MORE OR LESS THAN A DICTIONARY? WORDNET AS A MODEL FOR SERBIAN L2 LEXICOGRAPHY

Toma Tasovac

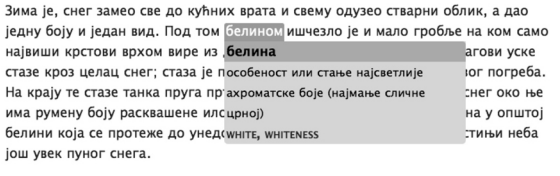
Center for Digital Humanities, Belgrade

Abstract: Transpoetika Dictionary is a Wordnet-based, bilingualized dictionary of the Serbian language, primarily aimed towards advanced students of Serbian as a foreign language (L2). The reason for basing a learner's dictionary on the model of a large semantic database is not necessarily self-evident: Wordnet excels in representing the complexity of the relations between entries, but not in the complexity of information (grammatical, syntactic, orthoepic) assigned to individual entries. The paper explores the advantages and shortcomings of a Wordnet-based L2 project and offers some solutions in extending Wordnet so that it can satisfy the needs of L2 lexicography.

Keywords: Serbian, L2, foreign language, Wordnet, lexicography, Transpoetika

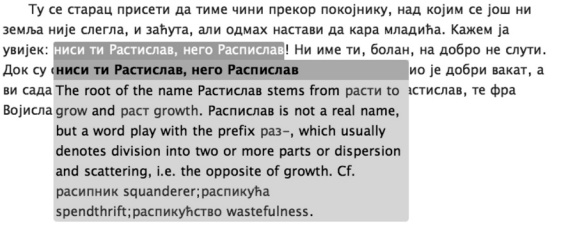
1. Introduction: the lexical turn

Transpoetika is a scalable, web-based, digital framework for editing and publishing annotated, fully-glossed study editions of literary works in the Serbian language. It currently consists of two tightly integrated modules: TransText (TT), an XML-based online editor and publisher, and TransDictionary (TD), a collaborative, WordNet-based Serbian-English dictionary. Each word in a published text is glossed by being linked to a particular sense of the corresponding TD entry.



Picture 1. Clicking on a word in a text shows the appropriate gloss

In addition to that, individual words, multiword expressions and sentences can be annotated with editorial comments.



Picture 2. Alt-clicking on an annotated phrase shows the textual commentary

The Transpoetika framework thus provides a set of tools to facilitate reading, comprehension and discussion of Serbian literary texts in an online environment. It is aimed primarily at advanced students of Serbian as a second language, heritage speakers and humanist scholars working with Serbian texts. Once its development has been completed, Transpoetika will be available online at http://transpoetika.org.

TD is the first electronic lexicographic resource being developed specifically for students of Serbian as a second language (L2)[1]. As a modern, freely accessible, web-based dictionary, it could become an important pedagogical asset to all the schools and universities wherever the language is taught. Foreign-language teaching in general has been increasingly focusing on the role of the lexicon and a need to move the pedagogical focus from lexicalized grammar to grammaticalized lexis (Nation, 1990; Lewis, 1993; Carter, 1998; Singleton, 1999; Lewis and Conzett, 2000; Schmitt, 2000; Nation, 2001).

This shift towards the lexicon can be seen – in broader terms – as part of the "lexical turn" in general linguistics, away from Chomsky's early syntax-based constraints of a universal grammar, which declared meaning as interpretive and immaterial to the study of language, towards the later Minimalist Program (MP), according to which language learning is primarily a case of learning a lexicon (Chomsky, 1995; Boeckx, 2006).

Most importantly, perhaps, the "lexical turn" in linguistics could be been a consequence of developments in corpus linguistics (Sinclair, 1991; Sinclair, 2000; Sinclair and Carter, 2004; Sinclair and Mauranen, 2006) and related studies of lexical or pattern grammars (Partington, 1998; Hunston and Francis, 2000; Sinclair, 2000), both of which have broadened the descriptive basis for dictionary making by applying computer technology to an increasingly large body of linguistic evidence.

