

Central zone of the field of antonyms in the English language

O. L. Lytvyn*

Institute of Foreign Philology, Lesya Ukrainka Eastern European National University, Lutsk, Ukraine

*Corresponding author. E-mail: oksanalytvyn8@gmail.com

Paper received 07.07.15; Accepted for publication 24.07.15.

Abstract. In this article an attempt is made to define the central zone of the field of antonyms in the English language by comparing and contrasting the data obtained by the process of analyzing illustrative material used in linguistic works written in English, and information about the frequency of these units in language obtained from a frequency dictionary – «A Frequency Dictionary of Contemporary American English». The results of the semantic and functional analysis, as well as quantitative calculations enabled us to arrive at the conclusion that the central zone of the field of antonyms in the English language is formed mainly by morphologically unrelated (absolute) adjectives and verbs which illustrate gradable and complementary antonymy, namely about 60 pairs of antonyms.

Keywords: *linguistic category, structure of the field, central zone of the field, antonymy, frequency*

The study of the structure, the content and the correlation between different fields in the lexical subsystem of language occupies one of the central places in modern linguistics (Karaulov, Abramov, Shchur, Denysenko etc.). However, it could be argued that even now, insufficient attention is being devoted to the structure and content of the fields consisting of abstract elements which are signs by nature – language units that are connected by a systemic relationship. These units form certain linguistic categories, particularly synonymy, antonymy, converseness, and hyper-hyponymy. Analysis of these fields makes it possible to obtain a clearer picture of the forms and the means for structuring knowledge about language in the consciousness of linguists, and that helps to expand the knowledge regarding the categorization and conceptualization of real-life phenomena. This is what forms the basis of **the current relevance** of our research.

In this article an attempt is made to define the central zone of the field of antonyms in English by comparing and contrasting the data obtained through the analysis of the illustrative material used in linguistic works written in English, cf [1], and the information about the frequency of these units in language. We assume that lexical units used by the scholars as representative examples of the category under investigation have a relatively high frequency. **The purpose** of the present research is to test our hypothesis. Semantic and functional, as well as statistical analyses constitute **the main methods** of our research.

Linguists regard antonymy as one of the main types of systemic semantic relations among language units. Researchers emphasize that the relationship of contrast is fundamental for human thinking [4; 6], and antonymy occupies a central place among all types of lexical semantic relations since it plays an important role in organizing the lexicon [2]. British scholar M. Lynne Murphy calls antonymy “the archetypical lexical semantic relation” [7, p. 169]. Notwithstanding the diversity in their approaches to the understanding of antonymy (both in its narrower and its wider senses) scholars agree that antonymy expresses conceptual opposition as realized in language.

A selection of examples provided by the authors of thirteen monographs, textbooks and linguistic encyclopedias written in English serves as **the material** for our research. In the linguistic works examined we found a total of 423 pairs of antonyms which were used as illustrative material, including 224 different pairs, not taking into account

cases of their recurrence. The analysis of the selected examples enabled us to draw certain conclusions. More than a third of all the pairs (35.46%) turned out to have been used only in one of the linguistic sources. On the basis of our calculations, we can affirm that in the opinion of the linguists, adjectives (59.1%) and verbs (26%) provide the best representation of the category of antonymy. Formal structural analysis demonstrated the numerical preponderance of morphologically unrelated (absolute) antonyms (85.34%) over derived ones.

