UDC 81'374

THE ACTIVE LEARNER'S CONSTRUCTION-COMBINATORY THESAURUES: USER-DRIVEN PRINCIPLES OF COMPILING (A COGNITIVE LINGUISTIC APPROACH)

Svitlana Zhabotynska
(Bohdan Khmelnitsky National University of Cherkasy, Cherkasy, Ukraine)
Yevhenii Plakhotniuk
(Kyiv National Linguistic University, Cherkasy, Ukraine)

S. Zhabotynska, Ye. Plakhotniuk. The Active Learner's Construction-Combinatory Thesaurus: user-driven principles of compiling (a cognitive linguistic approach). This article discusses the design of a new type of dictionaries, the Active Learner's Construction-Combinatory Thesaurus (ALCCT) intended for adult learners of the second (foreign) language. The ALCCT is an ideographic dictionary where phrases, understood as instantiations of constructions, are arranged in accordance with the cognitive ontology of a particular conceptual thematic field. As such, the ALCCT is a project compatible with cognitive lexicography, a contemporary branch of dictionary-making that adopts the findings of cognitive science, cognitive linguistics in particular. The article proposes a cognitive profile of the dictionary's target user, and makes it a departure point in elaborating the principles of compiling the ALCCT. They are defined as the principles of data selection, arrangement, and application. *Data selection* regards their thematic and formal coherence, their authenticity, and their prominence, or frequency. *Data arrangement* implies their relational coherence, and their elaboration. Relational coherence is realized through the lexicographic code, or the dictionary's overall design that develops at three hierarchical levels: those of macrostructure (a conceptual ontology of the theme), mediostructure (the key words evolving into phrasal sets), and microstructure (description of phrasal lemmas). Elaboration of the data is provided via the overarching structure mapped onto the three hierarchical structures of the lexicographic code, and concerned with etymological, cultural, grammatical, and phraseological (metaphorical) extensions. *Data application* reaches out to the communicative situations in which the ALCCT's resources can be used.

Key words: cognitive lexicography, Active Learner's Construction-Combinatory Thesaurus (ALCCT), user's cognitive profile, principles of compiling, data selection, data arrangement, data application.

С. Жаботинська, Є. Плахотнюк. Активний навчальний конструкційно-комбінаторний тезаурус: принципи укладання з урахуванням користувача (лінгвокогнітивний підхід). У статті запропоновано проєкт нового типу словника – активного навчального конструкційно-комбінаторного тезауруса (АНККТ), призначеного для дорослих користувачів, які вивчають іноземну мову. АНККТ є ідеографічним словником, в якому словосполучення, потрактовані як утілення конструкцій, упорядковано на підставі когнітивної онтології конкретної тематично визначеної поняттєвої царини. Тим самим АНККТ є проєктом, узгоджуваним із когнітивною лексикографією – сучасною галуззю словникарства, яка послуговується доробком когнітивної науки, зокрема, когнітивної лінгвістики. У статті потенційний користувач словника представлений у когнітивному профілі, який вмотивовує принципи укладання АНККТ. Вони визначаються як принципи відбору, упорядкування та застосування даних. *Відбір даних* ураховує їхню тематичну і формальну когерентність, автентичність та промінантність, або частотність. *Упорядкування даних* передбачає їхню реляційну когерентність і їхнє поглиблення. Реляційна когерентність утілюється в лексикографічному коді – цілісному ієрархічному дизайні словника, представленому макроструктурою (концептуальною онтологією

теми), медіоструктурою (ключовими словами у складі фразових сетів) і мікроструктурою (описом фразової леми). Поглиблення даних залучає надструктуру, яка проєктується на три ієрархічні структури лексикографічного коду і забезпечує етимологічне, культурологічне, граматичне та фразеологічне (метафоричне) поширення словникового матеріалу. Принцип *застосування даних* пов'язаний із виходом у комунікативні ситуації, які потребують мовного забезпечення, наданого в АНККТ.

Ключові слова: когнітивна лексикографія, активний навчальний конструкційно-комбінаторний тезаурус (АНККТ), принципи укладання, відбір даних, упорядкування даних, застосування даних.

С. Жаботинская, Е. Плахотнюк. Активный учебный конструкционно-комбинаторный тезаурус: принципы составления с учетом пользователя (лингвокогнитивный подход). В статье предлагается проект нового типа словаря – активного учебного конструкционно-комбинаторного тезауруса (АУККТ), предназначенного для взрослых пользователей, изучающих иностранный язык. АУККТ является идеографическим словарем, в котором словосочетания, трактуемые как воплощения конструкций, упорядочены на основе когнитивной онтологии конкретной тематически определенной понятийной области. Тем самым проект АУККТ согласуется с когнитивной лексикографией – современной отраслью составления словарей, использующей наработки когнитивной науки, в частности, когнитивной лингвистики. В статье потенциальный пользователь словаря представлен в когнитивном профиле, мотивирующем принципы составления АУККТ. Они определяются как принципы отбора, упорядочивания и использования данных. *Отбор данных* учитывает их тематическую и формальную когерентность, аутентичность и проминантность, или частотность. *Упорядочивание данных* предусматривает их реляционную когерентность и их углубление. Реляционная когерентность реализуется в лексикографическом коде – целостном иерхическом дизайне словаря, представленном макроструктурой (концептуальной онтологией темы), медиоструктурой (ключевыми словами в составе фразовых сетов) и микроструктурой (описанием фразовой леммы). Углубление данных осуществляется посредством надструктуры, которая проецируется на три иерархические структуры лексикографического кода и обеспечивает этимологическое, культурологическое, грамматическое и фразеологическое (метафорическое) расширение словарного материала. Принцип *использования данных* предполагает выход в коммуникативные ситуации, которые требуют языкового обеспечения, представленного в АУККТ.

Ключевые слова: когнитивная лексикография, активный учебный конструкционно-комбинаторный тезаурус (АУККТ), принципы составления, отбор данных, упорядочивание данных, использование данных.

1. Introduction

The anthropocentric orientation of nowadays linguistics echoes in the 'user perspective', or 'user-driven' approach in theoretical and practical lexicography (Tarp, 2008; 2011). This approach accentuates the necessity of making the dictionary more 'user-friendly', which is achieved via devising a 'target user profile' that guides the dictionary design (Tarp, 2008). A 'user-driven' approach is particularly important for compiling bilingual dictionaries intended for learners of the second / foreign language (L2). The available 'profiles' of dictionary users are mostly concerned with their "common sense" employed in decoding a dictionary entry (Tarp, 2008, pp. 41–43, 82–85). The respective strategies of compiling dictionaries tend to introduce some modest amendments to their text format or simplify the definiens. Most of such L2 dictionaries, consulted only sporadically, are de facto *passive*, i.e. alphabetically structured, and focused on the systemic properties of linguistic expressions. The user's cognitive capacities regularly employed in L2 acquisition and speech production remain under-addressed, since traditionally oriented dictionary-makers do not consider "mental processes in the brain" to be a matter of lexicography (Tarp, 2008, p. 132). Meanwhile, lexicography may benefit from the ideas of cognitive linguistics which dovetails mental and linguistic phenomena, and which may contribute to developing *active* dictionaries (Apresyan, 2010; Fuentes-Olivera & Bergenholtz, 2018) that aim to practically assist in L2 learning and teaching.

