

DICTIONARY TYPES AND THEIR MAIN PECULIARITIES

Komilova Tursunoy Rustamjon Qizi

Andijan state university first-year student of master's degree

Annotation: This article deals with the problems of lexicography and dictionaries. It also studies some special features of dictionaries and principles of compiling dictionaries.

Key words: lexicography, semantic, project, practical, theory, discipline.

Lexicography is the making of dictionaries. It is distinguished from lexicology as follows: The object of lexicology is the lexicon; its task is to investigate the structure of the lexicon. Lexicology is, thus, part of theoretical or descriptive (as opposed to applied) linguistics. The object of lexicography is the dictionary; its task is to make dictionaries. Lexicography is, thus, part of applied linguistics. Like other disciplines of applied science, lexicography could also be called a practical enterprise if it be understood that it is not purely practical (like, the writing of newspaper articles), but theoretically based, viz. based on lexicology.

There is also a discipline of linguistics called meta lexicography (or theory of lexicography). Its object is lexicography. While lexicography makes dictionaries, meta lexicography describes and prescribes how lexicography proceeds. It includes the history of lexicography. In terms of system linguistics, the lexicon of a language is an ideal entity that comprises the entire significant system of a language, under any aspect, to the extent that it is idiosyncratic rather than regular. The same would be true for an ideal dictionary, a perceptible representation of an ideal lexicon. Real dictionaries are limited in their purpose and scope. Therefore, a dictionary is a representation of part of the lexicon of a language for a certain purpose. The following aspects are relevant here:

Given that a language system is an ideal object, no description can fully represent it. No dictionary has ever fully represented the lexicon of a language. (The same is, of course, true for a grammar.) The purpose of the dictionary may involve the representation of the lexica of more than one language. Dictionaries have commonly been printed.

A dictionary contains linguistic knowledge, not world knowledge. The latter is the task of an encyclopedia. Nevertheless, delimitation between a dictionary and an encyclopedia is often problematic, as much of our world knowledge is coded in our language system. Most every kind of information in the world can be organized in the form of a (mostly alphabetic) list of items under which the information is grouped. There are lexica of mythology, of ancient philosophers and of French cuisine. All of these are encyclopedias rather than dictionaries.

The Neo-Latin word *dictionary* is a calque on the Greek *lexikón*. Just as the latter is derived from the noun *léxis* 'expression, saying, word', so the former is derived, by the suffix *-arius* 'repository of', from *dictio* 'expression, saying, word'. However, nowadays the words *lexicon* and *dictionary* are no longer synonymous, as explained above.

1. Levels of structure

The present section focuses on a general dictionary.

A dictionary is structured at three levels below the level of the dictionary as a whole:

The dictionary as a whole has a framing structure which comprises a set of main sections that correspond to the chapters of a book. A subset of these sections – mostly only one – comprises an entry list. The structure of each of these sections is a macrostructure (or the macrostructure of the dictionary). Each of the entries in a macrostructure has an internal structure of its own, which is the microstructure of the entries of that list.

The article of Mann and Schierholz published in *Lexicographica* 30 gives an overview in German and is a first attempt to structure the existing methods in our discipline [3. 88]. The other articles which are published in the *Lexicographica* yearbook treat methods in specific areas: the functional perspective, dictionary planning of online-

dictionaries, research in dictionary use, the saving of lexical data, translation science, specialized lexicography, pronunciation dictionaries, items giving the paraphrase of meaning Even this article cannot give a complete overview of all the methods but it provides a general overview of methods in lexicography and dictionary research, as well as in contiguous disciplines. Selected methods will be described in detail to show how to apply the methods [4. 134].

The basic aim of this article is to foster the work on methods and to encourage young researchers to show more interest in the methods of this research field. An overview of all methods should be theory independent or should be orientated to the different lexicographical theories, i.e. the "Lexicography and Dictionary Research" or the "Function Theory". Above this, all methods which are used in lexicographical practice should be enumerated and ordered.

This is rather complicated because one has to keep in consideration that some methods will be applied only with a special dictionary type or are only specific to one dictionary project. Besides this aspect regarding methods, it has to be noted that the individual theories are not critically reviewed with respect to existing methods, in "Lexicography and Dictionary Research", in the "Function Theory"¹, and in "Lexicography as an Art and Craft".

