

ЛЕКСИКОЛОГИЈА И ЛЕКСИКОГРАФИЈА
У СВЕТЛУ САВРЕМЕНИХ ПРИСТУПА

Зборник научних радова

SERBIAN LANGUAGE INSTITUTE OF SASA

LEXICOLOGY AND LEXICOGRAPHY
IN THE LIGHT OF CONTEMPORARY
APPROACHES

A collection of papers

Edited by:

Stana Ristić, PhD, scientific advisor

Ivana Lazić Konjik, PhD, senior research associate

Nenad Ivanović, PhD, research associate

Belgrade

2016

ISBN 978-86-82873-56-3

ИНСТИТУТ ЗА СРПСКИ ЈЕЗИК САНУ

ЛЕКСИКОЛОГИЈА И ЛЕКСИКОГРАФИЈА У СВЕТЛУ САВРЕМЕНИХ ПРИСТУПА

Зборник научних радова

Уредништво:

др Стана Ристић, научни саветник
др Ивана Лазић Коњик, виши научни сарадник
др Ненад Ивановић, научни сарадник

Београд
2016

Издавач:
Институт за српски језик САНУ

Рецензенти:
проф. др Андреја Желе
проф. др Вера Васић
проф. др Даринка Гортан Премк
др Људмила Ришкова

Коректура: аутори и Уредништво

Припрема за штампу:
Давор Палчић

Штампа:
Чигоја штампа

Тираж: 300

Издавање ове књиге финансијски је помогло Министарство просвете,
науке и технолошког развоја Републике Србије

НАПОМЕНА О АУТОРСКОМ ПРАВУ:

Ниједан део ове публикација не може се прештамповати, репродуковати или употребити у било
ком облику без писменог одобрења аутора као носиоца ауторског права.

COPYRIGHT NOTICE:

No part of this publication may be reprinted, reproduced or utilized in any form without permission in
writing from the author, as the holder of the copyright.

Danko ŠIPKA
Arizona State University
Danko.Sipka@asu.edu

A PROLEGOMENON TO AUTOMATIC DETECTION OF ERRORS IN THE TREATMENT OF LEXICAL ANISOMORPHISM

Based on the theoretical framework advanced in his monograph *titled Lexical Conflict: Theory and practice* (Cambridge University Press, 2015) the author proposes the search patterns aimed at detecting typical errors in the treatment of cross-linguistic lexical anisomorphism. Special attention is devoted to the development of regular-expression-like search patterns for the detection of errors in the treatment of multiple equivalence.

Keywords: multiple equivalence, lexicography, treatment errors, automatic detection.

Introduction

The present paper is intended as the first step in developing the mechanisms for automatic detection of errors in lexicographic treatment of multiple equivalence, the most prominent form of lexical anisomorphism. Multiple equivalence will be defined and its forms will be identified in the introductory section of this paper. The following section will be devoted to common errors in lexicographic treatment of multiple equivalence. The third section will address the possibility of automatic detection of the aforementioned errors, while the final section will outline the possibilities of further development of error-detection mechanisms.

The present paper construes multiple equivalence as a subset of cross-linguistic lexical differences. We will propose a taxonomy of the types of multiple equivalence and the strategies of its treatment in bilingual dictionar-

ies. In addition to the obvious interest this phenomenon has been commanding in lexicography (e.g., Zgusta 1971, Al-Kasimi 1977, Yong and Peng 2007), important direct and indirect contributions to the study of multiple equivalence can be found in various other intellectual traditions from ordinary language philosophy (most notably in Frege 1982 and Puttnam 1975), to linguistic anthropology (Goddard 2012), translatology (Pym 2009), and second language teaching (Kramersch 1993).