The development of a new, electronic lexicographic resource for the Serbian language is all the more important because mainstream lexicography in Serbia is seriously lagging behind contemporary digital trends.[2] The important computational resources (Krstev et al., 2008; Obradović and Stanković, 2008) which are under development at the Faculty of Mathematics – including the Serbian Morphological Dictionary (SMD) and Serbian WordNet (SWN), for instance – have so far not been used in the production of actual dictionaries for human use.

And even though modern lexicography is unimaginable without computer technology (Knowles, 1989; Meijs, 1992; Hockey, 2000), the sheer use of computers in producing a dictionary or delivering it electronically does not automatically transform a dictionary from "a simple artefact" to a "more complex lexical architecture," to use Sinclair's (2000) formulation. Calling dictionaries "simple artifacts" is itself a rhetorical oversimplification: there is nothing simple about a dictionary – whether we look at it as a material object, cultural artifact or a model of language. Yet the overall structure of how dictionaries are constructed – as extended word lists and prototypical hypertexts which make up a web of mutual references – has not changed in centuries. Most contemporary electronic dictionaries are simply digitalized transcriptions of their print editions. They have better and more efficient search mechanisms, but they rarely introduce features that are more complex in terms of the overall lexical structure.

TD, on the other hand, has been designed to take advantage of one such complex architecture: the hierarchical semantic network of Princeton Wordnet (Fellbaum, 1998). The Princeton Wordnet (WN) and its counterparts in other languages have been used in the context of computer linguistics and NPL, including word-sense disambiguation, homophone correction, document classification, document retrieval, open-domain Q&A and textual entailment. WN is also available online as a lexical resource for human use.[3] Yet I know of no other examples of WN-based L2 lexicographic resources.

This apparent lack of use in L2 lexicography may have to do with the inherent difficulty of aligning complex lexical databases. If monolingual mapping between words and concepts in a hierarchical system introduces a first order of complexity to a lexicographic project, cross-lingual mapping between word/concept pairs adds to it a second order of complexity, which is anything but trivial.[4] Wordnet architecture, however, has been applied to numerous other languages and used in projects like EuroWordNet (Vossen, 1998) and BalkaNet (Stamou et al., 2002; Tufis et al., 2004). Cross-linguistic issues, such as representing natural gender (Ordan and Wintner, 2005), dealing with lexical gaps (Bentivogli and Pianta, 2000; Bentivogli et al., 2002; Bentivogli and Pianta, 2003) or handling language-specific concepts (Крстев, 2004; Krstev et al., 2006) are well-documented and have been taken into consideration in the design of the TD database.

The development of TD on the basis of WN has been informed in particular by the research

done by the Human Language Technology Group (HLT) at Belgrade University (Vitas et al., 1998; Vitas et al., 2003). TD and the HLT's Serbian Wordent are separate projects with different goals, but complimentary in nature as far as the creation of a Serbian lexical database is concerned.

In this paper, I will focus on the reasons for choosing WN as the building block of the Transpoetika project, assess the advantages and shortcomings of basing an L2 resource on WN and, finally, offer some solutions as to how the WN lexical ontology could be extended so that it satisfies the specific pedagogical needs of an L2 dictionary.

2. Wordnet: more or less than a dictionary?

WN is a comprehensive, machine-readable lexical database of the English language. It is structured as a formalized and selective representation of the way lexical items are organized in the mind. In other words, it is a computational version of our "mental lexicon." [5]

WN was created as a resource that would help with psycholinguistic research: a kind of supplement to, rather than a complete replacement of, conventional dictionaries. That is why WN is a strange beast in the contemporary lexicographic jungle: both more and less than a dictionary. It is less than a dictionary because it does not contain some of the basic features expected from a lexicographic resource: definitions of all parts of speech, pronunciation, word stress, etymology, usage notes.[6] In fact, it is built on the assumption that its user is either a machine or a competent language user: which is why some words are not even defined. The assumption in WN is that a person who already knows the concept associated with a word form will be able to distinguish among its meanings from its synonyms (Miller et al, 5.).

At the same time, WN is more than a dictionary because it contains features that are not to be found in conventional lexicographic works. The basic lexical entry in Wordnet is not a word but a lexical concept manifesting itself in a set of synonymous lexical items (synset). Each synset occupies a praticular place in a network of lexical and semantic relations: nouns are organized in WN as topical hierarchies with lexical inheritance (hypernymy/hyponymy or IS-A relation and meronymy/holonymy or PART-WHOLE relation); adjectives and adverbs on the basis of bipolar oppositions (antonyms); and verbs are organized by a variety of entailment relations.