Exceptions are (a) the *Historical Dictionary Research* and (b) the *Research on Dictionary Use* where in the past the state of the art of the empirical methods was summarized by Ripfel and Wiegand, in Wiegand, in Welker. For the period from 1993 to 2012 a very good overview of the research projects concerning the use of electronic dictionaries including a lot of critical comments on the used methods is given by Töpel (2014b). Apart from these observations regarding the documentation of methods, one can find methods reflected more or less in the publications of lexicography. These are the publications where

- a) the subject matter itself is methodology,
- b) single methods are discussed critically,
- c) specific methods are used aligned to the subject matter,
- d) methods are applied without any reflection on their use but also without any consequences for the quality of the research results [3. 95].

The *linguistic theory* plays a decisive role since, not only in language dictionaries but also in subject dictionaries, linguistic expressions are described in/ with linguistic terms.

All linguistic procedures play a dominant role in lexicography and dictionary research. If such linguistic data are collected it shall be the basis for the writing of lexicographical items [4. 79]. This means that one has to consider;

- a) the procedures of phonetic transcription, procedures of segmentation and classification, many procedures in phonology and morphology, procedures to categorize the parts of speech, syntactical procedures for the categorization of valence partners, procedures of meaning analysis, content analysis, prototype theory and many more;

- b) the methods of observation and questioning, statistical procedures for the collection of linguistic data, for the evaluation of huge and small datasets, for samples as well as for the selection from huge data sets (e.g. textcorpus data);
- c) the knowledge the lexicographer has from experience related to their profession.

The methods of corpus linguistics must be used with special attention because linguistic data is collected on the basis of huge text corpora. The required size of the corpus, the balance of the corpus, the state of the corpus (original text or scanned text; with annotation or current text), the text varieties which exist in the corpus; but first the query procedures and the procedures of the citation evaluation must be taken into consideration and must be used with a critical reflection and perspective to the investigation aims [1. 97].

It must be distinguished in the lexicographical process between corpus controlled, corpus supported or corpus validated. Attention must be paid to the use of linguistic methods in that they serve very often only for the material processing and material preparation for the following lexicographical process. In each case it must be decided in which way data extracted from natural language sources have to be further worked on in the lexicographical process with linguistic methods.

Language planning has been defined as an organised activity to regulate language issues in a society. Such language issues may include "the creation of new alphabets, the codification of morphology, standardization, the development of 'plain language', spelling reform, language maintenance, and the

elimination of gender discrimination in language".

A common model used to describe the activities of language planners has been proposed by Haugen [2. 77]. *Status planning*, which is usually the agenda of politicians and bureaucrats, involves developing a national identity and language spread at national and international levels. *Corpus planning*, on the other hand, is usually the agenda of linguists, lexicographers and experts alike who intend "... (i) to give the language a terminology for scientific and technical purposes; (ii) to resolve normative/structural questions of correctness, efficiency, and stylistic levels; and/or (iii) to support an ideological cause by eliminating sexist, racist, or militaristic elements in the language" Selection refers to the decision of replacing one language with another or one variety with another when there are two conflicting norms. Examples of this can be seen in the replacement of Arabic with Hebrew in Israel, and the urban with rural dialects in Norway [4. 51].

Selection and codification will be successful only if they are followed by *implementation* and *elaboration*. *Implementation* usually takes the form of books, pamphlets, newspapers and textbooks. *Elaboration*, on the other hand, is "the continued implementation of a norm to meet the functions of a modern world" [11. 43]. The language form which has been selected and codified may be spread by individuals (e.g. a writer), a government agency or an institution. Status planning and corpus planning went hand in hand in Turkey, especially in the early years of language planning [3. 109].

Nationalism was "the central pillar of Kemalist ideology" and created a strong "demand for the purification of the Ottoman language by replacing its foreign elements with genuine Turkish words...". The selection of Turkish as the norm, and the recodification of the language using the Latin alphabet announced the start of an official language policy in 1928 [3. 99].

The implementation and elaboration would be carried out by a semi-official agency, the Turkish Linguistic Society, which was founded in 1932. TDK's activities included the elimination of Arabic and Persian elements in the language and the creation of a new Turkish lexicon based on a number of processes: a) collecting words from old Turkish texts (*tarama*) b) collecting words from folk vernacular (*derleme*) c) deriving new words using Turkish roots and derivational morphemes (*tUretim*) c) compounding (*birlestirme*) [1. 133].

