E . V . K u p r i j a n o v

National Technical University Kharkiv Polytechnic Institute

Electronic dictionary classification
as problem of modern computer lexicography

Купріянов Є. В. Класифікація електронних словників як проблема сучасної комп'ютерної лексикографії. Стаття присвячена проблемі розробляння типології електронних словників на сучасному етапі розвитку комп'ютерної лексикографії. Уточнено поняття *комп'ютерний словник*, виокремлено його головні ознаки, що необхідно враховувати під час побудови класифікації. Розглянуто існуючі класифікації електронних словників, запропоновані вітчизняними й зарубіжними науковцями, та виокремлено їх недоліки. Запропоновано власну класифікацію комп'ютерних словників та параметри, за якими вона має відбуватися.
Ключові слова: електронний словник, комп'ютерна лексикографія, текст, гіпертекст, типологія словників.

Куприянов Е. В. Классификация электронных словарей как проблема современной компьютерной лексикографии. Статья посвящена проблеме разработки типологии электронных словарей на современном этапе развитее компьютерной лексикографии. Уточнено понятие *компьютерный словарь*, выделены его основные признаки, которые необходимо учитывать при создании классификации. Рассмотрены классификации, предложенные отечественными и зарубежными учеными, а также выделены их недостатки. Предложена собственная классификация компьютерных словарей и параметры, по которым она должна осуществляться.
Ключевые слова: электронный словарь, компьютерная лексикография, текст, гипертекст, типология словарей.

Kuprijanov E. V. Electronic dictionary classification as problem of modern computer lexicography. The present article is devoted to electronic dictionary typology classification with regard to the recent developments in modern computer lexicography. The notion *electronic dictionary* is defined and its main signs to be taken in account for elaboration of dictionary classification. The classifications of national and foreign researchers are considered and their main disadvantages are revealed. The author's own classification of electronic dictionary is offered and parameters to be used while carrying out this classification.
Key words: electronic dictionary, computer lexicography, text, hypertext, dictionary typology.

The active development of electronic dictionaries and their broad application in research and practical activities require conducting theoretical lexicography studies. In this context, a most topical issue is building classification of electronic dictionaries and defining criteria by which they should be classified. There have been attempts to resolve this issue in several works by Ukrainian, Russian and foreign researchers, namely I. Zavaruyeva [1], Yu. Marchuk [3], S. Merkulova [4], R. Mysak [5], L. Nelubin [6], V. Perebyinis [8], N. Syvakova [9], V. Chernytsky [10], Carolin Müller-Spitzer [13], Gilles-Maurice de Schryver [12], Verónica Pastor [15] et al. However the electronic dictionary in classifications proposed is considered as a software product, not from lexicographic point of view.

The goal of our research is to propose lexicographic approach to electronic dictionary classification. To achieve this goal it is necessary to: 1) define the notion *electronic dictionary* and mark out its signs to be taken into account when building classification; 2) propose criteria necessary to distinguish different types of electronic dictionaries; and 3) elaborate our typology of electronic dictionaries on the basis of the criteria proposed.

There is no common understanding among the researchers of what electronic dictionary is. For example, Ye. Karpilovska and V. Perebyinis interpret this term as "a dictionary compiled by computer" [2, 8]. But this definition isn't correct since the computers are used to create not only electronic dictionaries but paper dictionaries too. In our opinion, though *computer* as a creation tool and working environment is a main sign of the notion analyzed, it cannot be regarded as determinative one. A rather broad definition is given by I. Zavarueva: "electronic dictionary is a computer database of the entries specifically coded to facilitate quick search of the words with regard to morphological form and with the possibility of searching word combinations (word usage) and changing the direction of translation (for example Ukrainian-Russian or Russian-Ukrainian)" [1]. The given definition marks out the following signs of the notion: *database, specifically coded entries* and *quick search*. Unfortunately this definition is not correct either because it is also applicable to the term *computer version* of a paper dictionary interpreted as

"a form of computer presentation of existing paper dictionaries and, therefore, this is nothing more but the lexicographic material transformed from paper into electronic form using computer tools" [8:54]. The same can be said about H. Nesi's definition: The term electronic dictionary (or ED) can be used to refer to any reference material stored in electronic form that gives information about spelling, meaning, or use of words [14:140].

A special consideration deserves the idea expressed by Ya. Pervanov electronic dictionary is "a new structured text having definite volume, aim and bearing definite idea" [7:54]. As C. Muller-Spitzer notes, in the context of electronic dictionaries, it is worth noting that "text" includes data represented in different media: as text, audio files, videos, graph-based views, etc. [13:2]. The signs considered by the researchers to be most characteristic to the electronic dictionaries and non-attributable to paper dictionaries are the following: 1) combination of text and hyper-text form of material representation [1, 11, 13]; 2) availability of verbal and non-verbal means of lexical unit description [13]; 3) search facilities: within dictionary itself (in entry, wordlist) and in other resources posted on the Internet [1, 15]. As for the latter it is worth noting that the electronic dictionaries are not isolated objects, merging with other linguistic resources (other online dictionaries) and thus forming the global lexicographic space. Due to this fact the scope of the data that can be obtained are not limited to a book volume or even to a library, it covers the global information resources, reducing considerably time for search and access to the information required.

