Viktoriia I. Davydenko

PhD in Philology, Associate Professor,
Department of Foreign Languages
for Specific Purposes,
Faculty of Foreign Philology,
Mykhailo Dragomanov State University
of Ukraine,
Kyiv, Ukraine
https://orcid.org/0000-0003-3498-5138
e-mail: v.i.davydenko@udu.edu.ua

CONTEMPORARY UKRAINIAN COMPUTATIONAL DICTIONARIES: HISTORY, ACHIEVEMENTS, AND CHALLENGES

Abstract

The article deals with the factors that led to the emergence of Applied Linguistics, Computational Linguistics, and Computational Lexicography – these are the processes of global digitalization of the world, the expansion of the scale of the scientific and technical revolution in the field of information systems and computer communication, the formation of the knowledge industry, requests on the development of radically new means of creating, storing, transforming and searching for information, as well as the integration of scientific fields. It is noted that at the beginning of the 21st century, lexicography incorporated centuries-old experience of theoretical linguistics and practical linguistics in the works of lexicography, and on the other hand, it received unprecedented development prospects associated with the use of computer technologies. It is emphasized that the social significance of lexicographic products is currently growing steadily. Modern science strives to embody all aspects of acquired knowledge in dictionary form, because dictionaries not only record the totality of certain knowledge, but also serve as a reliable tool for scientific mastering of reality. The works of Ukrainian and foreign scientists on the specific topic were analysed. It has been proven that computational dictionaries as a type of social, research or educational communication are the most popular reference materials, a source of access to any type of information (technical, encyclopaedic, linguistic, etc.)

due to the fixation of the modern level of knowledge, completeness, accuracy, reliability, a systematic approach to the organization and presentation of knowledge, accessibility, extreme "comfort" for use, etc. Dictionary works are also the goal and result of scientific work on linguistic and technical issues.

Keywords: *Applied Linguistics, Computational Linguistics, Computational Lexicography, dictionaries.*

1. Introduction.

Thanks to the processes of global digitalization of the world, "expanding the scope of the scientific and technical revolution in the field of information system and computer communications, the formation of the knowledge industry" (Shyrokov, 2011), requests on the development of radically new means of creating, storing, transforming and searching for information, as well as the integration of scientific fields, led to the emergence in the 20th century of Applied linguistics and Computational Linguistics, on the directions of which is Computational Lexicography. Currently, the mentioned scientific and practical areas of information technology are dynamically developing and are increasingly becoming part of our life.

At the end of the last century M. Peshchak (1999, p. 6) proved the interdependence of the development of language and linguistic science with the development of technology and engineering and technical theory about the production practice of society, she analysed the processes of such state of development of the literary language, when its use is transferred from the sphere of mass media (printing, radio, television, cinema, etc.) into the field of human-machine communication – computer automation of intellectual social activity. Here, the literary language, being by nature an exclusively humanitarian phenomenon, begins to work successfully in substance and in the technical field.

At the beginning of the 21st century, lexicography emerged as a direction of human activity, which, on the other hand, conceptualized and embodied in the works of lexicography the centuries-old experience of the theoretical acquisitions of theoretical linguistics and practical linguistics in the works of lexicography, and on the other hand, it received unprecedented development prospects associated with the use of computer technologies. The social significance of lexicographic products is steadily growing. Modern science seeks to embody all aspects of acquired knowledge in dictionary form; dictionaries not only record the totality of certain knowledge, but also serve as a reliable tool for scientific mastering of reality (Shyrokov, 2011, p. 3).

Applied and Computational Linguistics, and together with Computational Lexicography, "researching and developing the technology of constructing linguistic systems, make it possible to significantly increase labour productivity in the most diverse areas of social production, significantly speed up to the processing of linguistic information" (Partyko, 2008, p. 12).