Based on bottom-up metadata gathering of more than 100 languages from bilingual dictionaries, linguistic monographs, and structured interviews with dictionary compilers and users, Šipka (2015) identifies the following types of multiple equivalence: multiple equivalence in operators, entity-based multiple equivalence, concept-related multiple equivalence, tangled multiple equivalence, scalar multiple equivalence, zero-equivalence-based multiple equivalence, and partial-equivalence-based multiple equivalence (which includes application splits, organization splits, syntagmatic splits, network-based splits, frequency-based splits, and mental-imagery splits). One particular dictionary entry may contain more than one type of multiple equivalence. For example, there exists non-scalar multiple equivalence between American English *carburetor* and its three Croatian counterparts: *rasplinjač*, *karburator*, and *fergazer*, but these equivalents also feature a connotation difference (*rasplinjač* is considered bookish, *karburator* neutral, and *fergazer* archaic). There is furthermore a syntagmatic difference (*rasplinjač* is used in formal texts, the other two in informal texts), as well as a network difference: *rasplinjač* is a compound word, consisting of the prefix *ras-* ‘distribute’, the suffix *-ač* ‘agent’ and the stem *plin* ‘gas /fluid?’. Its literal meaning is ‘the agent of distributing gas’ and, as such, it features a network link with the word *gas*, while other two words, being borrowings from French and German respectively, do not have such links (however there is awareness of them being borrowed).

Lexicographic treatment of multiple equivalents is only partially influenced by the type of multiple equivalence, other relevant factors include dictionary type and scope, its microstructure, rendering medium, intended dictionary users and their needs, as well as general principles of lexicographic consistency. Needless to say the unequivocal differentiation of multiple equivalents is of primary concern in any dictionary. Šipka (2015) establishes the following types of valid lexicographic strategies (i.e., those strategies that are successful in resolving multiple equivalence): enumeration, decomposition, paradigmaticization, paraphrasing, specification, exemplification, co-textualization, contextualization, reconstruction, and cross-referencing. Each strategy exhibits a higher or lower degree of ap-

propriateness for one or more established types of multiple equivalence. Obviously, in real word these sound strategies are not always followed and the main thrust of our analysis here will be devoted to the deviation from valid strategies and the possibilities of their automatic detection.

Common errors in lexicographic treatment of multiple equivalence

We can distinguish the following commonly encountered erroneous strategies (i.e., the divergencies from the sound strategies).

- a. Strategy mismatch,
- b. Vagueness,
- c. Lack of explanation or specification,
- d. Redundancy,
- e. Underspecification overlaps,
- f. Overspecification gaps,
- g. Mislabeling,
- h. Inadequate exemplification,
- i. Language mismatch,
- j. Inconsistent segregation,
- k. Equivalent omission
- l. Overinterpretation

We will now discuss each of the aforementioned divergences in turn.

Strategy mismatch happens when one of the previously mentioned legitimate strategies is used in a place where another strategy would be more appropriate. The most common instance of this divergence from legitimate strategies is the use of enumeration in a situation when some kind of discrimination of the equivalents would be appropriate, e.g., through decomposition, specification, etc. For example, if an English-Slovak dictionary intended for production merely enumerates the two most frequent Slovak equivalents of the English noun *wall*, i.e.:

*wall N stena; múr

the English user will not know when to use which equivalent.

Vagueness is a problem area that possibly affects all kinds of entries, with single and multiple equivalence alike. The problem of vagueness is an issue demonstrated in each of the particular equivalents rather than in the inability to distinguish them. Consider the following hypothetical Serbian-English example:

***косац**, -сца #m 1 agricultural worker 2 a kind of bird 3 a sort of insect

We can see that the equivalents are separated from one another (which makes it possible for such vague references to be discriminating glosses) but no real equivalence is established.

Lack of explanation or specification comes in two different forms. Either the equivalents themselves (be it one or more of them) can be unspecified, which affects SL and TL speakers alike, or the devices which are supposed to distinguish the equivalents may be overly general, which affects SL speakers' productive abilities.

Let us consider the following hypothetical examples, the first Punjabi – English and the second English – Afar:

* **بہار** bahaar 1. spring 2. happiness

***acknowledge** [...] v.tr 1 eeleme / eemene [admit] 2 loowite [recognize]

The problem with the first hypothetical example is that both equivalents have multiple meanings – *spring* can refer to a season, water source, metal device, jump, etc. In this particular case the relevant meaning is that of springtime. In the second example, the difference is that the first verb is used about facts (e.g., one's own mistakes) and the second about people (e.g., someone as a member of an organization). However, the two synonyms of the English verb *acknowledge* can be used about facts and people alike, so their use remains unspecified.