Thus *electronic dictionary* is a special lexicographic characterized by non-linear textual structure (the scope of which depends on a user's queries), inside and outside search, harmonic combination of different types of information (phonetic, semantic, encyclopedic etc.) in one entry, verbal and non-verbal means of information representation and possibility of connecting with other information resources. From our point of view this definition reflects the main signs of the notion and they should be considered fundamental while classifying electronic dictionaries. Before we propose our dictionary typology we find necessary to survey existing classifications set forth in the works of key lexicographers.

For example, I. Zavaruyeva and Gilles Maurice apply in their electronic dictionary typology such parameters as: form (*online dictionaries* located in the Internet and *dictionaries in electronic form* distributed on *CD*), information arrangement (*textual* and *hypertextual* dictionaries) [1, 12]. Hypertextual dictionaries can be *creolized* (containing extra-linguistic elements such as pictures, audio and video) and *non-creolized dictionaries;* dichotomy "paper dictionary — electronic dictionary" (*based on a paper dictionary* and *newly developed*) [1].

According to V. Chernytsky, electronic dictionaries must be classified by the parameters "operational system" and "loading mode". Thus they are divided into: *those designed for MS DOS* and *those for Win-*

dows (multifunctional dictionaries); *non-residential* (with their own shell program) and *residential* (called from other applications, e.g. text processors) [10].

A word list arrangement is considered to be determinative by N. Sivakova for distinguishing different types of electronic dictionaries: frequency-ordered, alphabetically ordered, thesauruses, thematically grouped, concordances, special purpose dictionaries (meant for specific tasks, i.e. semantic synthesis), combined (arranged by several parameters, for example frequency-alphabetically ordered) dictionaries [9].

R. Mysak and Lehr use information medium and devices as parameters for their classification. Electronic dictionaries are divided into: "1) *computer dictionaries* (those set up on desktop computers or notebooks); 2) *pocket dictionaries* (recorded in pocket electronic devices etc.); 3) *mobile dictionaries* (used in mobile telephones). *Computer dictionaries* are subdivided into: 1) *stationary* (installed on computer hard disk); 2) *portable* (distributed on CDs); 3) *online dictionaries* (available and accessible in the Internet)." The combinations of these types are possible [5:52–53].

The above listed classifications have disadvantages, namely: 1) making great focus on technical aspects of electronic dictionaries and 2) ignoring parameters important for user such as vocabulary, lexicographic arrangement of material on macrostructure and microstructure elements and its representation form. Our idea is that the electronic dictionaries (not machine-oriented dictionaries) can be classified in the same way as paper dictionaries: encyclopedic (*Encyclopedia Britannica*, *Електронна енциклопедія українського козацтва, Энциклопедия Кругосвет*) and linguistic dictionaries (*Oxford English Dictionary, Інтегрована лексикографічна система «Словники України», Электронный словарь Ожегова*). Linguistic dictionaries can be monolingual (*СЛОВНИК.НЕТ, Толковый словарь Владимира Даля он-лайн*) and bi-, multilingual (*Пролинг УЛИС Online, ABBYY Lingvo*). To linguistic dictionaries can be attributed those including vocabulary/terminology of different areas (for example, *ABBYY Lingvo containing biology, medicine, machine building, engineering, building and other technical terms*) and those covering sing area (*Webopedia, Російсько-український коксохімічний словник, Англо-русский специализированный словарь нефтяной промышленности ECTACO*). As a result of recent developments in computer lexicography it has become possible to combine linguistic and encyclopedic information of lexical units in electronic dictionary. (*Електронний глосарій фітонімів, Словник гідротурбінних термінів Turbolex* та *Комплексний словник-довідник із фразеологізмів Луганської області*). As for word list arrangement, linguistic dictionaries can be alphabetically-ordered (like in English-Ukrainian dictionary «*Глоса*») and thesauri (*Тезаурус з комп'ютерної лексикографії, тезаурус з лінгвістичної термінології, GEMET Thesauri*). Regarding language varieties, electronic dictionaries fall into normative dictionaries, or liter-

ary language dictionaries (*Інтегрована лексикографічна система «Словники України», Толковый словарь Ожегова он-лайн*), regional dialect dictionaries (*Українські говірки Донеччини*) and social-group dialect dictionaries (*словник сленгу хіпі, словник злодійського жаргону* etc., available at slovnik.com.ua).

In some cases computer dictionaries are impossible to be categorized as they can combine at the same time the elements of different dictionaries. For example the electronic glossary of phytonym terms compiled by N. Sivakova includes the elements of special-purpose (phytonym term system is described), explanatory (a definition to each phytonym are given), translation (English and Russian equivalents are listed), etymologic (the reference about phytonym terms origin is provided), encyclopedic (the lifetime, geographical range and pictures of plants are indicated) and codifying (a great number of equivalents to Latin terms are codified in Russian and English languages) dictionaries. Another example of a complex dictionary is Thesaurus on Ecology (compiled by M. Koviazina) which combines translation dictionary and thesaurus.