Ukrainian scientist A. Zahnitko (2015, p. 26) notes that the modern system of linguistic sciences is uneven, covering classical theoretical linguistics, where the main areas of study are related to the consideration of phonology, morphology, syntax. Lexical semantics, to which modern scientific-linguistic paradigms have added cognitive linguistics, functional linguistics and generative linguistics; descriptive linguistics, which includes anthropological linguistics, evolutionary linguistics (comparative-historic linguistics, etymology), phonetics and sociolinguistics, and Applied linguistics is computational linguistics, mathematical linguistics, forensic linguistics, legal linguistics, language assessment and teaching methods / Language development), prescriptive linguistics, neurolinguistics, psycholinguistics, stylistics, etc.

2. Literature Review.

In Ukraine, T. Gryaznukhina, A. Zahnitko, N. Darchuk, Ye. Karpilovska, N. Klymenko, V. Perebiynis, M. Peshchak, V. Rusanivskyi, N. Snizhko, O. Taranenko, V. Shyrokov and others have made significant improvements to the development of the theoretical and applied foundations of compiling computational dictionaries.

A. Zahnitko determined the status of Applied linguistics at the current stage of the development of fundamental and practical sciences, revealed the interaction of this science with Computational linguistics and Computational lexicography, outlined two vectors of application for Applied linguistics: practical directions related to language learning, – lexicography, Linguistic didactics, terminology, translation studies, and directions related o practical functioning, – Computational linguistics with its main tasks of machine translation, automatic speech recognition, character recognition, automatic data editing, construction knowledge management systems, creation of electronic dictionaries, thesauruses, ontologies, corpus linguistics, as well as linguistic expertise and the science of organizing and standardizing scientific and technical terms (Zahnitko, 2015, p. 26). He also revealed the functions of language at the modern stage of development. Thus, "the following scientific disciplines are primarily oriented to the optimization of the communicative function: translation theory, machine translation, theory and practice of native and non-native language teaching, theory and practice of information and search systems, creation of informational (broader – artificial) languages, coding theory. Research in sociolinguistics (in particular, the study and justification of language policy), orthography and orthography, the theory of influence, political linguistics is correlated with the optimization of the social function of language (the communicative part). To optimize the epistemic (epistemological) function-oriented lexicography (in particular, Computational), research on terminology, corpus, and field linguistics. Optimization of the cognitive function of language is directly correlated with studies in Computational linguistics, psycholinguistics and aphasiology, quantitative linguistics, "linguistic criminology". Applied linguistics can be considered in the field of means of optimizing the functioning of language as a means of transmitting information. Here, translation studies dimensions come first (ibid., p. 26).

N. Pushyk analysed the content and close connection of Computational linguistics and artificial intelligence, singled out their interaction and justified from a linguistic point of view the need for a detailed study of Computational linguistics and the development of artificial intelligence. In particular, she noted that "language is closely and inextricably linked with thinking, concepts, categories, and symbols reflect the results of cognition, therefore the study of the laws of the existence of natural language is an important direction of artificial intelligence research, which continues the traditions of Computational linguistics and requires further interdisciplinary development. Key research in this field is devoted to the problems of comprehension, speech production, language learning, which reflect the classical tasks of artificial intelligence, such as perception, communication, knowledge, planning, thinking, and learning. N. Pushyk (2021, pp. 151–155) considers that the most important achievements in the field of Computational linguistics are: 1) automation of compilation and linguistic processing of machine dictionaries; 2) automation of the process of detecting and correcting errors when entering text into a computer; 3) automatic indexing of documents and information requests; 4) automatic classification and abstracting of documents; 5) linguistic provision of information search processes in monolingual and multilingual databases; 6) machine translation of texts from one natural language to another; 7) construction of linguistic processors that ensure communication of users with automated intelligent information systems in natural language or language close to natural; 8) extraction of factual information from differ texts.