Redundancy occurs when the equivalents have more information than needed to establish the equivalence or when discriminating glosses contain superfluous information. A very common source of such divergence happens when discriminating glosses are unnecessarily overdeveloped into monolingual definitions, e.g., in the following English-Spanish example, where English monolingual definitions are used instead of one-word synonyms:

***gallantry** [...] N (the qualities of a hero or heroine; exceptional or heroic courage when facing danger, especially in battle) valor m, valentía f; (a courteous or respectful or considerate act or manner) galantería f, cortesía f.

As can be seen, the entry is most confusing, as the superfluous information make its processing slow and complicated. The entry would become equally confusing if we were to substitute one synonym of the Spanish word *valor* (which perfectly suffices to establish the equivalence) with additional synonyms such as: *audacia*, *coraje*, *osadía*, *arroyo*, *arresto*, *intrepidez*, *entereza*, *atrevimiento*. This type of divergence does not affect the equivalence

per se – it just leads to an unnecessary increase in the processing time and results in the difficulty of dictionary use.

Underspecification overlaps are the first divergence characteristic specific for multiple equivalence, the one that we do not encounter in zero equivalence. This divergence occurs when one or more multiple equivalents or discriminating glosses remain overgeneralized, which creates an overlap between the equivalents, and in fact no overlap exists in reality. To exemplify this, let us use the example of the English word *sea dragon*. It refers to two different species, and a correct English-German equivalence in a production dictionary would be something like:

***sea dragon** [...] 1. Große Fetzenfisch [leafy sea dragon] 2. Seedrachen, Kleine Fetzenfisch [weedy sea dragon]

In a hypothetical example of an underspecification overlap, one would either use *Fetzenfish* as either equivalent 1 or 2 or use *Syngnathina* (i.e., the subfamily that includes both these fish species) as one of the distinguishing glosses, for example:

***sea dragon** [...] 1. Fetzenfisch [leafy sea dragon] 2. Seedrachen, Kleine Fetzenfisch [weedy sea dragon]

or

***sea dragon** [...] 1. Große Fetzenfisch [*Syngnathina*] 2. Seedrachen, Kleine Fetzenfisch [weedy sea dragon]

In either case one equivalent will overlap with the other – actually, the broader one, here number 1, would encompass the narrower one.

Overspecification gaps are the antipode of underspecification overlaps. In this case, the consequence of the divergence in question is that a part of the range of the SL word is not covered by the equivalents. Consider the following example. The English genus name *bluebell* can refer to two different genera and the correct rendering of that entry in a hypothetical English – Upper Sorbian botanical dictionary would be as follows:

*bluebell [...] 1. česnička [*Hyacinthoides*] 2. kitelnička [*Muscari*]

If, however, we replace the first genus name with a species name, all other species within that genus will not be covered, i.e., an overspecification gap will be created, e.g.,

*bluebell [...] 1. česnička [*Hyacinthoides hispanica*] 2. kitelnička [*Muscari*]

or

*bluebell [...] 1. zwónčkata česnička [*Hyacinthoides*] 2. kitelnička [*Muscari*]

The diversion of **mislabeled** takes place when a wrong or missing label (grammatical, usage, etc.) leads to the failure to segregate the equivalents. For example, the Scottish English word *Highers* has several German equivalents, which are differentiated geographically and stylistically, i.e., *Abitur* (in Germany, neutral), *Abi* (in Germany, informal), *Matura* (in Austria, neutral), *Maturität* (in Switzerland, neutral). Mislabeled such as:

***Highers** (Scottish) [...] *Abitur*, *Abi* (in Germany), *Matura* (in Switzerland), *Maturität* (in Austria)

would lead into the use of *Abi* in formal contexts (the lack of a label for marked usage is a type of mislabeling given that general use is assumed where no label is applied). This kind of entry would furthermore cause that two country-specific words would be used in wrong countries.