This article discusses the principles of compiling a particular kind of active dictionaries—the *Active Learner's Construction-Combinatory Thesaurus (ALCCT)* (Plakhotniuk, 2020b), which is an updated version of the combinatory thesaurus grounded on a conceptual ontology (Zhabotynska, 2010; 2019). Designing the ALCCT complies with the emergent field of cognitive lexicography (Ostermann, 2015) that bridges the theory and praxis of dictionary-making with cognitive linguistics and the broader field of cognitive science. In cognitive lexicography, the dictionary is expected to represent schematic patterns of cognition that are tracked in various linguistic data, and thus are relevant for language acquisition and speech production.

The discussion below focuses on a cognitive profile of the ALCCT user, a cognitive linguistic background of the dictionary design, the ALCCT's semiotic interpretation that provides a systemic approach to the user's needs, and the principles of compiling the ALCCT that satisfy these needs. The concluding discussion outlines the theoretical and practical implications of this study.

2. Cognitive profile of the ALCCT user

A 'user-friendly' dictionary should be compiled with regard to the user's cognitive capacities intended for processing linguistic information. Since the ALCCT is addressed to adult learners, these capacities are the ones inherent in the adult mind / brain. Its properties, explored by different branched of cognitive science, define the ALCCT's general objectives. In our research, they are further specified in a pilot survey on the needs and expectations of the ALCCT's potential adult users.

Neurolinguistics argues that language acquisition depends on both nature and nurture, i.e. language evolves at the intersection of biological and societal factors. Biological factors relate to the language faculty existing in the human mind, and societal factors are represented by the linguistic environment which activates this faculty (Zhabotynska, 2020, pp. 102–103). Language faculty, as a natural endowment, can be properly activated only at a particular age. For L1, this age (up to 8–12) is called *the critical period*, because a child not exposed to any language during this time will not be able to achieve an adequate proficiency in speaking and thinking. For L2, the same period is defined as *sensitive*, because L2, similarly to L1, is acquired unconsciously, with assistance of the procedural memory. The period after the age of 8–12 is called *post-sensitive*. Now, L2 is learned consciously, being assisted by the declarative memory (Lenneberg, 1967; McWhinney, 2005; Paradis, 2005; Zhabotynska, 2020, p. 103). As the ALCCT is intended for adults who learn L2 in the post-sensitive period, a profile of the ALCCT's user may incorporate the findings of cognitive and andragogic research on the brain / mind workings in adulthood.

According to andragogic studies, adult learners differ from children in such aspects as previous experience, internal motivation, need for implementing the social roles, awareness of the on-going rational cognitive activities and immediate application of new knowledge (Knowles, 1984). Studies of the post-sensitive period in language acquisition, as well as those concerned with the ageing brain and adult learning, provide evidence for the neurocognitive basis of this difference (for review see Zhabotynska & Plakhotniuk, 2016). The processes involved in adult learning are automated due to lateralization, myelination, development of the prefrontal cortex (the age of 14–21) and the default neural network, particularly in the medial prefrontal cortex (the age of 21–31). Adult learning becomes more efficient in terms of neural connectivity (Fair et al., 2008, p. 4030) and conscious conceptualization, or abstract thinking, as well as retrospective and creative use of information (Fair et al., 2008, pp. 4028–4029). Learning per se changes the language-related areas of the adult brain both functionally and structurally (Mårtensson et al., 2012). The role of comprehensive, enriched and meaningful input stimulating learning-based neuroplasticity in adulthood seems to be crucial and, as such, emphasized throughout the literature (Caine & Caine, 1994, pp. 30–33; Valipour & Asl, 2014).

Respectively, verbal and non-verbal mental representations are not chaotic. Instead, the researchers report on stable patterns of self-organization of information at conceptual and linguistic

levels, i.e. embeddedness and interconnectedness (Caine & Caine, 1994, p. 39), which provides evidence for a certain degree of iconic motivation between the external / formal linguistic patterns and internal / conceptual patterns (Perniss, Thompson, & Vigliocco, 2010). For instance, neural activations triggered by listening to audio-texts reveal semantic grouping throughout the cerebral cortex. Researchers associate such grouping with a certain interdependence between symbolic representations and bodily (perceptual) experience in mental schemata (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). Hence, information available in the ALCCD intended for adult learners should be provided in a systematic, structured, integrative and pragmatically driven way that is isomorphic to the way in which the adult brain / mind processes linguistic and conceptual information.

The results of our pilot survey on expectations of the potential ALCCT adult users are compatible with the conclusions of cognitive and andragogic studies as to workings of the adult brain / mind. The interviewed participants turned out to favor (i) a thematic arrangement of the dictionary instead of its alphabetical arrangement, (ii) exposure to the key words of the theme, to their synonyms, and to the phrases in which they are used, (iii) presence of syntactic patterns according to which these phrases can be transformed, (iv) availability of instruction as to combining the phrases into sentences that make up a text applicable in communication (Plakhotniuk, 2020a).

The above preferences of adult users are reflected in the ALCCT as a dictionary type: (a) it is an *ideographic (onomasiological) thesaurus* featuring a thematically homogeneous conceptual field; (b) it is a *combinatory thesaurus*: its units (lemmas) are phrases / word-combinations with the key words of the thematic field; (c) it is a *construction-combinatory thesaurus*: it provides the patterns of phrases as constitutive elements of sentences; (iv) it is *an active learner's thesaurus*: its design actively assists the learners in L2 acquisition and speech production (Plakhotniuk, 2020b). The ALCCT differs from the existing lexicographic projects of active dictionaries (see overview in Fuentes-Olivera & Bergenholtz, 2018). The ALCCT is to meet the user's primary, secondary and tertiary needs (Tarp, 2011, p. 283), defined so with regard to the order in which they are satisfied. The primary needs are concerned with the type of lexicographic data. The secondary needs associate with appropriateness of the dictionary's design which has to be compatible with the user's neuro-cognitive profile, thus facilitating L2 acquisition at the lexical and syntactic levels. The tertiary needs imply assistance of this dictionary in speech production, or developing thematically relevant texts and communicative skills.

Compiling the ALCCT according to the principles consistent with the ways in which the user's mind / brain processes linguistic and non-linguistic information presents a significant challenge for dictionary-makers. This challenge is attempted to answer in the cognitive linguistic conception termed Semantic of Lingual Networks (see the recent version in (Zhabotynska, 2018)) that underpins the combinatory thesaurus grounded on a conceptual ontology (Zhabotynska, 2010, 2015, 2019; Brovchenko, 2011; Radchenko, 2012, 2019). The ALCCT, which is an updated version of this thesaurus, has the same theoretical background that is briefly described below.