In order to evaluate the examples given in the linguistic works, we assigned a rank in the form of a simple fraction to each pair of antonyms where the number of occurrences of a certain antonymous pair in the form of an illustrative example is indicated in the numerator, and the total number of linguistic sources in the denominator. Our calculations show that one pair of complementary antonyms, *dead – alive*, occurred the most frequently among all the pairs, being present in all thirteen sources. Thus this pair has a ranking designation of 13/13. The gradable antonym pair *hot – cold* is in the second place: it was mentioned in twelve linguistic works and therefore we assigned a rank of 12/13 to this pair. The third place is occupied by the gradable antonym pair *good – bad*, which was used in nine of the sources as an example and consequently received a rank of 9/13. A ranking of 8/13 was allotted to the pairs *male – female* and *tall – short*, which were used in eight linguistic works. Two pairs mentioned in seven sources have a ranking of 7/13 – the antonymous pairs *long – short* and *true – false*. In six of the sources, the three pairs *big – small*, *up – down*, and *old – young* reappear. Therefore they were given a ranking of 6/13. The next group, comprising twelve pairs with a ranking of 5/13, are the lexeme pairs *clean – dirty*, *come – go*, *high – low*, *large – small*, *love – hate*, *married – single*, *north – south*, *open – shut*, *pass – fail*, *thick – thin*, *warm – cool*, and *wide – narrow*. Seven pairs were used in four sources, *beautiful – ugly*, *big – little*, *black – white*, *deep – shallow*, *fast – slow*, *honest – dishonest*, and *rich – poor*, so they received a ranking of 4/13. A ranking of 3/13 was given to thirteen of the pairs: *above – below*, *asleep – awake*, *clever – stupid*, *east – west*, *happy – sad*, *happy – unhappy*, *left – right*, *win – lose*, *man – woman*, *old – new*, *on – off*, *smooth – rough*, and *sweet – sour*. Thirty-two different pairs were found to have occurred in two of the sources examined and thus have a ranking of 2/13. They include *advance – retreat*, *back – front*, *easy – difficult*, *fresh – stale*, *here – there*, *like – dislike*, *marry – divorce*, and *wet – dry*. According to our

calculations the remainder of the illustrative examples (150 different antonymous pairs) were mentioned only in one linguistic source (*complex – simple, joy – sadness, lengthen – shorten, pretty – plain, useful – useless, widen – narrow* etc.).

Thus almost 65% of the total number of antonymous pairs occurred in more than one source as illustrative examples, including 74 different pairs (not taking into account instances of their recurrence). This suggests that two or more linguists consider these pairs to be sufficiently suitable representatives of the relation referred to as lexical antonymy. This fact makes it possible for us to conclude that, on the basis of the data obtained from the linguistic sources, pairs with a ranking of 2/13 or higher constitute the central zone of the field of antonyms. In other words, the central zone is formed by 74 different antonymous pairs with rankings between 2/13 and 13/13, among which two semantic types are the most prevalent: gradable (45.9%) and complementary (28.4%) antonyms. For the most part, gradable antonyms are adjectives, whereas complementary antonyms are mainly adjectives and verbs. The percentage of reverse and antipodal antonyms is considerably lower (16.2% and 9.5% respectively). It is to be noted that within the total selection of 74 pairs, adjectives (60.8%) and verbs (24.3%) predominate. Nouns (6.8%), adverbs (5.4%) and prepositions (2.7%) occur much less frequently. Morphologically unrelated antonymous pairs (86.5%) exceed the derived ones, with the ranking of all derived antonym pairs ranging between 1/13 and 3/13; only one pair (*honest – dishonest*) has a ranking of 4/13.

It was necessary to compare the results of the analyses carried out on the illustrative material in the linguistic works with existing data regarding the frequency of these units in language. This information about the frequency of the lexical units which make up the selected antonymous pairs was obtained from “A Frequency Dictionary of Contemporary American English” which provides a list of the 5,000 most frequently used words in the English language [5]. This dictionary is based on the data from a 385-million-word corpus (by 2008) “The Corpus of Contemporary American English” (COCA). COCA is the largest balanced corpus of American English that is publicly available [3].

The main index in this dictionary is a rank-ordered listing of the top 5,000 words (lemma) in English, starting with the most frequent word (the definite article *the*) which received rank frequency “1” in the dictionary. The following information is given in each entry: rank frequency (from 1 to 5000); lemma; part of speech; collocates, grouped by part of speech and ordered according to frequency. In addition, raw frequency (the total number of tokens in the corpus for the lemma) and dispersion (the index which shows the degree of word distribution in all of the registers in the entire corpus, from 0.00 to 1.00) are given. The final score in the dictionary, or the rank frequency, was calculated by multiplying raw frequency by dispersion.