Inadequate exemplification takes place where the examples do not help in distinguishing the equivalents. The following Russian-English example illustrates this kind of diversion:

До prep. gen. to, so far, as far as, till, until; – *последней копейки, капли*, to the last penny, to the last drop, to the dregs; – *которыхъ поръ*, – *чего* how far; – *сихъ пор*, thus far, till now, hitherto; – *тыхъ пор* then; – *настоящей минуты*, up to the present moment; *дожить* – старости, to reach old age; *что – меня или что касается – меня*, as for me, for my part; *онъ добежался – того, что упал*, he ran so much that he fell[...]

(Alexandrow, 1907:127)

As can be seen, the examples are not very helpful in differentiating the equivalents – it is impossible to generalize some kind of rule as to when to apply which equivalent based on the provided examples. The solution to the present diversion is to provide the examples which would be characteristic of each of the TL equivalents.

Unlike all previous problem areas, which typically affect individual entries, **language mismatch** is oftentimes an erroneous general lexicographic strategy. The following German – English entry illustrates the point:

Abschroten, (*w.*) *v. tr.* 1) to gnaw off, devour (as grubs, &c. little roots); 2) to take off the rough, to rough-hew, to cut or clip off, to work, hew, &c. rudely or coarsely, as for first purposes: to give the first form or shape, *among different trades*

(Flügel, 1894: 35)

The explanations in English are quite extensive yet the fact that they are in the TL makes it impossible for the German users (i.e., SL speakers here) to select the right equivalent.

Another problem that occurs globally is **inconsistent segregation**. This happens when some entries contain discriminating glosses while others do not. The consequence of such inconsistency is that the SL speakers cannot rely on the dictionary in their TL production. This diversion can be seen from the following two Zulu – English entries, which are just two spots in alphabetical order apart, where the first one contains the synonyms to discriminate the equivalents and the second does not.

i-mBaba (Bhaaba), n. Unsound, badly shaped pumpkin, generally discarded on the field (= i-nGxwele); hyaena (= Im-Pisi).

[...]

uku-Baba, n. Strongness; bitterness; sourness; acidity; pungency; itchiness, etc. See baba.

Bryant (1905: 13)

Equivalent omission comes about when only one equivalent is included in the entry in a situation where two or more equivalents would be warranted by their frequency or another selection criterion. The consequence is an illegitimate elimination of multiple equivalence. For example, the English word *bike* has two equally frequent meanings, i.e., that of bicycle and that of motorcycle. Equivalent omission would happen if the entry is to be rendered as:

**bike* N ~s biçikletë, = *bicycle*

the equivalence will be established correctly but the real-life case of multiple equivalence would be illegitimately transformed into a case of one-to-one equivalence.

Overinterpretation happens when relations that do not exist in reality are established between the equivalents. Shepherd (2012) provides the following Kurmanji Kurdish example:

In Kurmanji, the word *dirûnker* has two meanings, one being the occupation of tailor, the other, that of a laborer in a field/a reaper. *Dirûn* can mean both “to sew” and “to plant” and *-ker* is a suffix denoting “one who.” At first I thought that maybe the similarity is that they both involve putting something in something, as in with sewing, one puts a thread in fabric, with planting, one puts a seed in the ground. However, I’m not sure that *dirûnker* in the sense of a laborer is specific to the task of planting.

To exemplify this further, if a metaphorical link of this kind does not really exist, overinterpretation would be present in the following hypothetical entry.

**dirûnker* 1. reaper 2. (metaphorically) tailor

Possible detection mechanisms

Now that broadly encountered errors in the treatment of multiple equivalents are identified, we can address the key issue of this paper, namely the possibility of their automatic detection. The importance of such detection is self-evident. Establishing equivalence in a clear and consistent manner is at the heart of any dictionary project while differentiating between multiple equivalents represents the most urgent need of production dictionary users.

Automatic detection of errors in lexicographic treatment assumes a feasible and reasonably accurate search pattern of some kind. A search pattern, in turn, needs to be efficient and its construction should not be overly time-consuming. More precisely, when deciding if it makes sense to develop a search pattern one should be guided by the following consideration:

- a. Design feasibility,
- b. Design resources complexity,
- c. Comprehensiveness,
- d. Exclusiveness.