3. Cognitive linguistic background of the ALCCT

Semantics of Lingual Networks (SLN) has six theoretical statements based on the analysis of various linguistic phenomena. Four of these statements are immediately relevant for compiling the ALCCT (their further description corresponds to Zhabotynska, 2018).

(1) Conceptual structures that arrange the meanings of linguistic expressions are constituted by *basic propositional schemas (BPS)* which represent the most abstract conceptual categories and their relations. The BPSs are thematically grouped into five types: being, action, possession, identification, and comparison schemas.

– <u>Being schemas</u> include the *quantitative* (X is THAT MANY-Qn), *qualitative* (X is SUCH-Ql), *locative* (X exist THERE / LC-locative), *temporative* (X exists THEN / TM-temporative), and *mode of being* (X exists SO / MD-mode) schemas.

– Action schemas comprise the *state/process* (AG-agent acts), *contact* (AG-agent acts upon PT-patient / AF-affected), and *causation* (CR-causer makes FT-factitive) schemas.

– Possession schemas are represented by the *part-whole* (WH-whole has PR-part), *inclusive* (CR-container has CT-content / CT-content has CR-container), and *ownership* (OW-owner has OD-owned / OD-owned has OW-owner) schemas.

– Identification schemas are particularised as the *classification* (ID-identified = individual or kind is CL-classifier = kind or type), *characterization* (ID-identified = individual is CH-characteriser), and *personification* (ID-identified = individual is PS-personifier = a proper name) schemas. In English, CL is manifested with the indefinite article, and CH—with the definite article.

– Comparison schemas include the *identity / metamorphosis* (CV-comparative is [as] MS-correlate = another category of the same entity), *similarity / analogy* (CV-comparative is as AN-correlate = an entity from the same category), and *likeness / metaphor* (CV-comparative is as if MT-correlate = an entity from a different category) schemas.

The BPSs may get extension with additional argument roles: SC-circumstant (attendant, aid, counter-agent, instrument, mediator, means, and mode), ST-stimulus (cause and goal), PQ-prerequisite (condition and concession), RC-recipient (addressor, benefactor, and malefactor), LC-locative, and TM-temporative.

The BPSs integrate into an operational network employed in processing information about the objects of the experienced world. The number of BPSs is limited, but, arranged in various configurations, they structure an unlimited number of conceptual networks.

(2) A conceptual network may be transformed into a conceptual matrix, if the links between its nodes remain implicit.

(3) A conceptual network or matrix may be built at one level or several levels. In the latter case, the information evolves in-depth, being structured as 'networks-in-networks' or 'matrixes-in-matrixes'. The hierarchical conceptual levels are: a *conceptual thematic field* (all the information that is structured), the *domains* (focuses of the thematic field), *parcels* (focuses of the domains), and *concepts* which constitute parcels and which are structured as a set of properties. At all conceptual levels, the networks or matrixes are built with the BPSs that suchwise exhibit fractal properties. Thematically coherent information arranged with a conceptual network or matrix is defined as a *conceptual ontology*.

(4) The BPSs and their clusters are the meanings of syntactic *constructions*. In construction grammar, they are interpreted as schematic, or generalized linguistic forms that have their own schematic meanings existing independently of the words that fill out these forms (Goldberg, 1995; Ostman & Fried, 2004; Lyngfelt, Borin, Ohara, & Torrent, 2018 among others). The constructions whose meanings are represented by the BPSs are employed for both *categorization* and *re-categorization* of linguistic information. In case of re-categorization, the schematic meaning rendered by a BPS is manifested not by its own schematic form, but by the schematic form of some other BPS. For instance, the schematic form *NP$_2$ of NP$_1$*, the inherent meanings of which are represented by the possessive BPSs (*the page of a book, students of the group, a car of this owner*), can be used to explicate the other BPSs: qualitative (*beauty of the girl, a girl of beauty*), contact (*invitation of the student*), classification (*a game of chess*), likeness / metaphor (*a devil of a boy*), etc. In this case, the non-possessive propositional schemas are re-formatted as possessive, and their blended meaning integrates into the semiotic (syntactic) category of possession.

The above theoretical statements pairing linguistic and conceptual structures are relevant for developing user-driven principles of compiling the ALCCT. The system of these principles is prompted by the ALCCT's semiotic interpretation.

4. Semiotic interpretation of the ALCCT

Any dictionary is a text. Since the text can be viewed as a 'macrosign' (Vorobyova, 1993, p. 41), it agrees with the semiotic definition of a sign as the unity of a material form, the meaning which it evokes in the mind, and the function which it performs (Fig. 1).

In the dictionary, the *MATERIAL FORM* is the *lexicographic object*, or the data represented by particular linguistic expressions. In the ALCCT, they are phrases with the key words (nouns) that feature a particular conceptual thematic field identified as the dictionary's *MEANING*. In the ALCCT, which is an ideographic dictionary, the signified thematic field is arranged in accordance with a conceptual ontology that becomes a *lexicographic structure* providing a thematic and structural arrangement of the phrasal data. The latter obtain a *lexicographic description* that has its own design. Together, the lexicographic structure and lexicographic description make up a *lexicographic code* which is pivotal for the compiler. Lexicographic code is a system of methods employed for processing the lexicographic data of L2. The ALCCT's *FUNCTION* is assistance to users in L2 acquisition and speech production. The dictionary's multimodal (multisemiotic) text can be presented on a paper or digital carrier.
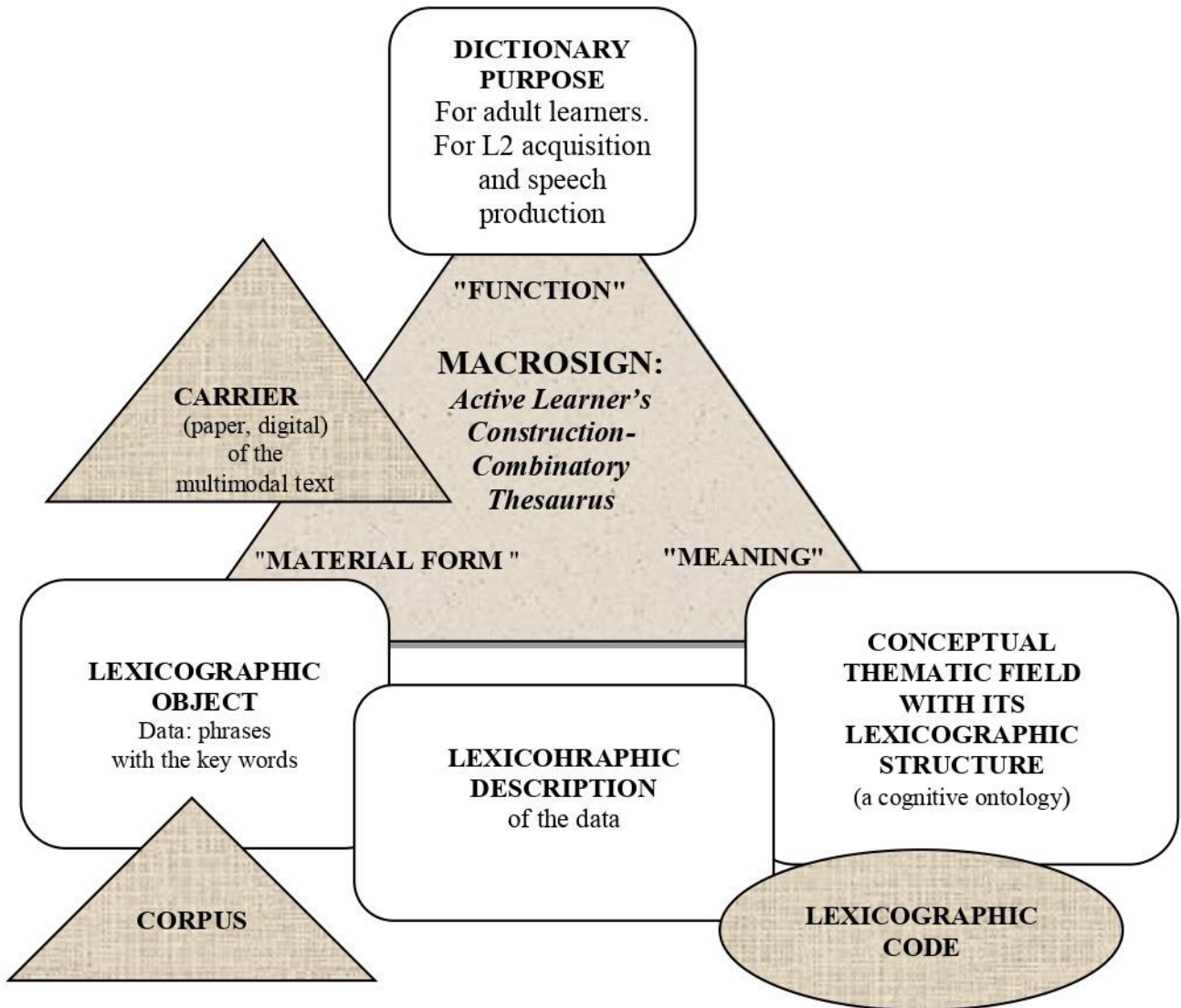
Figure. 1. Semiotic interpretation of the ALCCT

