

УДК 811.111'373
DOI: 10.15421/382105

М. Dobrushyna
М. Ю. Добрушина
М. Ю. Добрушина

Oles Honchar Dnipro National University
Дніпровський національний університет імені Олеся Гончара
Днепропетровский национальный университет имени Олеся Гончара

**THE CORPUS ANALYSIS OF THE NOMINATIVE FIELD OF THE
CONCEPT *LANGUAGE POLICY IN EDUCATION* AS A COMPONENT
OF THE CONCEPT *LANGUAGE POLICY***

**КОРПУСНЕ ДОСЛІДЖЕННЯ НОМІНАТИВНОГО ПОЛЯ КОНЦЕПТУ
LANGUAGE POLICY IN EDUCATION ЯК СКЛАДОВОЇ КОНЦЕПТУ
*LANGUAGE POLICY***

**КОРПУСНОЕ ИССЛЕДОВАНИЕ НОМИНАТИВНОГО ПОЛЯ
КОНЦЕПТА *LANGUAGE POLICY IN EDUCATION*
КАК СОСТАВЛЯЮЩЕЙ КОНЦЕПТА *LANGUAGE POLICY***

This article is dedicated to the corpus research in order to establish the nominative field of the concept. The relevance of the study is the need to examine the nominative field of the concept not only by conducting a semantic analysis of dictionary definitions, but also on the materials of texts created by native speakers. This corpus research is particularly concerned with the determination of the nominative field of the concept by analysing part of the English discourse. The article analyses the concept *LANGUAGE POLICY* in educational discourse on the basis of scientific articles and official texts with the particular focus on the study of the educational language policy. The purpose of this research is to establish discourse information about the concept *LANGUAGE POLICY IN EDUCATION*.

This study has been conducted on the basis of the created corpus of texts and analysed using the computer program AntConc. During this scientific research, 26 texts related to the given topic have been selected and on the basis of these texts the proprietary corpus of texts has been created.

This analysis has been conducted in three stages: during the first stage, the texts involved in this study have been identified. During the second stage, the word list has been developed and analysed. During the third stage, the results of the study have been presented schematically, where the key verbalizers of the concept *LANGUAGE POLICY IN EDUCATION* have been identified, and it has been found that both the key nominations and the close periphery are formed by lexical units that are directly related to education.

Keywords: concept *LANGUAGE POLICY*, corpus linguistics, discourse, educational discourse, nominative field, AntConc.

Здійснено корпусне дослідження з метою встановлення номінативного поля концепту. Актуальність дослідження полягає у необхідності встановлення номінативного поля концепту не лише шляхом проведення семантичного аналізу словникових дефініцій, а й на матеріалі текстів, створених носіями мови. Це корпусне дослідження передбачає визначення номінативного поля концепту шляхом аналізу частини англомовного дискурсу. У статті було проаналізовано концепт *LANGUAGE POLICY* в освітньому дискурсі на матеріалі наукових статей та офіційних текстів з дослідження освітньої мовної політики. Метою цієї розвідки було встановлення дискурсної інформації про досліджуваний концепт.

Це дослідження було проведено на базі створеного корпусу текстів та проаналізовано за допомогою комп'ютерної програми AntConc. Під час проведення цієї наукової

розвідки було виокремлено 26 текстів, пов'язаних із заданою тематикою, та на основі таких текстів було створено власний корпус текстів.

Цей аналіз було проведено в три етапи: на першому етапі було визначено тексти, які було залучено до цього дослідження. На другому етапі було встановлено та проаналізовано «список слів». На третьому етапі результати дослідження було проілюстровано у схематичному вигляді, де було визначено ключові номінанти концепту LANGUAGE POLICY IN EDUCATION та встановлено, що ключові номінації, так само як і ближню периферію, утворюють лексичні одиниці, які безпосередньо пов'язані зі сферою освіти.

Ключові слова: концепт LANGUAGE POLICY, корпусна лінгвістика, дискурс, освітній дискурс, номінативне поле, AntConc.

Осуществлено корпусное исследование с целью установления номинативного поля концепта. Актуальность исследования заключается в необходимости установления номинативного поля концепта не только путем проведения семантического анализа словарных дефиниций, но и на материале текстов, созданных носителями языка. Это корпусное исследование предполагает определение номинативного поля концепта путем анализа части англоязычного дискурса. В статье был проанализирован концепт LANGUAGE POLICY в образовательном дискурсе на материале научных статей и официальных текстов, посвященных исследованию образовательной языковой политики. Целью этой научной работы было установление дискурсной информации об исследуемом концепте.

Это исследование было проведено на базе созданного корпуса текстов и проанализировано с помощью компьютерной программы AntConc. Во время проведения этой научной работы было выделено 26 текстов, связанных с заданной тематикой, и на основе таких текстов был создан собственный корпус текстов.