V. Shyrokov, the founder and director of the Ukrainian Language and Information Fund of the National Academy of Sciences of Ukraine, described the cognitive, linguistic and system engineering principles of modern Computational lexicography, revealed the phenomenology of the lexicographic effect in information systems, noted modern software tools used in Computational lexicography, explained the method and basics of the formation of virtual lexicographic laboratories – an effective and promising instrumental means of supporting modern lexicography. He singled out four urgent and important problems of modern lexicography: the first is the problem of updating fundamental lexicons. After all, large dictionaries are, of course, exemplary lexicographic objects, created by large teams of highly qualified specialists; the second is the need to transfer the general cultural potential, as well as for use in automatic information processing systems, the third is the use of dictionaries in the formation of linguistic components of conceptual knowledge representation systems (such as ontology) and their use in knowledge mining tools; the fourth is the solution to the problem of multilingualism, which has become especially acute in the era of globalization and increasingly insistently requires powerful and flexible lexicographic tools for its solution (Shyrokov, 2011, p. 6).

S. Fokin considered the types, structure, characteristics, application of computational dictionaries, fundamental differences between electronic and paper lexicographic works. The scientist revealed the features and advantages of using bilingual and monolingual translation, onomasiological, terminological dictionaries, electronic concordances, auxiliary programs: text editors, checking the quality of translation work, natural language processing in the artificial intelligence system, etc. – precisely in the translator's activity (Fokin, 2019).

N. Snizhko claims that "the beginning of the 21st century is a period of active establishment of the methodology of integral linguistic research, systematization of knowledge about the world and languages in integrated lexicographic environments". The purpose of modern integral linguistic research is to clarify the development of Ukrainian conceptualization and categorization of reality, modelling of language and conceptual pictures of the world, and perception. Traditional interpretative lexicography, based on the theory of systematic arrangement of vocabulary, forms the basis of the formation of the leading directions of integral linguistic research. It is promising to develop integrated lexicographic systems, which concentrate various aspects of linguistic and micro-scientific data, as evidenced by such fields as lingoculturology, lingoconceptology, sociolinguistics, etc., the novelty of modern linguistic Ukrainian studies is in the active development of integral lexicographic environments (systems) capable to concentrate on rich multifaceted information about the word and confirm it with illustrative material (Snizhko, 2017, pp. 47–49).

3. Aims and Objectives.

The aim of the article is to study the achievements, understand the potential and outline the prospects of Ukrainian Computational Lexicography as a modern innovative resource. The objectives of the work are:

− to investigate the latest publications of Ukrainian and foreign scientists in the field of Applied Linguistics, Computational Linguistics and Computational Lexicography;

− to examine current state, possibilities, features of Ukrainian Computational dictionaries;

− to analyse ways of development and improvement of the specified lexicographic works.

4. Methodology.

We have analysed the works of Ukrainian and foreign scientists on Applied linguistics, Computational linguistics, Computational lexicography, analysed Ukrainian computational dictionaries, etc. Based on this, we described the current state, tasks, peculiarities, and prospects for the development of Computational lexicography and electronic dictionaries in Ukraine.

5. Results and Discussion.

Computational dictionaries as a type of social, research or educational communication are the most popular reference materials, a source of access to any type of information (technical, encyclopaedic, linguistic, etc.) thanks to the fixation of the modern level of knowledge, completeness, accuracy, reliability, a systematic approach to the organization and presentation of knowledge, accessibility, extreme "comfort" for use, etc. Dictionary works are also the goal and conclusion of scientific work on linguistic and technical issues.

Modern English computational lexicography is presented in;
– explanatory dictionaries of the English language: Oxford Learner's Dictionary, Cambridge Learner's Dictionary, Collins COBUILD Student's Dictionary, Longman Dictionary of Contemporary English, The Living Dictionary, etc.;
– educational dictionaries for certain groups of users;
– terminological dictionaries, especially on finance, business, banking, insurance medicine, logistics, and so on;
– dictionaries of slang, new words, the language of writers and politicians, phraseological units, etc.;
– dictionaries for entertainment, "the main object of which is the creation of reference books for hobbies and crosswords, as well as encyclopaedia dictionaries for movies and computer games, which are related together with new gaming and television projects" (Kalymon, 2019, p. 116).