The ALCCT's semiotic interpretation prompts the system of principles applied in the dictionary design. These principles will be specified below.

5. Principles of compiling the ALCCT

Compiling the ALCCT with regard to its semiotic aspects is guided by (a) the principles of data selection (concerned with the lexicographic object), (b) the principles of data arrangement (concerned with the lexicographic code), and (c) the principles of data application (concerned with the dictionary purpose).

The principles of data selection are represented by their thematic and formal coherence, their authenticity, and their prominence.

Thematic coherence of the data means that they name a particular thematic field relevant for everyday or professional communication (e.g. SCHOOL, TRAVELLING, AIRPORT, COURT, MARKETING, etc.). Besides, the data include not only the key words (nouns) of the thematic domain, but also the synonyms of these words. Each key word evolves into a set of phrases that specify schematic content of the BPSs and their extensions. Formal coherence of the data means that the phrases belong to particular structural types (e.g., Adj N_1—*prestigious school*, N_2N_1—*boy school*, N_1N_2—*school teacher*, Prep N_1—*at school*, N_1V N_2 – *school admits students*, N_2V N_1–*students attend school*, etc.) that are formal correspondences of particular BPSs and their extensions. Thematic and formal coherence of the dictionary's data agree with what text linguistics calls "referential coherence of the text", or the continuing reference to the same entities figuring in the text (Dirven & Verspoor, 2004, p. 186).

Authenticity of the data means that the thematically and structurally coherent phrases are retrieved from a corpus of L2 authentic texts describing a particular theme, or topic. Thus, the ALCCT that features language used in speech represents a "usage-based model of language" bridging linguistic competence and performance (Tomasello, 2003; Boyland, 2009). In learning L2, authenticity is of particular importance, since combinability of words in phrases tends to be language-specific. That is, a phrase in L1 may not be a word-for-word translation in L2—the fact which tends to be ignored by L2 learners.

Prominence of the data means that their retrieval from a specialized corpus considers a frequency factor that defines the learning priorities. The existing research shows that nearly 10% of the total lexicon is composed of words that are most frequent, deeply entrenched, and applicable in defining the rest of the word-stock. This part of the lexicon, first acquired in ontogenesis, becomes the so-called "minimum grounding set" (Vincent-Lamarre et al., 2016, pp. 636–637). Accounting for frequency effects, prototypicality and associative connections between lemmas as well as the use of highly productive lexical and syntactic patterns might enhance L2 acquisition by adults (Tomasello, 2003; Frost, Siegelman, Narkiss, & Afek, 2013). The ALCCT defines three frequency groups of phrasal data (differentiated with colors) which correspond to the users' proficiency levels in L2 within a particular theme, and facilitate the choice of learning priorities (cf. Frost, Siegelman, Narkiss, & Afek, 2013). Reference to data frequency allows the users to prioritize their lexical and syntactic choices, prevents them from the information overload, and visualizes L2 prototypical expressions.

The principles of data arrangement include their relational coherence, and their elaboration.

Relational coherence of the data agrees with what text linguistics defines as "relational coherence of the text", or comprehensive links between its referents (Dirven & Verspoor, 2004, p. 189). In the ALCCT, relational coherence provides conceptual and linguistic congruence of the lexicographic code, or the dictionary's overall design. This design develops at three major hierarchical levels: those of macrostructure, mediostructure, and microstructure. The first two correspond to the 'lexicographic structure' per se, while the third level corresponds to the 'lexicographic description'.