The first parameter of a possible search pattern is a simple check if the error in question lends itself to automatic detection. In some cases, there are no clear and consistent external markers of an error, which makes the whole enterprise of constructing a search pattern impossible. An additional factor here is the medium of the dictionary. A dictionary in the form of a relational database with clearly defined microstructural segments may lend itself to the design of a search pattern which is impossible in a poorly formatted textual dictionary.

If the answer to the first question above is positive, i.e., if the development of a search pattern is possible given the type of error, the next question should be what kinds of resources one would need to muster to design that particular search pattern. It may be so that gathering the resources for a search pattern is overly time consuming and that it makes more sense to search for the errors manually, by visually scanning the dictionary database.

The next step in building the search pattern is the question about the proportion of the errors in question that will be detected by that particular search pattern. Ideally, the search pattern should detect all instances of treat-

ment errors, but it may make sense to deploy the patterns that find most of them, especially if the design of the pattern is not overly time consuming. In this situation of underdetection, the cases that are not found by the search pattern may be addressed manually, left to be corrected in other editions of the same dictionary, etc.

Finally, we want the search pattern to exclude the cases which are not errors that we are looking for. Ideally, the output of the search patterns should not contain any non-errors. Again, in reality, we may want to allow a certain level of over-detection, especially if the development of the pattern is not resource-heavy.

We will now consider all aforementioned commonly encountered errors in the treatment of multiple equivalence applying the four search-pattern considerations (feasibility, resource complexity, comprehensiveness, and exclusiveness).

Strategy mismatch, as we saw in the previous section, is a rather broad category of errors as it assumes that one legitimate strategy is deployed in the situation where another strategy would be appropriate. The possibility of automatic detection will vary depending on which two strategies are being mismatched. However, the most common mismatch, which concurrently features furthest reaching consequences is the use of enumeration (i.e., providing a simple list of bare equivalents) in a production dictionary, where the user needs some kind of guidance in the selection of the appropriate equivalent. We will consequently address the possibility of automatically detecting that kind of strategy mismatch. This kind of error will manifest itself in the absence of discriminatory glosses. It is certainly possible to detect the lack of something but the question is then how resource-heavy, comprehensive, and exclusive the search pattern may be. These three parameters are heavily influenced by the format of the dictionary. We will use the example of an English – Slovak hypothetical entry:

*wall N stena; múr

where, in a production dictionary, we would expect something like

*wall N stena (a part of a building); múr (free-standing)

If we have a database with dedicated fields for equivalents, discriminatory glosses, and usage labels, we would just have to look for the records with two or more equivalents and with empty discriminatory glosses and usage labels fields (to account for the cases where the discrimination is performed by usage labels rather than glosses). The design of the search pattern is a simple database query, which does not require practically any resources,

and the search pattern is fully comprehensive and exclusive. On the other hand, if we are dealing with a dictionary which is a simple text rather than a database, we would need to see what the conventions for graphically marking microstructural elements are and then design a regular-expression search pattern based on those conventions. For example, in a dictionary which uses sharp brackets $\langle \rangle$ to render usage labels and angular [] to mark discriminatory glosses we would be looking for (in Perl notation): $[\wedge[\langle]+; [\wedge[\langle]+$, i.e., repetition of any number of any characters different from the opening sharp or angular bracket (which would find the first equivalent without a discriminatory gloss), followed by a semicolon and a space, followed by repetition of any number of any characters different from the opening sharp or angular bracket (which would find the second and any further equivalents without a discriminatory gloss). Obviously we may get some amount of over-detection, and if the use of the semicolon is not consistent, some amount of under-detection, but numerous errors will be detected using a simple non-time-consuming search pattern. A significant lesson learned from assessing the possibility of automatic detection of this particular type of errors is the importance of a clearly formatted dictionary. The less structured the dictionary microstructure, the less accurate the search patterns are.