The most authoritative and powerful educational British publishing houses are Cambridge University Press, Oxford University Press, Chambers, Collins, Longman, Macmillan, Pearson.

Scientists claim that "philological science over the centuries-old history of its existence and development has developed systematic approaches to define the concept of a dictionary. In the narrow sense, a dictionary (lexicological work) is a set of linguistic units arranged in a certain order, where their meaning is disclosed, various information is provided about them or a translation into another language, or information is included about the objects, phenomena, and facts that they denote. If the dictionary is understood as the result or research work on a certain problem, as the development of a certain linguistic problem or as a tool for solving new tasks, then the concept of "dictionary" expands in the direction of indicating a systematized collection of knowledge about a certain problem and acquires well-defined connotations to cognitive structures of the type of knowledge models (Shyrokov, 2011, pp. 15–16).

The theoretical lexicography provides the definition of a dictionary as a specific linguistic and informational object, into which the linguistic substance "tries" to be embodied in the process of unfolding the "subject – object" and "form – content" relations fundamental to language (ibid., pp. 3–4).

Computational lexicography has a powerful potential due to modern technical innovative methods and techniques. It has incomparable advantages over traditional dictionaries: it is manifested in the possibilities of searching not only by the title word, but by

any specified parts, active use of hyperlinks, compilation of dictionaries with previously unseen parameters as new lexicographic objects, etc.

Currently, the following powerful centres of research and creating of computational lexicographic works are known in Ukraine: the Ukrainian Language and Information Fund of the National Academy of Sciences of Ukraine, Laboratory of Computational Linguistics of the department of Contemporary Ukrainian language of Kyiv National University named after T. Shevchenko, Department of Lexicology and Computational Lexicography of the Institute of Ukrainian language of the National Academy of Sciences of Ukraine, Lviv Polytechnic University.

O. Sydorenko notes, the Department of Structural and Mathematic Linguistics of the Institute of Linguistics named after O. O. Potebnia of the National Academy of Sciences of Ukraine may be called the founder of Computational Linguistics in Ukraine. Here, the Morpheme-Word Formation Fund of the Ukrainian language was created, which has a branched architecture and consists of three main subfunds:1) a text base, which is currently equipped with procedures for orthographic control of texts, analysis of their morphological, syntactic and semantic structure; 2) the general register of Ukrainian words, based on the materials of the 5 most authoritative Ukrainian dictionaries-sources; 3) text processors performing morphological, syntactic and logical-semantic analysis of actual materials.

Based on the materials of the fund, the computer-based: "Словник символьних моделей морфемної будови слова" (Dictionary of symbolic models of the morpheme structure of the word), "Словник афіксальних морфем української мови" (Dictionary of affixal morphemes of the Ukrainian language), also published in paper form, "Кореневий гніздовий словник української мови" (Root Nest Dictionary of the Ukrainian language), "Ідеографічний словник іменників української мови" (Ideographic dictionary of nouns of the Ukrainian language), "Ідеографічний словник дієслів переміщення української мови" (Ideographic dictionary of verbs of movement of the Ukrainian language) were created (Sydorenko, 2010, pp. 524–528).

The Ukrainian Language and Information Fund of the National Academy of Sciences of Ukraine under the leadership of V. Shyrokov started in 1994 and coordinates the program to create a series of dictionaries of a new generation – "Dictionaries of Ukraine", which already has more than 70 works. Among them, it should be noted the first full-scale electronic dictionary of the Ukrainian language in Ukraine – the integrated lexicographic system "Dictionaries of Ukraine" appeared in 2001. A CD of this system was released, which has an alphabetical construction principle and a search system with a volume of about 152,000 tokens (also published in print). Users can work in 5 modes: "paradigm", "transcription", "phraseology", "synonymy" and "antonymy". Each of these sections is based on several of the most authoritative dictionaries of the corresponding variety. 11 versions of the specified system have been commercially released on laser discs since 2001.

Under the leadership of V. Shyrokov, the National Dictionary Base of Ukraine is being created – the development of an information-linguistic system and a linguistic corpus, which has a volume of more than 200 million words and forms and is the basis of modern lexicographic and linguistic research.