The ALCCT's **macrostructure** demonstrates arrangement of the entire conceptual thematic field signified with the lexicographic data. Here, the key concepts are linked within parcels, and the latter are linked within domains which constitute the thematic field (Figure 2). The relations between the concepts within a parcel, between the parcels within a domain, and between the domains within the thematic field are represented by BPSs that comply with the structured content. Hence, the ALCCT's macrostructure is a networks-in-the-network ontology that may be converted into a matrixes-in-the-matrix or networks-in-the-matrix ontology. A conceptual ontology that arranges the ALCCT's data provides a natural correspondence between the linguistic and conceptual fields. As Caine and Caine (1994) note, any lexicographic text is not a self-sufficient source of meaning in itself, but rather a form that is meant to activate and foster a definite knowledge structure. That is why processing of information in an active dictionary should be meaningful, or addressed to an inherently meaning-tuned user (Caine & Caine, 1994, pp. 100–101). The ALCCT's macrostructure definitely meets this requirement relevant for an idiographic (onomasiological) dictionary that demonstrates the 'meaning → form' perspective.
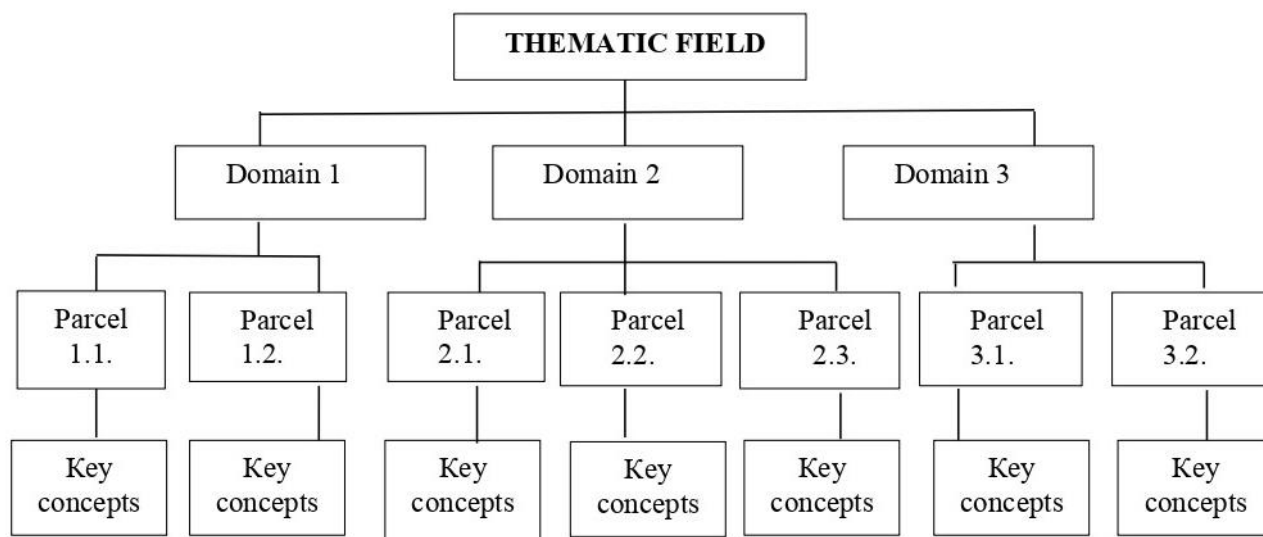

Figure 2. Macrostructure of the ALCCT (Zhabotynska, 2010, p. 81)

The ALCCT's **mediostructure**, iterated throughout the dictionary, arranges information about the key concepts as constituents of parcels in the macrostructure of the thematic field. The key concepts may exhibit variations that are linguistically captured by synonyms. In a synonymous group, the key word (a lexical lemma) that names the key concept, and the most frequent synonym(s) of the key word develop into phrases (phrasal lemmas). The two types of lemmas have their own patterns of representation. *Lexical lemmas* are described with regard to differential senses in the meanings of synonyms. Such senses are distinguished via the properties registered in the BPSs (e.g. contact BPS + MD-mode 'X teaches WHAT + HOW'– *school*: teaches all disciplines equally; *gymnasium*: teaches selective disciplines in-depth). *Phrasal lemmas* are arranged in *phrasal sets* (Figure 3).

The structure of phrasal sets, being constitutional for the lexicographic code of the ALCCT, accounts for its definition as a 'construction-combinatory thesaurus'. Here, word-combinations are considered as instantiations of constructions, or abstract syntactic forms that have their own schematic meaning.
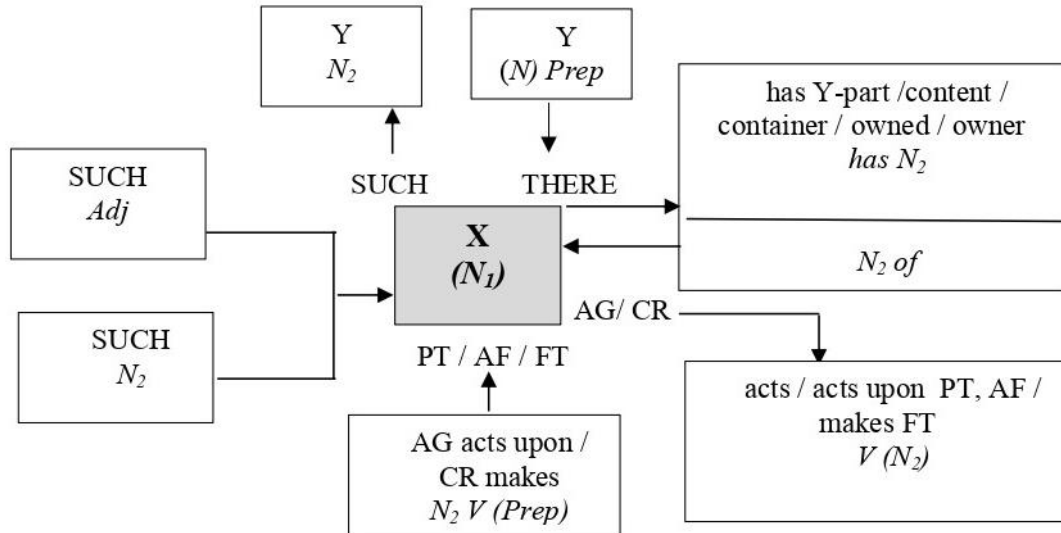
Figure 3. Mediostructure of the ALCCT: arrangement of a phrasal set
(Zhabotynska, 2019, p. 21)

In the ALCCT, a phrasal set has its *thematic tuning*, i.e. the phrases subsumed by a particular construction are thematically grouped. For instance, the phrases which instantiate the qualitative BPS with the key word SCHOOL (SCHOOL is SUCH > SUCH SCHOOL) further split into those where the logical predicate represents (1) the students' age, (2) the taught subjects and the student body, (3) the students' sex, (4) the way of funding, (5) evaluation, etc. (Figure 4).
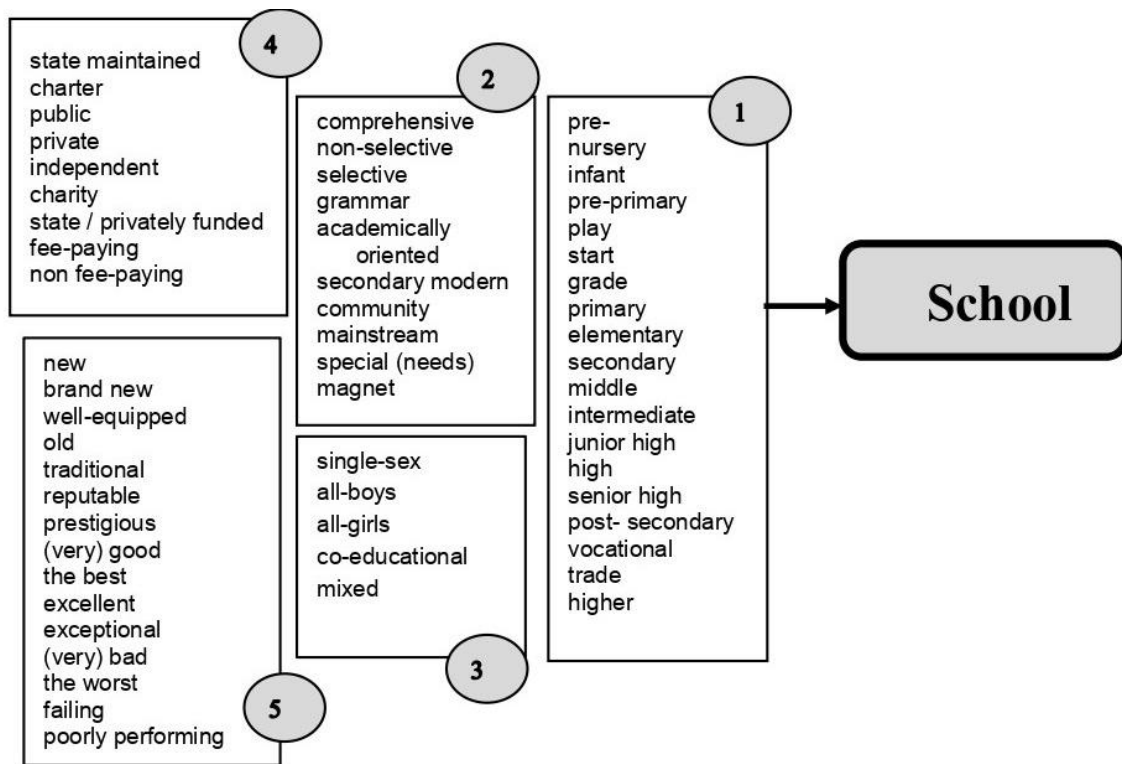