Этот анализ был провен в три этапа: на первом этапе были определены тексты, которые были задействованы в этом исследовании. На втором этапе был установлен и проанализирован «список слов»; на третьем этапе результаты исследования были предоставлены в схематическом виде, где были определены ключевые номинанты концепта LANGUAGE POLICY IN EDUCATION и установлено, что ключевые номинации, так же как и ближнюю периферию, образуют лексические единицы, которые непосредственно связаны со сферой образования.

Ключевые слова: концепт LANGUAGE POLICY, корпусная лингвистика, дискурс, образовательный дискурс, номинативное поле, AntConc.

Nowadays corpus linguistics and the construction of text corpora play a pivotal role in conducting linguistic research. The use of electronic corpora (both those which are replenished over a long period of time, and those which are created by the researcher to study a phenomenon) simplified the procedure for conducting research in linguistics, which reduced the time of such research and reduced the costs of achieving one or another goal of the corpus linguistic analysis. The relevance of the analysis of language data with the help of corpus research is gaining momentum today, as attracting a large amount of material for multifaceted analysis helps to create a real picture of the functioning of all units of language in speech [7; p. 9]. The need for analysis and generalization of the results of the study of language material is one of the key factors in the corpus study of the text [3; p. 83].

Corpus linguistics is a branch of computational linguistics which deals with the development of general principles of construction and use of linguistic corpora (text corpora) with the use of computer technology [9; p. 7]. According to M.A.K. Halliday, the study of language on the basis of discourse is important for corpus linguistics. Corpus linguistics sees language as a social phenomenon [5; p. 97–98]. In the study “Lexicology and Corpus Linguistics” it is explained that the English discourse is a set of texts produced over the centuries by members of the English discourse community [5, p. 100]. It is noted that the extremely large amount of discourse makes comprehensive research impossible. Even if the study of discourse is limited to preserved texts, this

discourse is too large to make it the object of a research. It is said that it will never be possible to study all extant texts [5; p. 100].

This research has been carried out as part of corpus linguistics and provides for the establishment of the nominative field of the concept LANGUAGE POLICY in the educational discourse on the basis of scientific articles on the educational language policy. It should be stated that in the given research the concept LANGUAGE POLICY is referred to linguistics and has been analysed as a language phenomenon.

The given research is part of establishing the nominative field of the concept LANGUAGE POLICY in linguocognitive and linguodiscursive aspects, because the nominative field is reproduced not only by semantic analysis of dictionary definitions, but also on the materials of texts created by native speakers. The key method of the given analysis is the corpus analysis. This research has been conducted on the basis of the corpus of texts which has been exclusively created and analysed using the computer program AntConc. The object of the research is a corpus of scientific texts and articles on language policy in education, the subject is the concept LANGUAGE POLICY IN EDUCATION, which is used in this type of articles and texts. The aim of the article is to organize and establish key verbalizers of the concept LANGUAGE POLICY in scientific articles dedicated to the study of the educational language policy.

The establishment of the nominative field of the concept LANGUAGE POLICY IN EDUCATION on the basis of scientific articles and official texts dedicated to the study of educational language policy has been carried out in several stages.

During the first stage, the corpus of articles concerning the study of the educational language policy has been developed. Thus, 26 articles that directly related to the analysis of the principles of language policy in the European Union as a whole, some European Union countries and the United States have been selected. Also the attention has been drawn to some official documents that had been implemented in the European Union and the United States of America in the framework of language policy in the field of education, and which are directly related to the given topic, such as Position Statement on Language Policy in Higher Education in Europe, Language Policy of the European Schools. The selection parameters of the materials are not strictly limited: the establishment of the nominative field of the concept under study can be actualized both with the help of texts published in 2020 [4] and texts which have been the subject of research for a long time [6]; the key requirement of the study is to establish verbalizers of the concept with the help of the following parts of speech: noun, verb and adjective. Pronouns, numerals and adverbs have not been analysed because in the context of the research the given parts of speech do not provide a meaningful component.

The rationale for constructing the proprietary corpus of texts is due to the fact that not all texts are in known corpora. In the preceding paragraphs it has been mentioned that 26 texts that are freely available on the Internet have been selected. All texts have been reformatted to *.txt. format.

During the second stage of the research the attention has been paid to the WORD LIST, which is inherent in all given texts. During the establishment of the list of significant words, we have identified 16104 lexical units in total, which occurred in all texts 350603 times. In this study, we have focused on the first thousand lexical units that have been identified with the help of the program AntConc. AntConc is a freeware corpus analysis toolkit for text analysis and concordancing designed by Dr. Laurence Anthony, a Professor in the Faculty of Science and Engineering at Waseda University,

Japan [1]. This concordance is aimed to be the primary analytical tool used by vocabulary researchers [2; p. 79].