Miller calls the traditional dictionary entry – with its Aristotelian manner of defining terms by means of establishing a superordinate (*genus proximum*) and its *differentia specifica*, but without exploring other types of semantic relations – "woefully incomplete" (Miller et al, 11). Admittedly, a great deal of factual information is packed into a prototypical dictionary entry, but a defined term often remains in isolation and insufficiently connected or embedded into the language system as a whole. The promise of digital lexicography stems not only from the transformation of the production medium, but also from the technological feasibility of complexity. A dictionary, which is modeled like a relational database, can accommodate a much wider range of lexical and semantic relations between words than a print dictionary ever could without becoming bloated or unmanageable.

As a dynamic thesaurus that not only defines meanings but also relations between words and concepts, WN tries to bridge the gap between factual and structural information. Being a digital resource, it avoids the pitfalls of redundant entries, which appear in alphabetized print thesauri, and redundant lookup strategies, which affect topical thesauri.[7] A traditional dictionary may tell you what a car is, or even perhaps what its synonyms are, but it will not tell you what types of cars exist (hyponyms), what are parts of a car (meronyms), what are its sister terms (hyponyms of hypernyms) etc. By explicitly encoding lexical and semantic relations, WN is making it possible – and easy – for users to access words

they do not necessarily know by making it trivial to execute database queries of the type "What are parts of *?", "What does * entail?" and so forth.

This type of complex lexical and semantic architecture will not only be useful to L2 students, but may, in fact, help overcome what Shvedova calls one of the paradoxical apects of a (traditional) dictionary entry: the lexicocentrism of the entry vs. the class-centrism of language itself (Шведова, 1988). By treating all of its entries as members of semantic classes, and by placing these classes in hierarchies, Wordnet maps linguistic meaning into a consistent, logical metasystem. This, in turn, makes it possible to treat a dictionary entry not only as a depository of lexicocentric information about a given headword, but as a point of departure in the user's interaction with the linguistic system in general.[8]

3. Transpoetika: with and beyond Wordnet

TD belongs to the category of reception-oriented bilingualized dictionaries. This means that for any given Serbian entry in the dictionary, the user can expect to find both English equivalents and Serbian definitions of the headword. Bilingualized (also sometimes referred to as hybrid, glossed, translated or semi-bilingual) dictionaries are usually the result of partial or full adaptation of existing monolingual dictionaries, whereby L1 equivalents are added to L2 definitions and examples in an attempt to bridge the gap between monolingual and bilingual dictionaries (Cowie, 1987; Reif, 1987; Stein, 1990; Hartmann, 1994a; Hartmann, 1994b; Nakamoto, 1994).

A more recent empirical study on dictionary use (Thumb, 2004) concluded that the bilingualized learners' dictionary was "highly usable and useful because of its compatibility with the language needs of learners" (108) while showing that the users used both definition languages, although with varying frequency, depending on their level of proficiency and the word in question. There is inconclusive evidence suggesting that bilingualized dictionaries are superior to bilingual dictionaries in language learning (Pujol et al., 2006). While leaner's' dictionaries are sometimes believed to require definitions to be expressed in simpler language than those for native speakers or – in more extreme cases – using a limited or restricted vocabulary system (Kirkpatrick, 1985), the point seems to be moot in a bilinguialized dictionary, in which the monolingual definition is always accompanied by the translation equivalents in the target language.

All the Serbian senses in the TD are manually aligned with WN so that English equivalents of Serbian words are actually, for the large part, given by WN. In other words, any given entry in the Serbian dictionary is a member of a Serbian synset which expresses the same lexical concept as its corresponding English synset. If a corresponding English synset is missing (for instance, {зимљив, зимогрожљив, зимогрозан, зимоморан (који је осетљив на хладноћу)}, it is added to the TD database as a non-lexical term in English {sensitive to cold}.

In their interaction with the language system, foreign – as opposed to native – dictionary users are significantly affected by the insufficiencies of their normative competence. That means that familiarity with a wide range of lexical, grammatical, syntactic, orthographic and orthoepic norms, which is a given for most native speakers of a language, may not be present in the foreign user's mind (Wiegand, 1985).