Figure 4. Tuning of phrasal sets in the ALCCT
(a fragment of the phrasal set in Zhabotynska 2019, p. 23)

A tuned phrasal set arranged around a key word creates a *construction-combinatory portrait* of this word. Since such a portrait retains multiple instantiations of the same constructions as schematic form-meaning correspondences, it may foster the required automation of rote memorization and learning based on the pattern recognition and reproduction (see Frost, Siegelman, Narkiss, & Afek, 2013).

The ALCCT′s **microstructure** arranges information in the dictionary entries that describe phrasal lemmas. This description includes (a) translation of the phrase into the native language of dictionary users, (b) examples of sentences with this phrase, and (c) transformations (TRF) of this phrase caused by re-categorization of the respective construction. For example, *reputable school—* TRF: *school (that) has a good reputation, school with a good reputation*. The problem that arises thereby is caused by prominence of the syntactic form inherent for a particular BPS, and prominence of the syntactic form(s) which this BPS has adopted due to re-categorization. An adopted syntactic form may become more entrenched, which is demonstrated by frequency of its use in speech. In this case, an adopted syntactic form becomes the phrasal lemma, and the initial syntactic form is listed among transformations of this lemma. For example, *boy school—*TRF: *school teaches only boys*, *school for boys only*.

Elaboration of the data takes place at the forth, additional level of the ALCCT′s design. This level represents the **overarching structure** that maps upon the three core structures of the lexicographic code and provides additional information relevant for the constituents of these structures. The additional information for lexical lemmas concerns their particular *etymology*. Phrasal lemmas may require *culture-specific comments* (e.g. specific kinds of schools typical of Great Britain and the USA). Phrasal sets are supplied with a *syntactic constructor* retaining the guidelines as to combining phrases into sentences of different types and different degrees of complexity (see the instances of exercises in Zhabotynska, 2015, pp. 50–52; 2019, p. 25; Plakhotniuk, 2015, pp. 59–70). Some domains in the conceptual ontology that structures the entire ALCCT's thematic field may be employed in conceptual metaphors, where they become either the metaphorical target (e.g. SCHOOL is as if X) or metaphorical source (e.g. X is as if SCHOOL). The phraseological linguistic expressions, or idioms, brought under particular conceptual metaphors, are represented in the '*metaphorical repository*' included into the overarching structure of the ALCCT.

The principles of data application reach out to the communicative situations where the ALCCT's data can be used. The dictionary has a system of 'Let's talk' assignments targeted at individual phrasal sets, and clusters of phrasal sets (a) within one and the same parcel, (b) within one and the same domain, and (c) within several domains of the entire conceptual ontology of the theme. The 'Let's talk' assignments propose dictionary users to employ the expressions from one or several phrasal sets in various simulated communicative practices relevant for a thematically focused interaction. The communicative assignments are to be engaged after the user's work with the syntactic constructor. The number of involved phrasal sets depends on complexity of a communicative task. The latter may also extend into the field of creative writing, where the user may employ the ALCCT's metaphorical repository. To devise the communicative assignments, the ALCCT's compiler should consult experts in the field for which the thesaurus is intended.

The principles of compiling the ALCCT suggest multimodality of the resultant text: besides the verbal part, it has visual constituents—conceptual graphics and pictorial illustrations. Provided the text carrier is digital, the ALCCT may be supplied with videos and other Internet resources. *Conceptual graphics* is used to represent the ALCCT's ontology, and to visualise arrangement of phrasal sets which is compatible with the formal arrangement of a sentence (see the examples in (Zhabotynska, 2010, 2015, 2019; Brovchenko, 2011; Radchenko, 2012, 2019)). *Pictures* may illustrate some lexical and phrasal lemmas, especially those that are culture-specific. *Videos* are of particular help for scaffolding the communicative situations. Converging evidence asserts that visual perception improves comprehension and boosts learning. Visualization of the inherent

conceptual properties of linguistic expressions or extralinguistic objects stimulates associative memorization, helps in understanding complex ideas, and increasing the mind's productivity and creativity (Hay, Kinchin, & Lygo-Baker, 2008; Li & Jeong, 2020, p. 2). Therefore, a multimodal format of the ALCCT adds to making it more 'user-friendly'.

6. Concluding discussion

The ALCCT project is consistent with the contemporary theoretical conceptions of natural language generation, and practical approaches to L2 teaching and learning.

The models of natural language generation distinguish several aspects in 'the speaker's blueprint' (Levelt, 1998): (1) conceptual planning (CONCEPTUALIZER)—conceptualizing the event and forming a preverbal message, (2) grammatical encoding (FORMULATOR)—mapping the preverbal message onto lexicon and syntactic structures, and (3) morpho-phonological and phonetic encoding (ARTICULATOR)—formalization and verbalization of a linear message. In (Guhe, 2003), conceptualization is viewed as an incremental process that reduces the complexity of computation (pp. 31, 54) by parallel processing of information that regards: (a) construction / segmentation—mapping what is perceived to concepts from long-term memory, (b) selection of the events that are to be verbalized (macroplanning), (c) linearization—ordering selected events appropriate to the goal of the discourse (macroplanning), and (d) generation / structuring a preverbal message—mapping the conceptual representation that has been handled so far to the semantic content that can interface with the linguistic formulator (microplanning) (p. 31). The information models at this stage involve semantically underspecified "referential nets" of incremental elements. The activation value assigned to each element determines its salience (p. 110).