Vagueness Most prototypical cases of this divergence are easily detectable by looking for the phrases such as “(a) kind of”, “(a) sort of”, etc. This is a simple search pattern using literals. Obviously, there are cases which are not going to have common external marks of vagueness, which case the search pattern not to be fully comprehensive. Similarly, the aforementioned phrases can be legitimate in some glosses, etc., so if the microstructure is not segmented properly, the pattern will underperform in its exclusiveness parameter.

Lack of explanation or specification can be detected automatically by looking into one-word equivalents and glosses respectively. An on-line tool for this kind of detection, based on WordNet (see Miller, 1995) is available at: <http://www.asusilc.net/dict>. For example, entering the following Serbo-Croatian – English headword and one of the equivalents: *kran/crane* returns the following:

! /crane/=> |crane|large long-necked wading bird of marshes and plains in many parts of the world|lifts and moves heavy objects; lifting tackle is suspended from a pivoted boom that rotates around a vertical axis|stretch (the neck) so as to see better; “The women craned their necks to see the President drive by”

Three different meanings of the word *crane* listed in WordNet are listed to alert the compiler of the dictionary to additional meanings of the headword and the fact that the equivalent is insufficiently specific. For more details about this tool see Šipka (2007a).

Automatic detection of **redundancy** in the case of overabundant synonyms would typically rely on searching for a series of words separated by commas and/or semicolons. Any such series with three or more words in it should definitely be extracted from a production dictionary. Similarly, any long string of words (e.g., five or more) would need to be extracted and examined.

Underspecification overlaps and **overspecification gaps** are practically impossible to detect automatically, or, in other words, building the tools for the automatic extraction of such cases would be overly time- and labor-intensive to be feasible. These two divergences are based on the semantic range of the equivalents and, in order to detect any possible problems, one would need to have full semantic mapping of each equivalent and the ability to compare them.

Automatic detection of **mislabeled** would rely on comparing the labels of the equivalents with those from comprehensive monolingual dictionaries. All cases where the label used in the authoritative monolingual dictionary differs from that in the bilingual dictionary in question would be extracted. This may include the cases where the monolingual dictionary does not have a label and the bilingual one does, those where the label is present in the monolingual dictionary but absent from the bilingual dictionary, and finally those where the labels differ (once we normalize the system of labels).

Automatic detection of **inadequate exemplification** would search for examples without glosses, i.e., a short equivalent string of one or two words followed by one or more examples.

Automatic detection of **language mismatch** can be performed by language detection of the discriminating glosses, which is relatively easy for medium- and high-density languages (e.g., German, English, Russian, Danish, etc.) using monolingual spell-checking databases. Low-density languages, e.g., Somali, Chuvash, etc. would need to rely on ad-hoc created stemmers based on the material of the same bilingual dictionary that is being examined.

Automatic detection of **inconsistent segregation** relies on looking for a series of words separated by commas and/or semicolons with no glosses between them.

Obviously, in a production dictionary, discrimination would need to be consistent as to enable the segregation of equivalents in all entries.

Equivalent omission is rather difficult to automatically detect. A rather intricate mechanism can be put in place, which would compare SL monolingual entries with the analyzed bilingual dictionary and extract those entries which have several meanings in the monolingual dictionary but only one equivalent in the bilingual dictionary. Obviously, in many instances, one TL equivalent will legitimately cover all SL meanings, so numerous extracted cases would have to be discarded before one gets to the real divergences.

In **overinterpretation**, as in the cases of underspecification overlaps and overspecification gaps, automatic detection is not feasible.

Outlook

We saw that not all cases of divergences lend themselves to the development of automatic detection mechanisms. In some cases the cases do not offer an adequate amount of relevant information, in others the amount of resources needed to develop the detection mechanism is too high to warrant a feasible development. In addition, even there where such mechanism is possible and feasible, the mechanism is typically not ideally comprehensive and exclusive. Nevertheless, the construction of search patterns for automatic detection of errors in the treatment of multiple equivalence is imperative, given the serious consequences they may have in dictionary usage. There are two major fronts where this development ought to be directed in the future. The first lies in a broader spread of rigorous segmentation of dictionary microstructure (e.g., that every bilingual dictionary is rendered as a database with a separate field for one particular microstructural element). As we have seen in the previous section, this will facilitate the construction of the detection mechanisms. The second development front should lie in the refinement and expansion of the detection mechanisms per se. Such development should rely on facilitating the acquisition of the information needed to develop the search pattern as well as on maximizing the comprehensiveness and exclusiveness of the patterns. The present prolegomenon is meant to be the first step toward the aforementioned two lines of development in automatic detection of lexicographic errors in the treatment of multiple equivalence.