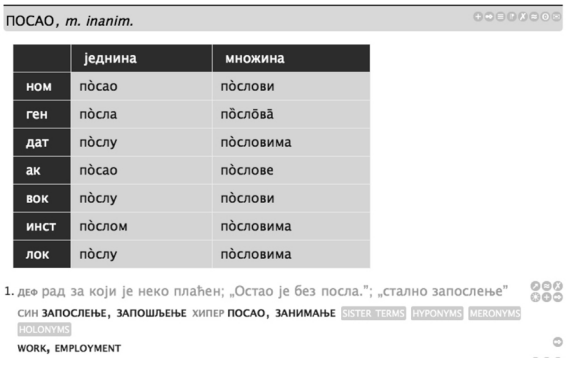
With the goal of turning a Wordnet-based lexical database into a full-fledged online dictionary for foreign users, the WN architecture in TD has been extended by a module for adding function words (pronouns, abbreviations, prepositions, numerals, particles, conjunctions and interjections), a detailed description of entryword grammar and a labeling system for differentiating among members of the same synset.

4. Extending Wordnet

Function words do not form vertical semantic hierarchies. At the same time, some function words can be described as lexically synonymous.

For instance, the genitive prepositions {до, поред, покрај, крај, искрај, украј (предлог који уз именицу или израз у генитиву означава место у непосредној близини нечега)} correspond to {beside, by, next to, at the side of, by the side of, alongside of, side by side with}, while {осим, сем, поред, поврх, врх (предлог који уз именицу или израз у генитиву означава додавање, прибрајање или додатно укључивање нечега)} corresponds to {beside, apart from, aside from, besides, other than, in addition to}. The same applies to some conjunctions, for instance {ако, уколико, у случају да, под условом да, под погодбом да}, and even interjections {jao, joj, oj}.

TD includes four types of grammatical information that we may expect to find in a dictionary: inflection, parts of speech, syntactic operation (e.g. transitive/intransitive) and implicit grammar through examples (Kirkpatrick, 1985; Whitcut, 1985). But, unlike print dictionaries, which are limited by space, TD can include full inflectional paradigms for nouns, adjectives and verbs.



Picture 3. Declension of the masculine inanimate word *посао*.

The synset labeling system is of particular importance in the TD context: firstly, because WN synsets can contain members whose connotational value diverges from other members of the same synset, even though they undoubtedly represent the same concept; and, secondly, because TD entries do not list English-language equivalents for individual lexical items, but rather a conceptual intersection between two languages in an instance of use. While an L1 student may intuitively grasp the difference between members of a synset, an L2 student needs a critical apparatus to help with the differentiation.

Labeling is a common lexicographic technique of succinctly indicating pragmatic information (Burkhanov, 2003), i.e. specifying sociocultural parameters of meaning, including relations between interlocutors, and their social and cultural roles, attitudes, values and beliefs (Апресян, 1988; Wierzbicka, 1992; Marmaridou, 2000). In the context of learners' dictionaries, it is especially important to use a labeling system to mark words which are formal or literary, informal or slang and offensive or taboo (Kirkpatrick, 1985). The WN model, however, with its emphasis on conceptual equivalence, has not been designed with pragmatics or discrimination of near synonyms in mind. English WN synsets sometimes mix registers by clustering together words with different levels of formality {lie, prevarication}, domain of use {dog, domestic dog, Canis familiaris} or regionally distinct orthographic variants (British finalise vs. US finalize).

While it would be wrong to say that pragmatic information is completely missing from WN, its presence is certainly not systematic. In some cases, WN groups together words of a particular register, as is the case with the noun synset {(slang for sexual intercourse) fuck, fucking, screw, screwing, ass, nooky...}, which is a hyponym of the more formal {(the act of sexual procreation between a man and a woman) sexual intercourse, intercourse, sex act, copulation...}. But WN is highly inconsistent in this aspect: the verb synset for the sense "have sexual intercourse with" includes not only the euphemistic "be intimate", formal "have intercourse", Biblical "know" but also colloquial (or – depending one's level of linguistic squeamishness – vulgar)

"screw, fuck, have it off..." If Serbian synsets followed the practice of ignoring the differences between conceptual and pragmatic granularity, the learners of Serbian as L2 would be denied a more complete picture of lexical entries in the dictionary.