Thus, the models of natural language generation emphasize the role of conceptualization, or processing the information that is to be manifested with linguistic expressions. Meanwhile, the ways in which this information is processed remain unspecified. In the ALCCT, information processing is effected via constructions that integrate pre-verbal conceptual schemas, or BPSs, with their formal manifestations, or syntactic schemas. Besides, the BPSs are involved in developing a conceptual ontology that arranges the total scope of information in the ALCCT. While constructions contribute to exposure of linguistic information (HOW to say), a conceptual ontology gets beneficial for exposure of non-linguistic information (WHAT to say), which is especially important when the scope of information is sufficient or / and the information is new to the learners. Cognitive studies argue that in the brain / mind the information recall and connection of the working memory with the long-term memory is based on the associative map-like activation (Caine & Caine, 1994, p. 42–44). A conceptual ontology that arranges information in the ALCCT makes the associative activation structured and thus facilitates comprehension and memorization of the intended content. Hence, the theoretical framework employed in compiling the ALCCT may contribute to understanding the nature of relations between conceptual representations and their linguistic manifestations. And conversely, this framework may benefit from the new findings in the field of natural language processing.

The ALCCT, which represents both content and language, agrees with CLIL as one of the most popular contemporary approaches in language teaching and learning. CLIL (Content-and-Language Integrated Learning) emerged in the USA and Europe around 1990s as a continual teaching of curricular content through the medium of a foreign language and foreign language through content (Cenoz, 2015, p. 12; Castellano-Risco, Alejo-González & Piquer-Píriz, p. 6). Although the balance at any one time may vary, the assumption is that overall, a CLIL program will equally focus on content and language and will be referenced to both a foreign language and a content subject curriculum (Kiely, 2011). It is reasonable, therefore, to accept that the language aspect of a CLIL program will also be content driven, in that it will be generated from the specific needs of the particular subject taught and will assist students in better dealing with the requirements

of the subject (Ioannou-Georgiou, 2012, pp. 498–499). Meanwhile, the critical remarks address, on the one hand, insufficiencies in language teaching. They are caused by absence of linguistic expertise of the non-native language instructors, relatively late age of the learners (Dalton-Puffer, 2011, pp. 183–184), and—in general—lack of a systematic, as assumed, "content-driven language aspect of CLIL programs" (Ioannou-Georgiou, 2012, pp. 498–499). On the other hand, the critical remarks also address insufficiencies in content teaching. The scholars note that published materials targeted for CLIL teachers sometimes water down the content subject and treat it in a FL-oriented manner. So, if specific guidelines are not given, CLIL risks to become a time-consuming, ineffective, and frustrating experience (Ioannou-Georgiou, 2012, 497–498).

Presumably, the ALCCT employed in CLIL may become the 'missing link' that balances language-and-content learning and teaching. The ALCCT provides phrasal coverage of particular coherent content which, being rich, demonstrates algorithmic arrangement consonant with the mind's natural logic. The language teachers not quite knowledgeable in a specific professional domain may use the ALCCT as a source of structured subject-related data. The content teachers without an adequate linguistic background may use the ALCCT as a source of linguistic data (lexical, grammatical, and communicative) for teaching language employed in their professional field. If the field has no ready-made ALCCT yet, the teacher can compile it using the principles set out in this article. Such a dictionary may provide linguistic scaffolding of a particular class or the entire topic taught in L2.

ABBREVIATIONS
ALCCT—Active Learner's Construction-Combinatory Thesaurus
BPS—basic propositional schemas
CLIL—Content-and-Language Integrated Learning
SLN—Semantics of Lingual Networks

REFERENCES
Apresjan, Ju. D. (2010). *Prospekt aktivnogo slovarja russkogo jazyka* [Project of the active dictionaly of Russian]. Moskva: Jazyki Slavjanskoj Kul'tury (in Russian).

Boyland, J. T. (2009). Usage-based models of language. In D. Eddington (Ed.), *Experimental and quantitative linguistics* (pp. 351–419). Munich: Lincom.

Brovchenko, I. V. (2011). *Konceptual'naya metafora v klishe angloyazychnogo nauchnogo teksta* [Conceptual metaphor in clichés of the academic text] (Unpublished candidate dissertation). Bohdan Khmelnitsky National University, Cherkasy, Ukraine (in Russian).

Caine, R. N., & Caine, G. (1994). *Making connections: Teaching and the human brain*. Menlo Park, CA: Addison-Wesley.

Castellano-Risco, I., Alejo-González, R., & Piquer-Píriz, A. (2020). The development of receptive vocabulary in CLIL vs EFL: Is the learning context the main variable? *System 91*:101163, 102–263. https://doi.org/10.1016/j.system.2020.102263

Cenoz, J. (2015). Content-based instruction and content and language integrated learning: the same or different? *Language, Culture and Curriculum, 28*(1), 8–24. https://doi.org/10.1080/07908318.2014.1000922

Dalton-Puffer, Chr. (2011). Content-and-Language Integrated Learning: From practice to principles? *Annual Review of Applied Linguistics, 31*, 182–204. Retrieved from https://www.unifg.it/sites/default/files/allegatiparagrafo/20-01-2014/dalton-puffer_content_and_language_integrated_learning_from_practice_to_principles.pdf

Dirven, R., & Verspoor, M. (2004). Cognitive exploration of language and linguistics. 2nd ed. Amsterdam: John Benjamins.

Fair, D. A., Cohen, A. L., Dosenbach, N. U., Church, J. A., Miezin, F. M., Barch, D. M., Raichle, M. E., Petersen, S. E., & Schlaggar, B. L. (2008). The maturing architecture of the

brain's default network. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(10), 4028–4032. Retrieved from https://pubmed.ncbi.nlm.nih.gov/18322013/

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological science, 24*(7), 1243–1252. https://doi.org/10.1177/0956797612472207

Fuentes-Olivera, P. A., & Bergenholtz, H. (2018). Dictionaries for text production. In: P.A. Fuertes-Olivera (Ed.), *The Routledge handbook of lexicography* (pp. 267–283). Abingdon: Routledge.

Goldberg, A. E. (1995). *Constructions. A construction grammar approach to argument structure.* Chicago, IL: The University of Chicago Press.

Guhe, M. (2003). *Incremental conceptualization for language production*. (Unpublished doctoral dissertation). Hamburg University, Hamburg, Germany.

Hay, D., Kinchin, I., & Lygo-Baker, S. (2008). Making learning visible: The role of concept mapping in higher education. *Studies in Higher Education. 33*(3), 295–311. https://doi.org/10.1080/03075070802049251

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Ioannou-Georgiou, S. (2012). Reviewing the puzzle of CLIL. *ELT Journal*, *66,* 495–504. https://doi.org/10.1093/elt/ccs047

Kiely, R. (2011). Understanding CLIL as an innovation. *Studies in Second Language Learning and Teaching, 1*(1). https://doi.org/10.14746/ssllt.2011.1.1.9

Knowles, M. S. (1984). *Andragogy in action. Applying modern principles of adult education*. San Francisco, CA: Jossey Bass.

Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: John Wiley & Son.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Li, P., & Jeong, H. (2020). The social brain of language: grounding second language learning in social interaction. *Science of Learning 5*(1), 1–9. https://doi.org/10.1038/s41539-020-0068-7

Lyngfelt, B., Borin L., Ohara, K., & Torrent, T. T. (Eds.), (2018). *Constructicography: Constructicon development across languages. Constructional Approaches to Language*. Amsterdam: John Benjamins.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *Neuroimage, 63*(1), 240–244. https://doi.org/10.10.1016/j.neuroimage.2012.06.043

MacWhinney, B. A. (2005). Unified model of language development. In G. J. MacWhinney, A. M. Kroll, & B. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 49–67). Oxford: Oxford University Press.