The frequency of use of the first lexical unit, the definite article *the*, which has not been taken into account because it does not meet the established requirements of the given research, is 22733 times, the frequency of use of the last, thousandth lexical unit *improving* is 44 times. The first significant lexical unit is the lexeme *language* with a frequency of 5475.

In the following section we have discussed the word list in more detail.

Thus, there have been examined the following nouns: *language* (5475+*languages* 1941), *education* (2521), *policy* 1503+*policies* 408), *children* (1408 + *child* (417)), *learning* (992), *school* (869+ *schools* 659), *students* (798), *reform* (770+*reforms* 198), *teachers* (675+ *teacher* 494), *parents* (611), *development* (531), *research* (520), *system* (502+*systems* 156), *time* (476), *level* (446), *world* (428), *curriculum* (419), *professionals* (393), *skills* (393), *support* (375), *family* (343+*families* 184)), *university* (336), *years* (325+*year* 225), *people* (322), *France* (321), *countries* (310+*country* 234), *Europe* (301), *part* (298), *implementation* (290), *multilingualism* (289), *practices* (282+*practice* 273), *diversity* (281), *context* (272), *knowledge* (272), *example* (263), *change* (253), *approach* (252), *ECEC* (Early Childhood Education and Care (251), *century* (243), *student* (241), *data* (239), *work* (237), *role* (235), *goals* (233), *framework* (232), *number* (228), *minority* (227), *communication* (226), *needs* (221), *mother* (220), *activities* (218), *levels* (216), *instruction* (215), *speaking* (215), *capacity* (213), *model* (213), *state* (208+*states* 202), *home* (205), *case* (203), *place* (202), *bilingualism* (201), *parent* (201), *focus* (200), *medium* (200), *terms* (200), *government* (199), *society* (198), *section* (197), *rights* (196), *programs* (194+*program* 165), *setting* (194), *staff* (193), *access* (192), *human* (187), *understanding* (186), *speakers* (185), *city* (183), *quality* (183), *culture* (182), *literacy* (182), *content* (178), *planning* (177), *assessment* (176), *term* (176), *community* (175), *pupils* (175), *training* (174), *fact* (172), *information* (172), *council* (169), *competence* (163+*competencies* 140), *members* (162), *German* (161), *means* (160), *analysis* (159), *institutions* (159), *ministry* (159), *resources* (158), *association* (155), *studies* (155), *life* (154), *classes*

(153), *identity* (152), *project* (147), *report* (146), *changes* (144), *groups* (144), *point* (144), *org* (143), *universities* (142), etc.

In the given research there have been examined the following adjectives: *English* (1735), *French* (1103), *Bilingual* (638), *professional* (601), *linguistic* (557), *European* (545), *new* (533), *educational* (478), *national* (478), *different* (471), *multilingual* (383), *cultural* (374), *foreign* (330), *international* (326), *political* (313), *secondary* (309), *social* (307), *primary* (284), *important* (269), *academic* (232), *economic* (213), *official* (210), *global* (208), *general* (191), *public* (183), *local* (182), *regional* (168), *specific* (168), *young* (160), *own* (156), *institutional* (149), *major* (146), *possible* (142).

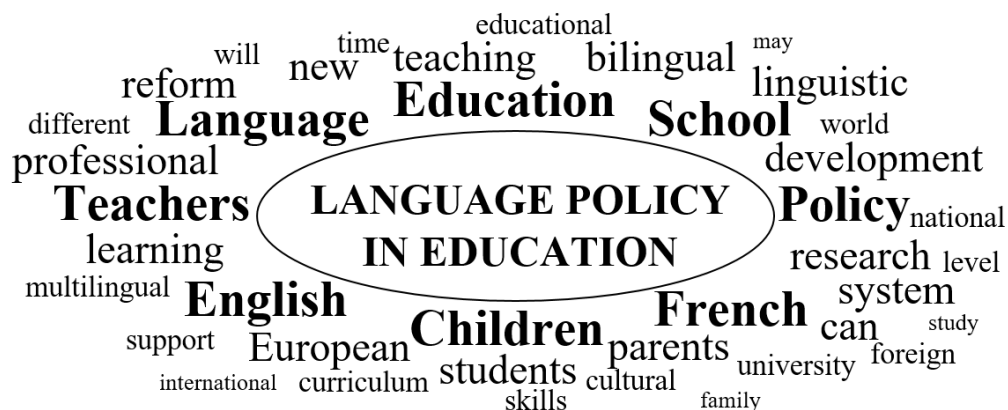
The given research has provided us with the following verbs: *study* (331), *provide* (183), *learn* (178), *develop* (162+*developed* 152), *taught* (158), *understand* (150).