In order to identify the frequency of the lexical units found in the linguistic sources examined, we assigned each of them to a specific group based on its rank in the dictionary. For this we divided all the lemmas in the dictionary into five groups – A, B, C, D and E. If a word belongs to

the first thousand units in the dictionary in terms of its ranking we have placed it in group A; for example, the adjective *good* has a rank of 111 in the dictionary, therefore it belongs in group A; the verb *repair*, with a rank of 3998, belongs in group D. Words that are not included in the register (the 5,000 most frequently used words), and which thus have a final score above 5,000, were placed in group F. By way of example, the adjectives *unmarried, rude* and *evergreen*, the noun *clergy*, and the verb *embark* are not presented in the dictionary; consequently they are included in group F.

It is thus possible to designate each pair of antonyms using the two capital letters that correspond to the classification the two elements which compose it; for example, the antonymous pair *wide – narrow* can be represented using the combination BC, based on the rankings of its components in the dictionary (*wide – 1181; narrow – 2014*). Among 224 different pairs of antonyms there are pairs the components of which belong to the same group. They constitute almost 40% (89 different pairs) of the examples. Among them are the pairs *old – young* (AA), *converge – diverge* (FF), *light – dark* (BB), *maximum – minimum* (DD), and *shout – whisper* (CC). It should be noted that pairs of the type AA constitute a little over a fifth of the examples (20.54% – 46 pairs). This means that the lexical units which form these pairs belong to the first thousand most frequently used words in the English language (involving not only the raw frequency in the corpus, but also the dispersion index, which shows the equability of word distribution in all of the registers in the entire corpus). The antonymous pairs of the type FF, the components of which were not included in the dictionary, constitute slightly less (13.39% – 30 pairs). The percentage of other examples is considerably lower (BB – 3.57%; CC – 1.34%; DD – 0.89%;). No pairs of the type EE were found at all.

Pairs within which the elements belong to different groups constitute about 60% of the examples (135 pairs). For example, the antonymous pair *rich – poor* is assigned to the category designated BA, since its elements, based on their rankings in the dictionary, belong to different groups (*rich – 1090; poor – 718*).

It is clearly evident that lexical units composing the selected pairs of antonyms have differing degrees of divergence in terms of their rankings in the dictionary. Sometimes there is a very significant divergence between the rankings of the elements of the antonymous pairs, such as *deep – shallow* (BE; 1179 – 4178), *arrive – depart* (AE; 813 – 4862), and *towards – away* (CA; 2067 – 742). With this fact in mind, in order to evaluate the significance of the antonymous pairs and to rank the pairs, but not their individual components, in terms of frequency, we have introduced an index referred to as the index of divergence by ranking. The index of divergence for a pair of antonyms is calculated as the arithmetic mean of the rankings which the elements of a given pair have in the dictionary, and is represented using the group designations A, B, C, D, and E; a subscript letter _p (for the English word ‘pair’) is added to it, indicating that this designation using the subscript letter _p, as in A_p, B_p, C_p, D_p, or E_p, denotes a pair, but not a separate lexical unit. Let us consider one example: we represent the antonymic pair *healthy – ill* using the designation BD, based on the rankings which its elements have in the dictionary (*healthy – 1515; ill – 3016*). In order to compare

this antonymous pair with the other pairs we define the index of divergence of its elements in terms of their ranking by calculating the arithmetic mean between 1515 and 3016, which gives us a value of 2265.5. Thus the index of divergence for the pair *healthy – ill* is $C_p - 2265.5$.