Ostermann, C. (2015). *Cognitive lexicography: A new approach to lexicography making use of cognitive semantics*. Nürnberg: Universität Erlangen.

Ostman, J.-O., & Fried, M. (Eds.). (2004). *Construction grammars. Cognitive grounding and theoretical extensions*. Amsterdam: John Benjamins.

Paradis, M.A. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamis.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology, 1*, 1–15. https://doi.org/10.3389/fpsyg.2010.00227

Plakhotniuk, Ye. (2015). *Dorobok kohnityvnoi linhvistyky u navchanni anhlijskoi movy doroslykh slukhachiv* [Accoplishments of cognitive linguistics in teaching English to adult learners] (Unpublished Master's thesis). Bohdan Khmelnitsky National University, Cherkasy, Ukraine (in Ukrainian).

Plakhotniuk, Ye. (2020a). An active dictionary: user needs and expectations. In *Nova Philologia, 80*. Zaporizhzhia: Zaporizhia National University. (In print).

Plakhotniuk, Ye. (2020b). Blend-dictionaries in a semiotic-cognitive perspective: An active learner's construction-combinatory thesaurus "TRAVELLING". *Science and Education: A New Dimension, 63* (235), 44–50.

Radchenko, O. Yu. (2012). *Konceptual'na metafora v terminosistemi marketingu* [Conceptual metaphor in the terminology of marketing] (Unpublished candidate dissertation). Bohdan Khmelnitsky National University, Cherkasy, Ukraine (in Ukrainian).

Radchenko, O. (2019). Marketing terminology thesaurus: The competition segment. *Lege Artis. Language Yesterday, Today, Tomorrow*, *IV* (1), 133–169. Retrieved from https://lartis.sk/wp-content/uploads/2019/06/Radchenko_Issue-1_2019.pdf

Tarp, S. (2008). *Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner's lexicography*. Tübingen: Max Niemeyer Verlag.

Tarp, S. (2011). Reflections on lexicographical user research. *Lexikos*, *19.* 275–296. https://doi.org/10.5788/19-0-440

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Valipour, V., & Asl, H. D. (2014). Differences between children and adults in foreign language pronunciation and grammatical rules learning. *Indian Journal of Fundamental and Applied Life Sciences, 4* (S3), 195–198. Retrieved from: https://www.cibtech.org/sp.ed/jls/2014/03/JLS-026-S3-036-VALIPOUR-DIFFERENCES-LEARNING.pdf

Vincent-Lamarre, P., Massé, A. B., Lopes, M., Lord, M., Marcotte, O., & Harnad, S. (2016). The latent structure of dictionaries. *Topics in Cognitive Science, 8*(3), 625–659. Retrieved from https://pcl.sitehost.iu.edu/papers/Lamarre2016.pdf

Vorob'eva, O. P. (1993). *Tekstovye kategorii i faktor adresata* [Textual categories and the addressee factor]. Kiev: Vyshcha Shkola (in Russian).

Zhabotynska S. (2010). Principles of building conceptual models for thesaurus dictionaries. *Cognition, communication, discourse, 1,* 75–92. Retrieved from https://drive.google.com/file/d/0BzbZ3piwVKrtVlI5cVlXTmZINW8/view

Zhabotynska, S. (2015). *English clichés of scholarly texts: A thesaurus.* (*Materials for teaching Academic writing*). Kyiv: Ukrainian Association of Cognitive Linguistics and Poetics.

Zhabotinskaja, S. (2018). Generativizm, kognitivizm i semantika lingval'nyh setej [Generative linguistics, cognitive linguistics, and semantics of lingual networks]. In Ye.A. Karpilovska (Ed.), *Doctrina multiplex, veritas una. Uchen'bahato, istyna odna: zbirnyk prac' do yuvileyu Izabelly Rafayilivny Buniyatovoyi* (pp. 99–141). Kyiv: Un-t im. B. Hrinchenka (in Russian).

Zhabotyns'ka, S. (2019). Semantyka linhval'nyh merezh u navchal'nomu kombinatornomu tezaurusi [Semantics of lingual networks in the learner's combinatory dictionary]. *Studia Philologica*, *2,* Kyivs'kyi universytet imeni B. Hrinchenka, 17–27 (in Ukrainian).

Zhabotynska, S. (2020). Dominance of Ukrainian in a bilingual setting: neurocognitive factors. In O. Bilash (Ed.), *Research initiative on democratic reforms in Ukraine (RIDRU)* (pp. 92–115). Edmonton: University of Alberta. Retrieved from https://www.ualberta.ca/canadian-institute-of-ukrainian-studies/centres-and-programs/ulec/ulec-news/ridru_publication.pdf

Zhabotyns'ka, S., & Plakhotniuk, Ye. (2016). Vyvchennya inozemnoyi movy doroslymy: nejrokohnityvnyj rakurs [Learning a foreign language by adults: a neurocognitive aspect]. *Movoznavchyj visnyk, 21*, 7–17 (in Ukrainian).

Zhabotynska, Svitlana—Doctor of Philology, Professor, Bohdan Khmelnitsky National University of Cherkasy, Ukraine (81 Shevchenko Blvd, Cherkasy 18031, Ukraine); e-mail: saz9@ukr.net; ORCID: 0000-0001-9841-6335.

Жаботинська Світлана Анатоліївна – доктор філологічних наук, професор, Черкаський національний університет імені Богдана Хмельницького, Україна (Бульвар Шевченка, 81, м. Черкаси 18031, Україна); e-mail: saz9@ukr.net; ORCID: 0000-0001-9841-6335.

Жаботинская Светлана Анатольевна – доктор филологических наук, профессор, Черкасский национальный университет имени Богдана Хмельницкого, Украина (Бульвар Шевченко, 81, г. Черкассы 18031, Украина); e-mail: saz9@ukr.net; ORCID: 0000-0001-9841-6335.

Plakhotniuk, Yevhenii—PhD student, Kyiv National Linguistic University, Kyiv, Ukraine (73 Velyka Vasylkivska Str., Kyiv 03680, Ukraine); e-mail: eugeneplakhotniuk@gmail.com; ORCID: 0000-0002-4413-1112.

Плахотнюк Євгеній Ігорович – аспірант, Київський національний лінгвістичний університет, Україна (вул. Велика Васильківська 73, м. Київ 03680, Україна); e-mail: eugeneplakhotniuk@gmail.com; ORCID: 0000-0002-4413-1112.

Плахотнюк Евгений Игоревич – аспирант, Киевский национальный лингвистический университет, Киев, Украина (ул. Большая Васильковская, 73, г. Киев 03680, Украина); e-mail: eugeneplakhotniuk@gmail.com; ORCID: 0000-0002-4413-1112.