The linguistic portal mova.info presents:

‒ frequency dictionaries;

‒ "Електронний граматичний словник української літературної мови (словозміна) I етап" (Electronic grammatical dictionary of the Ukrainian literary language (word change) 1 stage);

‒ "Синтез англійських парадигм" (Synthesis of English paradigms – the electronic program was created for English nouns and verbs);

− "Електронний словник мови Тараса Шевченка" (Electronic dictionary of the language of Taras Shevchenko);

− "Відкритий словник" (дискусійна платформа) (Open dictionary (discussion platform));

− "Перекладний чотиримовний словник наукової термінології" (Translated four-language dictionary of scientific terminology);

− "Глоса" (Glosa);

− "Труднощі англійського слововживання" (Difficulties of English word usage);

− "Українсько-італійський словник" (Ukrainian-Italian Dictionary);

− "Словник порівнянь" (Dictionary of comparisons);

− "Тезаурус з комп'ютерної лексикографії" (Thesaurus of computational lexicography);

− "Короткий українсько-сербський словник сполучуваності слів" (Short Ukrainian-Serbian word conjugation dictionary);

− "Словник іншомовних слів" (Dictionary of foreign words) etc.

6. Conclusions.

Computational lexicography is closely related and successfully interacts with such modern sciences as Applied linguistics, Computational linguistics, Mathematical linguistics, Artificial intelligence, etc., as well as with Cognitive, Functional, Generative, Evolutionary linguistics, and so on.

At the beginning of the 21st century, Lexicography embodied the century-old experience of theoretical and practical Lexicography in the lexicographic works. Computational lexicography successfully combined these achievements with the latest contemporary computer technologies. Thus, Computational lexicography has much greater opportunities in compiling modern lexicographic works, reference books, dictionaries of new generation, creating products according to many given parametric characteristics.

Computational lexicography is rapidly developing and constantly improving: its methods, approaches, technologies, techniques, modelling, etc. are changing due to contemporary challenges, aims, objectives and requests.

Recently, the field of application of computational dictionaries in Linguistics, Corpus linguistics, Linguistic didactics, Terminology, Translation studies, as well as in machine translation, automatic recognition of speech and symbols, data editing, construction of knowledge management systems, etc. has been expanding.

Significant achievements of contemporary Ukrainian studies in the field of Computational lexicography are: National Dictionary Base of Ukraine, Integrated lexicographic system "Dictionaries of Ukraine"; computational dictionaries etc.

Despite the achievements of Computational lexicography in Ukraine, the branch faces the tasks of improving existing works, as well as creating new dictionaries as objects with various parametric characteristics that cannot be compiled in a paper version.

R e f e r e n c e s

Zahnitko, A. (2015). Status prykladnoi linhvistyky v systemi linhvistychnykh nauk / Anatolii Zahnitko, Illia Danyliuk, Zhanna Krasnobaieva-Chorna, Oksana Putilina, Hanna Sytar. Paragyhmalno-katehoriini osnovy prykladnoi linhvistyky: Monohrafiia [Paradigm-categorial foundations of Applied linguistics: Monograph]. Vinnytsia: "Tov Nilan-LTD", 26–47. [in Ukrainian].

Kalymon, Yu. (2019). Kompiuterna leksykohrafiia: vyklyky ta perspectyvy [Computational lexicography: challenges and prospects]. *Aktualni pytannia inozemnoi filolohiii*, 10, 112–118. [in Ukrainian].

Partyko, Z. V. (2008). Prykladna I kompiuterna linhvistyka: Vstup do spetsialnosti: Navchalnyi posibnyk [Applied and Computational Linguistics: Introduction to the speciality: Study guide]. Lviv, 224 p. [in Ukrainian].

Peshchak, M. M. (1999). Narysy z kompiuternoi linhvistyky: Monohrafiia [Essays on Computational linguistics]. Uzhhorod, 200 p. [in Ukrainian].