On the basis of the index of divergence the antonymous pairs were divided into five groups. The first and most numerous group constitutes almost a quarter of all the pairs (53 different pairs – 23.66%). These are the pairs with an index of divergence of A_p . Among these, two subgroups can be distinguished. The first subgroup contains the pairs of the type AA (46 pairs), such as *hot – cold*, *big – small*, *up – down*, *east – west*, *remember – forget*, and *after – before*. The second subgroup includes the pairs of the type AB (4 pairs) and BA (3 pairs), such as *love – hate*, *rich – poor*, *all – none*, *slowly – quickly*.

Antonymous pairs with the index of divergence B_p constitute the second group, including 33 different pairs (14.73%). These are pairs of the types AB and BA, BB, AC and CA, BC and AD, such as *dead – alive*, *male – female*, *married – single*, *happy – sad*, *raise – lower*, *include – exclude*, and *towards – away*.

The third group, which constitute 12.05% of all the examples, is formed of the 27 different pairs of antonyms that have the index of divergence C_p . Pairs of the types BC and CB, BD, BE and EB, AE and EA, CC, AD and DC are in this group; among them are *brilliant – stupid*, *defend – submit*, *frequently – rarely*, *arrive – depart*, *fat – thin*, *increase – decrease*, *marry – divorce*, *near – far*, *happy – unhappy*, *beautiful – ugly*, *deep – shallow*, and *clean – dirty*.

Antonymous pairs in the fourth group are less numerous; it contains 13 different pairs (5.8%) with the index of divergence D_p . Pairs of the type DE and ED, CD and DC, DD, CE and BE are in this category, including pairs such as *asleep – awake*, *criticize – praise*, *damage – repair*, *formal – informal*, *friendly – hostile*, *full – empty*, *interesting – boring*, and *permit – forbid*.

According to our analysis, no pairs with an index of divergence E_p were found. The fifth group contains pairs to which the index of divergence F_p can be conventionally assigned. These are antonymous pairs of which neither component was present in the frequency dictionary, and which have a final score higher than 5,000. These pairs are represented using the designation FF. The percentage which they represent within the total body of the examples is not high – 13.4% (30 different pairs). Among them are *inflate – deflate*, *polite – rude*, *adore – despise*, *darken – lighten*, *deciduous – evergreen*, *ferocious – meek*, *forwards – backwards*, *knit – unravel*.

One anomalous group takes in antonymic pairs whose index of divergence is impossible to define even conventionally in terms of the fact that one of its elements was not included in the register of the frequency dictionary. This takes in the pairs of the type AF and FA, CF and FC, BF and FB, FE and EF, FD and DF. Thus their index of divergence might range somewhere within the groups C_p , D_p , or E_p , or it might exceed 5000. This group constitutes slightly less than a third of the antonymous pairs (30.36% – 68 different pairs), including pairs such as *accept – turn down*, *attractive – unattractive*, *joy – sadness*, *maximize – minimize*, *sleep – insomnia*, *ascend – descend*, *cruel – kind*, *clever – stupid*, *left – right*, and *honest – dishonest*.

In our opinion, the antonymous pairs in which the components have a low degree of divergence in terms of their rankings in the dictionary should be considered individually. It is also necessary to calculate the difference between the rankings of the elements in the antonymous pair that we can define in the frequency dictionary. This provides the relative frequency of the antonyms within the antonymous pair. Among the examples, 22 such pairs were identified (almost 10% of the examples), in which the difference between the rankings of the components does not exceed 100.

The pair *girl – boy* has the smallest difference between the rankings of its components. The difference between the rankings of these lexical units in the dictionary is 1 (*girl* – 381; *boy* – 382). The three next highest positions are occupied by the pairs *large – small* (*large* – 225; *small* – 207), *man – woman* (*man* – 95; *woman* – 114), and *here – there* (*here* – 96; *there* – 117). The degree of divergence between the rankings of their elements is 18, 19, and 21 respectively. There is a much greater disparity between the rankings of the components of the remainder of the antonymous pairs.