As a result of the analysis of the first thousand words, it has been determined that the number of nouns, adjectives and verbs is 789 out of 1000 (78.9%). It is worth noting that during the corpus analysis it has been found that some lexical units are cognate words, or may have different grammatical forms, such as: *institution* – *institutions*, *system* – *systems*, *teach* – *taught*, etc.

During the third stage of the research the nominative field of the concept LANGUAGE POLICY IN EDUCATION on the basis of the abovementioned verbalizers

has been established. The field structure has been build (Fig. 1). The field structure is designed on the basis of the brightness of the corresponding cognitive features in the structure of the concept. Cognitive features have been distributed across the field zones according to the degree of brightness. The result of the field stratification of the concept is the representation of the structure of the concept as a field structure – the core, the close / far periphery, and the interpretation field [8].

Fig.1. Schematic reproduction of the nominative field of the concept LANGUAGE POLICY IN EDUCATION on the material of texts.



Thus, in Fig. 1 the following schematic reproduction has been proposed: all the given nominations of the concept are located in close proximity to each other due to the fact that only 40 out of 789 nominations have been depicted in this field structure, which form the close periphery of the concept. The frequency of use of nominations has also been represented: those nominations that are used more than 1000 times – are the key nominations of the concept and they compose the nominative field of the concept LANGUAGE POLICY IN EDUCATION (Language (7416), Education (2521), Policy (1911), Children (two grammatical forms of the noun have been involved: Child / Children (1825), English (1735), School (1528), Teachers (two grammatical forms have been involved together with the singular noun teacher (1164), French (1103)).

To conclude, we have determined that when conducting corpus research, it is necessary to take into account what part of the discourse falls under the research, in order to provide discursive information about the lexical units that are the basis of the research. The aim of the article that implies the organization and establishment the key verbalizers of the concept LANGUAGE POLICY IN EDUCATION has been achieved. It should be noted that according to the conditions of the study, the given research has been focused on the lexical units that belong to three parts of speech and which have been analysed with the help of the concordance AntConc.

Thus, it has been found that among the nominations of the concept LANGUAGE POLICY IN EDUCATION a significant number are those that belong to the field of education: *Education, School, Teachers*. Since this study aimed to identify verbalizers of the concept LANGUAGE POLICY IN EDUCATION, it should be emphasized that commonly used nominations inherent in the concept LANGUAGE POLICY, such as *multilingual, international, European, cultural*, etc., are an integral part of verbalisation of the concept in various aspects of the research.

It should be emphasized that although the concept LANGUAGE POLICY IN EDUCATION has been studied exclusively on the basis of scientific articles and official texts dedicated to the study of the educational language policy, the vast majority of its nominations are general verbalizers for the concept LANGUAGE POLICY. With regard to the key nominations of the concept LANGUAGE POLICY IN EDUCATION, it should be noted that both the key nominations and the close periphery are formed by lexical units that are directly related to education: *teachers, children, curriculum, students*, etc.

The prospect of further research of this material involves the systematization of data for a detailed view of the use of the concept LANGUAGE POLICY.

References

1. AntConc//Software. Retrieved from <https://www.laurenceanthony.net/software.html>
2. Anthony, L. (2017). Corpus Linguistics and Vocabulary: A Commentary on Four Studies. *Vocabulary Learning and Instruction*, 6 (2), 79–87. DOI: 10.7820/vli.v06.2.Anthony
3. Demska, O. (2011). *Tekstovyi korpus: ideia inshoi formy*. Kyiv. 342 p.
4. Fernando, M. R. (2021). *Implementing Deeper Learning and 21st Education Reforms. Building an Education Renaissance After a Global Pandemic*. Springer International Publishing. DOI 10.1007/978-3-030-57039-2
5. Halliday, M. A. K. (ed.). (2004). *Lexicology and corpus linguistics*. Bloomsbury publishing.
6. Helot, Christine. (2003). Language Policy and the Ideology of Bilingual Education in France. *Language Policy*. 2. 255–277.
7. Kovbasiuk, L. A. (2017). *Korpusna lingvistyka ta hermanistyka: teoretychni zasady i perspektyvy*. Naukovi zapysky Nizhynskoho derzhavnoho universytetu im. Mykoly Hoholia. Filolohichni nauky 1, 9–14.
8. Popova, Z. D., Sternin, I. A. (2007). *Kognitivnaya lingvistika*. Moskva: AST: Vostok – Zapad.
9. Zakharov, V. P., Bogdanova, S. Yu. (2011). *Korpusnaya lingvistika: uchebnik dlya studentov humanitarny'kh vuzov*. Irkutsk: IGLU, 161 s.

Надійшла до редакції 23.10.2020