For example, when the Serbian word "домовина" is entered in the TD database, its meaning "земља одакле неко потиче или где је рођен" is aligned against the Wordnet concept "the country where you were born." Having established the conceptual intersection, we can – at that point or any time later in the dictionary-writing process – when we encounter a Serbian word with the same meaning, add it to the existing sense to form a Serbian synset for the given concept: {домовина, завичај, отаџбина, постојбина, отачаство, земља рођења, земља порекла}.

The consequences of this approach will be multiple:

a) any member of the given synset will list all the other members of the synset as synonymous;

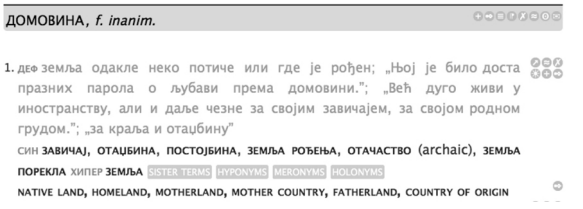
b) any member of the given synset is defined in its own dictionary entry by means of the Serbian lexical concept "земља одакле неко потиче или где је рођен";

c) any member of the given synset is translated by all the members of the corresponding English lexical concept "the country where you were born", i.e. {fatherland, homeland, motherland, mother country, country of origin, native land}.

If we take a closer look at members of the Serbian synset, we will see that their connotational value is not the same. While most lexical items in this synset are neutral, отачаство is stylistically marked as archaic. To make the Serbian synset productive for L2 users, we need to indicate which properties make individual words stand out from other lexical items in the synset.

In TD, labels are used to mark temporal, geographic, stylistic, grammatical and functional specificity of synset elements. For instance:

```
<entry>
    <sense>
        <def>земља одакле неко потиче или где је
        рођен</def>
        <form type="syn"><orth>домовина</orth>
        </form>
        <form type="syn"><orth>отаџбина</orth>
        </form>
        <form type="syn"><orth>постојбина
        </orth></form>
        <form type="syn"><orth>земља рођења
        </orth></form>
        <form type="syn">
            <orth>отачаство</orth>
            <usg type="time">archaic</usg>
        </form>
        <form type="syn"><orth>земља порекла
        </orth></form>
    </sense>
</entry>
```

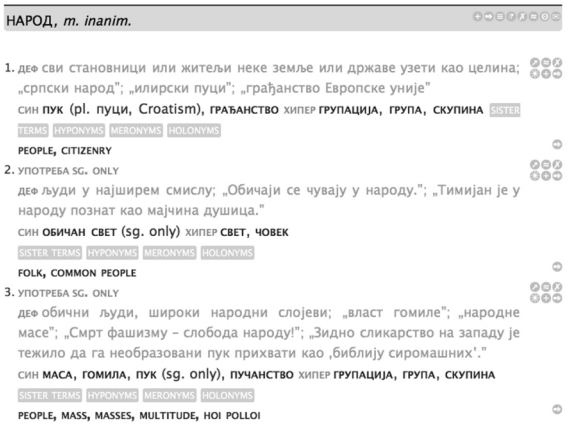


Picture 4. Dictionary entry for *домовина*.

Each word can be labeled with multiple labels, so that a sufficient amount of information can be provided to distinguish a particular element from others in the same synset. For example:

```
<entry>
    <sense>
        <def>сви становници или житељи неке
        земље или државе узети као целина;
        „српски народ"; „илирски пуци"</def>
        <form type="syn"><orth>народ</orth>
        </form>
        <form type="syn">
            <orth>пук</orth>
            <usg type="gram">pl. пуци</usg>
            <usg type="geo">Croatism</usg>
        </form>
        <form type="syn"><orth>грађанство
        </orth></form>
    </sense>
</entry>
```

In this particular case, the word "пук" is marked as a Croatism, but the dictionary label also indicates that the plural form in this particular sense is "пуци" as opposed to "пукови."



Picture 5. Dictionary entry for *народ*.

Labels are cross-referenced across the database, so that users can easily find out what other words are marked by a given label.

The labeling system in the TD achieves the stipulated goal of retaining the complex lexical architecture of a WN database, while making the particulars of that architecture susceptible to a finer-grained analysis suited to L2 needs.