All 22 of the antonymous pairs identified, in which the elements have a low degree of divergence between their rankings, belong to the type AA. Among them gradual antonyms (11 pairs) predominate, and there are significantly fewer complementary (5 pairs), antipodal (4 pairs), and reverse (2 pairs) antonyms.

Our analysis of the data relating to the frequency of the language units under study enabled us to draw certain conclusions. About 13% of the antonymous pairs were not included in the frequency dictionary, half of them being derived. The pairs in which one of the components was not included in the register of the dictionary (more than a third of them being derived antonyms) constitute about 30% of the examples.

It is necessary to note that almost 40% of the antonymous pairs (86 different pairs) have an index of divergence A_p or B_p . In our opinion, these pairs, based on the data from the dictionary, form the central zone of the field of antonyms. Gradable (41.9%) and complementary (27.9%) antonyms predominate among them. The majority of the gradable antonyms are adjectives; complementary antonyms are mainly composed of adjectives and verbs, and less frequently by nouns. The percentage of reverse and antipodal antonyms is 17.4% and 8.1% respectively. It is also necessary to point out that adjectives constitute almost half (47.7%) and verbs about a third (30.2%) of the total group of 86 pairs. Nouns (9.3%), adverbs (5.7%), prepositions (4.7%), pronouns, and numerals (1.2% each) occur much less frequently. Morphologically unrelated antonyms (96.5%) exceed derived ones, the former generally having a higher rank in the frequency dictionary than the latter. Almost a third of the derived antonymous pairs (30.6% – 15 different pairs) were not included in the frequency dictionary. In addition, the analysis of the frequency of the components in the derived antonymous pairs made it possible for us to identify one notable feature: an element with a negative prefix or suffix has a much lower ranking in the dictionary or is not included in its register at all.

Conclusions. On the basis of the study and comparison of the data obtained in the course of analyzing the illustrative material in the linguistics sources as well as the in-

formation regarding the frequency of the same language units in the frequency dictionary (a total of 224 different pairs), certain conclusions can be drawn.

Having compared the indices of divergence of the pairs in the dictionary and their ranks in linguistic works, we identified certain correlations. Practically all antonymous pairs with the highest levels in the rankings, from 13/13 to 5/13, have an index of divergence A_p and/or B_p (with the exception of one pair, *clean – dirty*, with a ranking of 5/13, which was designated using the index of divergence C_p). The pair of complementary antonyms *dead – alive* (AB) that was found to have occurred the most frequently and received the highest ranking, 13/13, has an index of divergence B_p (1136). The pair of gradable antonyms in the second highest position, *hot – cold* (AA), with a ranking of 12/13, is designated with the index of divergence A_p (813.5). The pair *good – bad* (AA), in the third-highest position with a ranking of 9/13, has an index of divergence A_p (200.5). The next two pairs *male – female* (BB) and *tall – short* (BA) with rank 8/13 received the index of divergence B_p (1504.5 and 1097 respectively). The two pairs of lexemes which occur next in the listings, *long – short* (AA) and *true – false* (AC), with a ranking of 7/13, have an index of divergence A_p (445) and B_p (1517) respectively. All three pairs with the ranking of 6/13 (*big – small*, *old – young*, and *up – down*), based on their index of divergence, are in the group A_p . These pairs fall into the type AA and occupy rather high positions in terms of the index of divergence (186; 198 and 85.5 respectively). Among the twelve antonymous pairs with a ranking of 5/13, six pairs belong to the group A_p , and four pairs to the group B_p ; one pair has an index of divergence C_p ; and one pair of the type AF also belongs here. In addition, the index of divergence A_p and B_p was also used in designating several pairs with rankings between 4/13 and 1/13.

Pairs with a ranking of 1/13 constitute more than a half of the pairs with the index of divergence C_p (15 pairs). There are significantly fewer pairs with a ranking of 2/13 in this group (only six of them). The index of divergence C_p was also assigned to two pairs with a ranking of 3/13 (*happy – unhappy* and *smooth – rough*), three pairs with a ranking of 4/13 (*beautiful – ugly*, *deep – shallow*, and *fast – slow*) and one pair with a ranking of 5/13 (*clean – dirty*).