The *Semantic dictionary of Ukrainian Language* (*Семантичний словник української мови*) represents a special type of electronic dictionary. Its title is rather relative because it contains not only linguistic information about word (hyper-hyponym characteristics) but also encyclopedic information about an object denoted by the word (geographical range and usage of a plant). The dictionary also shows how the object is reflected in Ukrainian culture (for example, a *dog-rose* as a symbol of love, beauty and ornament of tableware), folklore (for example, a *periwinkle* in a song "Halia carries the water") and everyday life (a *periwinkle as name of cake*).

Thus, the electronic dictionaries are proposed to be classified both by conventional parameters: 1) vocabulary (*general-purpose* or *special-purpose dictionaries*), 2) number of languages (*monolingual, bilingual and multilingual dictionaries*), 3) destination (*translation, explanatory dictionaries* etc. or *complex dictionaries*); 4) adherence to language norms: (*dictionaries of literary or spoken language*); and by criteria peculiar only to computer lexicography: 1) linguistic (*textual* and *hypertextual dictionaries*, with hypertext linking the entries and outer language resources, such as Wikipedia, Lingvo.pro etc.); 2) dichotomy "paper dictionary — electronic dictionary" (*based on a paper dictionary* and *newly developed*); 3) availability of terms used in one or several areas in case of terminology dictionaries (*dictionaries containing terms to be used in a single or several areas*); 4) information form: *textual dictionaries, audio dictionaries and video dictionaries*. The example of audio dictionary is a dictionary of *Ukrainian dialects of Donetsk region* and the example of a video dictionary is *online encyclopedia of distinguished people* the video materials of which are devoted to outstanding painters, scientists, military leaders and politicians.

Literature

1. Заваруева И. И. Об одной возможной классификации электронных словарей / И. И. Заваруева // Вісн. Харк. нац. ун-ту імені В. Н. Каразіна. Сер. : Філологія. — 2007. — № 765, вип. 50. — С. 67—70.

2. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. / Є. А. Карпіловська, — Донецьк: ТОВ «Юго-Восток, Лтд», 2006. — 188 с.

3. Марчук Ю. Н. Вычислительная лексикография / Ю. Н. Марчук. — М., 1973. — 183. с.

4. Меркулова С. В. Автоматизированные толковые словари и контекстное определение : дисс... канд. филол. наук : спец. 10.02.21 «Структурная, прикладная и математическая лингвистика» / С. В. Меркулова. — М., 1999. — 189 с.

5. Мисак Р. Комп'ютерні словники : класифікація та укладання / Роман Мисак // Проблеми української термінології : [зб. наук. праць]. — Львів : Національний університет «Львівська політехніка», 2008. — С. 52—55.

6. Нелюбин Л. Л. Перевод и прикладная лингвистика / Л. Л. Нелюбин. — М. : Высшая школа, 1983. — 208 с.

7. Перванов Я. А Заметки по электронной лексикографии [Електронний ресурс] / Я. А. Перванов. — Режим доступу : http://arxiv.org/ftp/arxiv/papers/ 1107/1107.1753.pdf

8. Перебийніс В. І. Традиційна та комп'ютерна лексикографія : [навч. посіб.] / В. І. Перебийніс, В. М. Сорокін — К. : Вид. центр КНЛУ, 2009. — 218 с.

9. Сивакова Н. А. Лексикографическое описание английских и русских фитонимов в электронном глоссарии : автореф. дис. … канд. филол. наук : спец. 10.02.21 «Прикладная и математическая лингвистика» / Н. А. Сивакова. — Тюмень, 2004. — 28 с.

10. Черницький В. Б. Комп'ютерна лексикографія : [навч. посіб.] / В. Б. Черницький. — Нац. ун-т кораблебудування ім. адм. Макарова. — Миколаїв : НУК, 2004. — 81 с.

11. Dziemianko A. On the use(fulness) of paper and electronic dictionaries [Electronic resource] / Anna Dziemianko. — Access mode : https://www.academia.edu/3195210/On_the_use_fulness_of_paper_and_electronic_dictionaries.

12. Gilles-Maurice de Schryver. Lexicographers' dreams in the electronic-dictionary age [Electronic resource] / Gilles-Maurice de Schryver. — Access mode : http://tshwanedje.com/publications/dreams.pdf.

13. Müller-Spitzer C. Textual structures in electronic dictionaries compared with printed dictionaries: a short general survey [Electronic resource] / Carolin Müller-Spitzer. — Access mode : http://www1.ids-mannheim.de/fileadmin/lexik/owid/ Preprint_Mueller-Spitzer_HSK5.4-Article11_01.pdf.

14. Nesi H. The Use and Abuse of EFL Dictionaries. How learners of English as a Foreign Language Read and Interpret Dictionary Entries (Lexicographica Series Maior 98)/ H. Nesi. — Tübingen: Max Niemeyer, 2000. — 149 p.

15. Veronica Pastor. Search techniques in electronic dictionaries: a classification for translators [Electronic resource] / Veronica Pastor, Amparo Alcina. — Access mode : http://repositori.uji.es/xmlui/bitstream/handle/10234/32816/45010.pdf? sequence=1.