [Perhaps, it needs no introduction. It deals will almost all the traditional areas of linguistics in descriptive point of view]


Doroszewski, W. (1973) Elements of Lexicology and Semiotics. Mouton: The Hague. [It fabricates a nice discussion to explore the interface between lexicology and semiotics]


Householder F.W. and S. Saporta (Eds.) (1962) Problems in Lexicography. Bloomington IN: Indiana University Research Center for Language and Semiotic Studies. [Another useful book that deals with the problems a dictionary maker often faces while trying to compile a dictionary – either general or special kind]

Ilson, R.F. (Ed.) (1986) Lexicography: An Emerging International Profession. Manchester: Manchester University Press. [It focuses on lexicography as a profession that opens up scope for survival and growth for the new generation of linguists. Also it provides some basic ideas about the techniques of dictionary compilation]


Jackson, H. (1988) Words and Their Meaning. London: Longman. [It is a book more in the nature of lexicology rather than lexicography. If explores the interface between form and meaning that often plagues lexicographers]


Landau, S.I. (2001) *Dictionaries: The Art and Craft of Lexicography*. (2nd Ed.) Cambridge: Cambridge University Press. [Perhaps, the best book I have ever read in the area of lexicography. In a captivating narrative technique, the author addresses all the major issues of dictionary compilation including with the use of electronic corpora in the work]


McDavid, R.I. and A.R. Duckert (Eds.) (1973) *Lexicography in English*. New York: Academy of Sciences. [It carries detailed descriptions about the history and present activities of lexicographical works in English language]


Singh, R.A. (1982) *An Introduction to Lexicography*. Mysore: Central Institute of Indian Languages. [Perhaps, it is the first book written in India to address all basic aspects of lexicography with reference to English and Indian languages]


Urdang, L. (1963) “Review of Problems in Lexicography”. *Language*. 39: 586-594. [It is one of the most insightful writings to address the problems and hurdles of dictionary compilation]


Wright, J. (1898-1905) *English Dialect Dictionary*. London: Oxford University Press. [One of the first dictionaries that analyses a large set of words collected from the British dialects of late 19th century]


Zgusta, L., et al. (1971) *Manual of Lexicography*. The Hague: Mouton and Company. [It is one of seminal works which is highly referred to in any discussion of lexicography. A must read for all lexicographers]

**Biographical Sketch**

**Niladri Sekhar Dash** has been working in the area of corpus linguistics and language technology for more than fifteen years at the *Indian Statistical Institute*, Kolkata. His first book “Corpus Linguistics and Language Technology: With Reference to Indian Languages” (2005: Mittal Publications, New Delhi) is widely acclaimed as one of the most exhaustive works in this area and used as a course and reference book in several universities and research institutes in India and abroad. He has two more books in this area published in Bengali which are appreciated as first works in Bengali. To his credit Dash has more than fifty research papers published in national and international journals. He has taught as a visiting
faculty at Madras University, Chennai (India), Jadavpur University, Kolkata (India), Punjabi University, Patiala (India), and North Bengal University, Darjeeling (India). He has acted as a Co-Investigator in the TDIL Project of the Ministry of Information Technology, Govt. of India, besides acting as an Expert in the ASI@IT&C project of European Commission, and in the Indo-African project of the International Scientific Research Network, Brazil. Dash specialises in the area of corpus linguistics, language technology, and lexicography.