5. Conclusions and future work

Wordnet was designed to overcome the inherent deficiencies of lexicocentric dictionary entries. It introduced a new, complex, lexical architecture as a way of providing structural information about words and their roles in the production of meaning. But, as a dictionary with a specifically synchronous, monolinguistic and psycholinguistic bent, WN leaves out a number of standard lexicographic features that are of particular importance for L2 users.

By including phonological, morphological and pragmatic features in a WN-aligned, Serbian semantic database, TD is an attempt at merging the rigorously hierarchical architectural complexities of WN with complex lexicographic treatment of individual entries in the production of a comprehensive L2 dictionary of the Serbian language.

Thanks to its modular architecture, TD can be gradually extended to include additional features. The extensions, discussed in this paper, should therefore be seen as only initial steps in the direction of making TD a full-fledged dictionary of Serbian for L2 users. Further areas of development will include, for instance, modules for marking grammatical patterns and morphogenetic (etymological) relations between words, as well as phraseology.

[1] While an increasing number of researchers is working on a range of topics in Serbian as L2, lexicographic resources for Serbian as a foreign language are still too few and mostly theoretical. The specification of a Council of Europe threshold level of language proficiency for the Serbian language was completed in 2002, but the results of this research project were never published (Subotić, personal correspondence). On the project itself, see van Ek and Trim, 1991; Суботић, 2004; Vasić et al., 2008. For the principles of designing a minimal dictionary of Serbian as L2, see Dražić, 2008; for a model of a learner's explanatory-combinatorial dictionary, see Milićević, 2008; and on Serbian as L2, in general, see Дешић, 2007; Golubović and Raecke, 2008.

[2] Neither the still unfinished, massive Dictionary of the Serbian Literary and Folkloric Language (САНУ, 1959), nor the six-volume Dictionary of the Serbocroatian Literary Language (MC, 1967), nor its recent one-volume counterpart (MC, 1997), have been made available in the electronic format, whether online or offline.

[3] See http://wordnet.princeton.edu/perl/webwn

[4] There are two general approaches to building cross-lingual Wordnets: the so-called "merge approach" whereby a specific language WN is built from scratch and then aligned to other Wordnets via an Inter-Lingual Index (ILI) (Vossen, 1998), and the so-called "expand approach," whereby one starts from the English WN, then maps meanings of a different language to it (Bentivogli et al., 2002). Our model of creating the Transpoetika Wordnet is informed by the "expand approach".

[5] On the mental lexicon see Singleton, 1999; Aitchison, 2003; Bonin, 2004.

[6] WN defines nouns, verbs, adjectives and adverbs, but leaves out all function words. Psycholinguistic studies have shown that function words are stored in the mind not as lexical, but syntactic items (Garret 82).

[7] If three words (W_1, W_2, W_3) are synonymous, an alphabetic print thesaurus would have to have them listed three times, under entries for each member of the set of synonyms. In a topical thesaurus, a user would have to look up a word in an index, and only then find the word in the thesaurus proper.

[8] Because of this basic requirement of systematicity – which traditional dictionaries do not have to observe – Wordnet sometimes expresses concepts in terms of artificial lexical items as a node in its network, for example "bad person". It uses compound terms (such as "Greek deity" or "animal product"), which could be predicted from their constituent elements, but they serve an important organizational role "by adding much needed structure to the middle ontology" (Hayes et al., 2005).

Bibliography

Aitchison, Jean. 2003. *Words in the Mind: An Introduction to the Mental Lexicon*. Malden, MA: Blackwell Publishing.

Bentivogli, Luisa, and Emanuele Pianta. 2000. Coping With Lexical Gaps When Building Aligned Multilingual Wordnets. *LREC-2000* 993-997.

Bentivogli, Luisa, Emanuele Pianta, and Christian Girardi. 2002. MutiWordNet: Developing an Aligned Multilingual Database. In *1st International Global WordNet Conference, January 21-25, 2002: Proceedings*, Mysore: Central Institute of Indian Languages.

Bentivogli, Luise, and Emanuele Pianta. 2003. Beyond Lexical Units: Enriching Wordnets With Phrasets. *Research Note Sessions of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL'03)* 67–70.