After 1935 "many long lists and entire books full of proposed new vocables for all the walks of life were published" (Tietze 1962) as well as a monolingual dictionary in 1944, which has been an influential agent in establishing the new 'norm'. While Turkish language planning is commonly labelled as a successful movement, it did not enjoy a smooth transition. The outline of the microstructure of a dictionary presented below obeys the following principles:

It is oriented towards the description of a language on the background of another language. We are, thus, in principle speaking of a bilingual dictionary, although there may be more than one relevant background language in the working context. It attends the needs of a fieldworker who describes a language he does not master himself. It provides a maximum model which will not be fully needed in every single case.

It is, however, easier to leave cells provided by the structure blank than to expand the structure after one has started with too simple a model. Likewise, it is simpler to secondarily conflate the contents of two fields in one than to split up the content of a field into two. The microstructure is explained with a (relational) lexical database in mind. The complete set of fields, with their numbers, names, explanation, examples and linking relations to other parts of the overall linguistic description, is listed in a table on a separate page. Each entry of a dictionary must be identifiable, both by references from within the dictionary and by references from outside (e.g. the accompanying grammar). If the lexical database is a relational database, then each entry will have an ID in the technical sense. However, since that is a user-unfriendly number, it is normally not suitable for human users. Instead, for the identification of an entry, these rely on the following pieces of information:

1. Lemma

The lemma is given in standard orthographic representation. Sometimes, the lemma representation itself is used for additional purposes, e.g. to indicate stress, syllable boundaries or word-break points for hyphenation. However, the present comprehensive microstructure provides dedicated fields for such purposes.

If the lemma is a segmental sign, but not an independent word form, it is flanked by a hyphen (or two hyphens) at the side where it is bound, like English *cran-* and *-ize*. If it is a morphonological process or a suprasegmental sign, such as German megaphone marking plural or Yucatec Maya high tone marking

deagentive formation, a suitable notation has to be developed. Such conventions are explained in the appropriate main section of the dictionary.

2. Homonym number

Homonyms are, of course, separate entries distinguished by numbers; see the separate section for details. The same goes for the readings of a polysemous entry; see next item.

3. Sense number

As shown in the section on lexical relations, a relational lexical database provides an elegant way of both keeping the distinction between homonymy and polysemy flexible and of providing its own partial microstructure for each of the readings of a polysemous item. These readings are numbered from 0 (mother entry) to n.

4. Citation form

Normally, the lemma should be given in the citation form that is traditional in the speech community. In that case, the content of this field is identical to the content of field 1. However, the dictionary also contains lemmas – especially dependent forms (roots, stems, affixes ...) that would not be cited in their naked form by non-linguists. In such cases, the field ‘citation form’ must be filled in. For instance, Yucatec Maya verb roots are never cited as such, but with a default stem extension, for instance: lemma kul- (‘sit’), citation form: kutsal. This field is also the reference point for field6 below.

B. Expression

In every natural language that is at all represented in a dictionary, an expression has at least two significantia, a phonological and a graphemic one.

5. Phonological representation

The phonological representation is given in IPA. If the citation form differs from the lemma, this field may refer to either of them. In practice, filling in this field may be limited to such cases where the phonological representation is not derivable by rule from the orthographic representation.

6. Sound

This is a link to a sound file. There are at least two possibilities here:

There is a separate sound file for each lemma. This presupposes that there is a representation – a citation form – of the lemma that may be pronounced naturally in isolation. Consequently, what is pronounced here is the content of field. If there is no such sound file for each lemma, this field may contain one or more pointers to text recordings. Technically, such a pointer identifies a sound file and specifies a start point and an end point contained in that file, in milliseconds.

7. Phonological variants

This field accounts for idiosyncratic phonological variation. For instance, the first phoneme of economic may be /ε/ or /i/.

Other variants are derivable by phonological rule. For instance, the rule of syncope in German predicts that if there is a lemma Wanderer, there will be a variant Wandrer. This rule is stated in the phonology section of the dictionary and then renders superfluous the enumeration of all the variants it generates.