Pushyk, N. V. (2021). Kompiuterna linhvistyka ta "shtuchnyi intelect" [Computational linguistics and artificial intelligence]. *Young Scientists*, 2 (90). February, 151–155. doi: https://doi.org/10.32839/2304-5809/2021-2-90-29

Shyrokov, V. A. (2011). Kompiuterna leksykohrafiia [Computational lexicography]. Kyiv, 351 p. [in Ukrainian].

Snizhko, N. (2017). Nova dzherelna baza ukrainskoi leksykohrafii v systemi intehralnykh linhvistychnykh doslidzhen [New source base of Ukrainian lexicography in the system of integral linguistic research]. *Human. Computer. Communication*. 20–22 September. Lviv, Ukraine, 47–52. [in Ukrainian].

Sydorenko, O. M. (2010). Ukrainska kompiuterna leksykohrafiia yak vazhlyvyi innovatsiinyi chynnyk navchalnoho protsesu [Ukrainian computational lexicography as an important innovative factor in the educational process]. *Aktualni problemy slovianskoi filolohii.* Vyp. XXIII. Chastyna 1, 524–528. [in Ukrainian].

Fokin, S. (2019). Kompiuterni innovatsiini tekhnolohii v perekladatskii ta perekaldoznavchii diialnosti: Navchalnyi posibnyk [Computational innovative technologies in translation and translation studies: Study guide]. Kyiv, 254 p. [in Ukrainian].

Research initiative. (n.d.). Retrieved October 10, 2023, from: http://www.mova.info

Research initiative. (n.d.). Retrieved October 10, 2023, from: https://www.ulif.org.ua

Research initiative. (n.d.). Retrieved October 10, 2023, from: https://collins.co.uk/collections/collins-cobuild-dictionaries-for-learners

Research initiative. (n.d.). Retrieved October 10, 2023, Research initiative. (n.d.). Retrieved October 10, 2023, from: https://dictionary.cambridge.org/dictionary/learner-english

Research initiative. (n.d.). Retrieved October 11, 2023, from Longman Dictionary of Contemporary English, https://www.ldoceonline.com

Research initiative. (n.d.). Retrieved October 11, 2023, from: https://www.oxfordlearnersdictionaries.com

Бібліографічний опис:

Анотація

У статті розглянуто чинники, що призвели до появи прикладної лінгвістики, комп'ютерної лінгвістики і комп'ютерної лексикографії, – це процеси глобальної діджиталізації світу, розширення масштабів науково-технічної революції у галузі інформаційних систем та комп'ютерних комунікацій, становлення індустрії знань, запити на розвиток кардинально нових засобів створення, зберігання, перетворення й пошуку інформації, а також інтеграція наукових галузей. Зазначено, що на початку XXI століття лексикографія втілила у творах словникарства багатовіковий досвід теоретичних надбань лінгвістики і практичного лексикографування, а з іншого боку, – отримала досі небачені перспективи розвитку, пов'язані із застосуванням комп'ютерних технологій. Наголошено, що наразі відбувається неухильне зростання соціальної значущості лексикографічної продукції. Сучасна наука прагне втілити у словниковій формі всі аспекти отриманих знань, тому що словники не лише фіксують сукупність певних знань, але й слугують надійним інструментом наукового освоєння дійсності. Проаналізовано роботи українських і зарубіжних учених із зазначеної теми. Доведено, що комп'ютерні словники як тип соціальної, дослідницької або навчальної комунікації є найпопулярнішими довідниковими матеріалами, джерелом доступу до будь-якого виду інформації (технічної, енциклопедичної, лінгвістичної та ін.) завдяки фіксації сучасного рівня знань, повноті, точності, достовірності, системному підходу до організації й представленню знань, доступності, надзвичайній "комфортності" для користування тощо. Словникарські праці також є метою й підсумком наукової роботи над лінгвістичними і технічними проблемами.

Ключові слова: прикладна лінгвістика, комп'ютерна лінгвістика, комп'ютерна лексикографія, словники.