The complete group of pairs having an index of divergence D_p , with the exception of the pair *asleep – awake* with a ranking of 3/13, is made up of pairs that received a ranking of 1/13, such as *criticize – praise*, *damage – repair*, *formal – informal*, *friendly – hostile*.

It is necessary to note the fact that the antonymous pairs

with the lowest rankings in the linguistic works – two pairs with a ranking of 2/13, and the other 28 pairs with a ranking of 1/13 – belong to the group of antonyms with the index of divergence F_p (where neither component of the pair was included in the dictionary).

The pairs with a ranking of 1/13 in linguistic works constitute almost 78% of the examples (53 different pairs) in the group where one of the elements of the pair is not included in the dictionary. There are significantly fewer pairs with a ranking of 2/13 (10 different pairs – 14.7%). Thus, almost the entire group (93%) consists of pairs which received rankings of 1/13 and 2/13 in linguistic works. Three pairs with a ranking of 3/13 (*clever – stupid*, *left – right*, and *sweet – sour*), one pair with a ranking of 4/13 (*honest – dishonest*) and one pair with a ranking of 5/13 (*open – shut*) also are included here.

The facts which have been presented demonstrate clearly that there are certain correlations between the rankings in linguistic works and the frequency of the pairs under study, since the pairs with rankings between 2/13 and 13/13, based on the index of divergence, mainly belong to the groups A_p (40.5%), B_p (18.9%) and C_p (16.2%). The pairs in which one of the components was not included in the dictionary (the other component mainly belonging to the first three thousands of words in the dictionary) constitute 20.3% of them.

The analysis of the semantic types of antonyms showed the numerical dominance of gradable and complementary antonyms both among the pairs in the linguistic works, and in the frequency dictionary. Reverses and antipodals were much fewer in number in both sources. Lexical semantic analysis demonstrated the dominance of the same lexical-grammatical classes both in the linguistic sources (adjectives – 60.8%; verbs – 24.3%), and in the frequency dictionary (adjectives – 47.7%; verbs – 30.2%). In terms of the formal structural approach, morphologically unrelated antonyms exceed derived ones, the former having higher rankings in linguistic works and higher indices of divergence in the dictionary.

On the basis of our assessments, the central zone of the category of antonymy in English is formed by approximately 55-60 pairs of antonyms which, on the one hand, have a high frequency, and on the other, are actively used as representative examples by the authors of linguistic works. The research which has been conducted enables us to affirm that the central zone of the field of antonymy is formed primarily by morphologically unrelated adjectives and verbs which illustrate gradable and complementary antonymy.

REFERENCES

1. Lytvyn, O.L. Lexical content of the category of antonymy (based on a selection of English-language linguistic works) / O.L. Lytvyn // New Philology. Collection of scientific papers. – Zaporizhzhya : ZNU, 2014. - № 64. – P. 49-54.
2. Chaffin, R., Herrmann D.J. The similarity and diversity of semantic relations / R. Chaffin, D.J. Herrmann // Memory and Cognition 12 (2), 1984. – P. 134-141.
3. Corpus of Contemporary American English (COCA) / URL: <http://corpus.byu.edu/coca/>.
4. Cruse, D.A. Lexical semantics / D.A. Cruse. – Cambridge : Cambridge University Press, 1987. – 311 p.
5. Davies, M., Gardner D. A frequency dictionary of contemporary American English : word sketches, collocates, and thematic lists / M. Davies., D. Gardner – London, New York : Routledge, 2010. – 357 p.
6. Lyons, J. Semantics / J. Lyons. – Vol. 1. – Cambridge : Cambridge University Press, 1977. – 375 p.
7. Murphy, M.L. Semantic relations and the lexicon : antonymy, synonymy, and other paradigms / M.L. Murphy. – New York : Cambridge University Press, 2003. – 296 p.