8. Orthographic variants

The standard graphemic representation already serves as lemma. This field is needed for alternative spellings found in the corpus. For example, the lemma encyclopedia contains encyclopedia in this field. The lesser the degree to which the language is standardized, the more variation there is in available texts (e.g. of earlier periods); and the fewer dictionaries there are for the language, the more important does it become to display the variants in this field. Additionally, such orthographic variants which are based on regular phonological variants do not need to be noted individually.

C. Language variety

The language that constitutes the object of the description including the dictionary determines the corpus set up for the research. It must be defined in the introduction to the dictionary. The definition will necessarily, and explicitly, excludes certain kinds of linguistic variation (cf. the section on variation). For instance, diachronic variation may be limited to one of the stages traditionally defined for the language. All the variation not explicitly excluded will be represented in the corpus and have to be categorized. This information is also called diasystematic marking.

The possible contents of the following fields may be defined as range sets.

9. Dialect

Assuming that the lexicon is not confined to one dialect, dialect appurtenance of the lemma is indicated here. One possible value of this field is 'standard'.

10. Sociolect

Again, this will usually be marked only if the lexical item is special. Relevant values include particular age groups or professions.

11. Style

This concerns style, register, connotations and any kind of pragmatic information. The same restrictions as in the previous fields apply. Relevant values include 'ritual', 'formal', 'vulgar'.

12. Stage

A lexicon is usually confined to one stage of a language. Other stages may come in two ways:

If the text corpus includes older texts, it may feature obsolete items.

Because of the presence of diachronic in synchrony, some elements of the inventory of a given state are archaic, others are current, others are fashionable. This section of the microstructure concerns both the internal structure of the stem representing the lexeme – its inflectional and derivational morphology – and its distribution in syntax and phraseology. The concepts used are those introduced systematically in the grammar coupled with the lexicon; and they will appear in print in the grammar section of the published dictionary. Whenever such a term appears in a lexical entry, reference to that grammar (section) is implied.

13. Proper name

What is meant here is the proper name of a grammatical morpheme. For instance, the proper name of the English suffix *-ize* is 'verbalizer', and the proper name of English's *'s* is 'genitive'. Consequently, the possible contents of this field are unique (i.e. there is no range set), and only a portion of the entries of the lexical database will be specified for this field, viz. the grammatical formatives.

14. Syntactic category

From among the grammatical categories of a lexical item, this field is dedicated to its syntactic category qua distributional category (for morphological categories see #18). This is understood as a narrow subcategory of a part of speech, e.g. 'proper noun', 'transitive verb with additional prepositional complement'. The taxonomy implied here will be explained in the grammar. A lemma which belongs to diverse syntactic categories is considered polysemous. Each category then constitutes a record.

15. Morphological structure

This field contains the immediate constituents of the lemma stem; as long as binarism obtains, there are two of them. In the case of a compound, they are two stems; in the case of a derivative, they are a stem and some derivational operator which may or may not be segmental. The items listed there are identical to certain lemmas of the database.

Database solution

Relational set up a cross-table for derivational relations: column 1: ID of complex stem; column 2: ID of first constituent; column 3: ID of second constituent

Free field structure hyperlinks from the constituents to their target records

16. Word formation

This field contains the technical term for the word-formation process that formed the lemma stem, e.g. bahuvrihi, causative, denominal, deverbal, intensive etc. Possible entries in this field are taken from a range set defined in the grammar, where the word-formation processes of the language are dealt with systematically. In this field, the last word formation process applied is indicated, i.e. the process which was applied to the components of field to form the stem of the lemma. In the case of a derivationally complex lemma, other word formation processes may have created stems that are part of it, in particular those of field. Such processes are not indicated here, since they may be seen by following the links of the latter field.

17. Derivatives

In this field, the set of lemmas is referenced which have the current lemma in their field. Thus, references between the current field and that field are mutual. Database solution relational the cross-table already mentioned for field 15 provides the content both for that field and for the present field free field structure

series of hyperlinks, to each of the derivative lemmas. An alternative would be not to have entries for productively derived stems in the dictionary. Then the present field would contain the derivation schemata which are applicable to the lemma stem.