REFERENCES

- Alexandrow, A. 1907. *Complete English-Russian dictionary*. London: D. Nutt
Al-Kasimi, A. M. 1977. *Linguistics and Bilingual Dictionaries*. Leiden: E.J. Brill.
Bryant, Alfred T. 1905. *A Zulu – English Dictionary*. Pinetown: Mariannhil Mission Press.
Flügel, Felix 1894. *A Universal English – German and German – English Dictionary*. Braunschweig: George Westermann.

- Frege, G. 1892. Über Sinn und Bedeutung. In: *Zeitschrift für Philosophie und philosophische Kritik* 100, 25–50.
- Goddard, C. 2012. Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. *Linguistics*, 50–3, 2012, 711–743.
- Kramersch, C. 1993. *Context and culture in language teaching*. Oxford: Oxford University Press.
- Putnam, H. 1975. Mind, Language and Reality. *Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.
- Pym, A. 2009. Natural and Directional Equivalence in Theories of Translation. In: *Metalanguage of Translation*. Y. and L. van Doorslaer (eds.). Amsterdam: John Benjamins, 81–104.
- Shepherd, L. 2012. *Private e-mail communication*.
- Šipka, Danko 2007a. Ambiguous definitions: A case study. *Poznan Studies in Contemporary Linguistics* 42, 269–278.
- Šipka, D. 2015. *Lexical Conflict: Theory and practice*, Cambridge: Cambridge University Press.
- Yong, H., J. Peng. 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam: John Benjamins.
- Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.

Данко Шипка

ПРОЛЕГОМЕНОН АУТОМАТСКОМ ОТКРИВАЊУ ГРЕШАКА
У ОБРАДИ ЛЕКСИЧКОГ АНИЗОМОРФИЗМА

Резиме

Полазећи од теоријског оквира предложеног у својој монографији под насловом *Lexical Conflict: Theory and practice* (Cambridge University Press, 2015) аутор предлаже обрасце за откривање типичних грешака у обради међујезичког лексичког анизоморфизма. Посебна пажња посвећена је формулисању образаца претраживања сличних регуларним изразима за откривање грешака у обради анизоморфизма заснованог на вишеструкој еквиваленцији. Посебно су обрађене појаве неприлагођене стратегије, нејасности, недостатка објашњења, редунданције, подспецифицираности, надспецифицираности, погрешног етикетања, неадекватног опримерења, помијешаних језика, недоследног одвајања, ненавођења еквивалената и неосноване интерпретације. Кад су у питању обрасци откривања ових грешака, обрађени су следећи њихови аспекти: изводивост, комплексност ресурса за њихову израду, њихова обухватност и искључивост.

CIP – Каталогизација у публикацији
Народна библиотека Србије, Београд

811.16'373(082)

811.16'374(082)

ЛЕКСИКОЛОГИЈА и лексикографија у светлу савремених приступа : зборник научних радова / уредништво Стана Ристић, Ивана Лазић Коњик, Ненад Ивановић. – Београд : Институт за српски језик САНУ, 2016 (Београд : Чигоја штампа). – 577 стр. ; 24 cm

На спор. насл. стр.: *Lexicology and Lexicography in the Light of Contemporary Approaches*. – Радови на више језика. – Тираж 300. – Напомене и библиографске референце уз текст. – Библиографија уз већину радова.

ISBN 978-86-82873-56-3

1. Ристић, Стана, 1950– [приређивач, сакупљач]
 2. Лазић Коњик, Ивана, 1978– [приређивач, сакупљач]
 3. Ивановић, Ненад, 1979– [приређивач, сакупљач]
- а) Словенски језици – Лексикологија – Зборници
б) Словенски језици – Лексикографија – Зборници

COBISS.SR-ID 228321036