18. Morphological categories

The nature of the inflectional categories to be specified here depends on the language. Examples are noun class, gender, possessive class, verbal voice, inflection class. An inflecting word of a language may fall into diverse morphological categories at once, e.g. voice x, conjugation class. Some may be syntactically relevant lexical classes such as the gender of a noun; others may be purely morphological classes such as inflection classes such as inflection classes. It is practical to set up a separate field for each of these categories.

19. Irregular inflection

If the stem has inflected forms not derivable by rules pertaining to its inflection class, those forms are listed here. They may be both stem allomorphs such as worse, appearing in this field of the lemma bad and irregular forms of the inflection paradigm, e.g. oxen appearing in this field of the lemma ox.

20. Construction

This field contains the syntactic and semantic construction frame (for a verb: its valiancy frame), including selection restrictions. A case in point is different constructions of complement for complement-taking verbs. This is a specification of the information contained in. It should be represented by a formal notation, e.g. [\sim X]Y, where \sim indicates the position of the lemma, X represents relevant syntactic constituents or properties of the context, and Y is the syntactic category of the construction.

21. Phraseology

This field lists collocations in which the lemma is involved. These may be any kind of fixed expressions, including phrases, idioms and proverbs. If one has decided to bestow lemma status to such complex expressions, then this field contains links to such lemmas.

22. Meaning definitions

Semantic information on the lemma is provided in different languages and from different points of view. The sum of the information contained in this subset of fields is highly redundant. It is, however, useful for different kinds of dictionaries to be output from the database. The basis of the methodology of dictionary definitions is the logic and methodology of the definition in general. See the website devoted to this topic. For the present purpose, the meaning is specified in plain prose. In specifying properties of an argument of a relational lexeme, care must be taken to distinguish selection restrictions for a possible argument from semantic features of the lexeme itself. For instance:

anziehen (a) put on [clothes]

sichanziehen (b) put on clothes

In the German example a), '[clothes]' represents a selection restriction concerning the direct object of a transitive verb. In example b) instead, 'clothes' represents part of the meaning of an intransitive verb. If the brackets were missing in a), it would not be clear that 'clothes' does represent a selection restriction on a direct object.

Native definition

This field contains a specification of the meaning of the lemma in plain prose of the same language, as it would be the case in a monolingual dictionary.

User language definitions

The following fields specify the meaning in a set of languages that may be relevant in the working context. These may include the following languages:

English, because that is the general lingua franca in which either the entire dictionary or, at least, extracts from it may be published; the regional lingua franca, i.e. the language in which the linguist begins his fieldwork and in which the dictionary may be published, too; the native language of the linguist, because that is the language he fully controls. To the second of these fields may be added another one, called 'native translation', which contains literally the explanation that the informant gave. This differs from field #22.2, which generally contains the lexicographer's definition. Since polysemous lexemes are split up over so many database records (see the relevant discussion), each lemma has only one sense or meaning. There is, thus, no necessity to provide for a special substructure of these fields.

Whenever a token of the lemma stem appears in a text provided by an interlinear morphological gloss, the gloss is the same for all tokens of that type. This is achieved by retrieving it from the lexicon.¹ In principle, the gloss is provided in the same user languages as before. However, since only linguists are interested in interlinear glosses, an English gloss may suffice. Care must be taken concerning the relation of the gloss to the lexical item. An item of an interlinear gloss corresponds to an item in the text that is taken as a whole, i.e. not analyzed morphologically. That is, minimally, a morph, whenever the lemma of a lexical entry consists of a morpheme, no problem arises for the gloss field. The same holds, in principle, if the lemma is a complex stem. Problems do arise if the lemma is an inflected (citation) form, because then it will contain inflectional morphemes beside the stem. The gloss, however, must render the stem.

In conclusion we are going to say that we have investigated the problems of dictionary and its types. We have also studied the peculiarities of dictionaries in Modern English and Uzbek.

REFERENCES:

1. Barantsev C.T. English Lexicology in Practice, Kiev, 1954 p.43, 87, 138.
2. Ginzburg R. A. Course in Modern English Lexicography .Moscow, 1979. P.98-99.
3. Marchand H. Studies in modern lexicography. Munich, 1998. P.148-153.
4. Malkiel J. Genetic "Current trends in linguistics; vol. M.,1988. P. 28.
5. Marchand H. The category and types of present-day English dictionaries. Wresbaden, 1960. 18-24p.