Boeckx, Cedric. 2006. *Linguistic Minimalism: Origins, Concepts, Methods and Aims*. Oxford linguistics. Oxford: Oxford University Press.

2004. *Mental Lexicon: Some Words to Talk About Words*. Edited by Patrick Bonin. Hauppauge, N.Y: Nova Science Publishers.

Burkhanov, Igor. 2003. Pragmatic specifications: Usage indications, labels, examples; dictionaries of style, dictionaries of collocations. In *A Practical Guide to Lexicography*, ed. Piet Van Sterkenburg Sterkenburg, 6, 102-111. Amsterdam; Philadelphia: John Benjamins Pub.

Carter, Ronald. 1998. *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.

Cowie, Anthony Paul, ed. 1987. *The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985*. Tübingen: Niemeyer.

Dražić, Jasmina. 2008. Principles for the Design of the Monolingual Minimal Dictionary of Serbian as a Foreign Language. In *Bosnisch. Kroatisch. Serbisch. B.K.S. als Fremdsprachen an den Universitäten der Welt*, eds. Biljana Golubović, and Jochen Raecke, 43-50. München: Verlag Otto Sagner.

Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Mass: MIT Press.

Golubović, Biljana, and Jochen Raecke, eds. 2008. *Bosnisch. Kroatisch. Serbisch. B.K.S. als Fremdsprachen an den Universitäten der Welt*. München: Verlag Otto Sagner.

Hartmann, R. R. K. 1994a. Bilingualised Versions of Learners' Dictionaries. In *Wörterbücher und ihre Benutzer*, eds. Ekkehard Zöfgen, and Gert Henrici, 206-219.

Hartmann, R. R. K. 1994b. The Learner's Dictionary - Unilingual or Bilingual? In *Entering Text*, eds. Lynne Flowerdew, and Anthony K. K Tong, 239-246. Hong

Kong: Language Centre, The Hong Kong University of Science and Technology.

Hayes, Jer, Nuno Seco, and Tony Veale. 2005. Creative Discovery in the Lexical "Validation Gap". *Computer Speech & Language* 19, no. 4: 513-523.

Hockey, Susan M. 2000. Dictionaries and Lexcial Databases. In *Electronic texts in the humanities: Principles and Practice*, 146-171. Oxford & New York: Oxford University Press.

Hunston, Susan, and Gill Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Studies in corpus linguistics. Amsterdam: Benjamins.

Kirkpatrick, Betty. 1985. A Lexicographical Dilemma: Monolingual Dictionaries for the Native Speaker and for the Learner. In *Dictionaries, Lexicography, and Language Learning*, ed. Robert Ilson, 7-13. Oxford; New York: Pergamon Press in association with the British Council.

Knowles, Francis E. 1989. Computers and Dictionaries. In *Wörterbücher: ein internationales Handbuch zur Lexikographie*, eds. Franz Josef Hausmann, Oskar Reichmann, and Herbert Ernst Wiegand, 1, 1645-1672. Berlin; New York: W. de Gruyter.

Krstev, Cvetana, Ivan Obradović, and Duško Vitas. 2006. Developing Balkan Specific Concepts Within BalkaNet - A Multilingual Database of Semantic Networks. *5th International Conference Formal Approaches to South Slavic and Balkan Languages, FASSBL 2006* 94-98.

Krstev, Cvetana, Duško Vitas, and Gordana Pavlović-Lažetić. 2008. Resources and Methods in Morphosyntactic Processing of Serbo-Croatian. In *Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003*, ed. Gerhild Zybatow, 3-17. Frankfurt am Main; New York: Peter Lang.

eds. 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Hove, England: Language Teaching Publications.

Lewis, Michael. 1993. *The Lexical Approach: The State of ELT and a Aay Forward*. Hove, UK: Language Teaching Publications.

Marmaridou, Sophia. 2000. *Pragmatic Meaning and Cognition*. Pragmatics & Beyond. Amsterdam ; Philadelphia: J. Benjamins Pub. Co.

Meijs, Willem. 1992. Computers and Dictionaries. In *Computers and Written Texts*, ed. Christopher Butler, Applied Language Studies, 141-165. Oxford, UK & Cambridge, USA: Blackwell.

Milićević, Jasmina. 2008. Towards an Explanatory-Combinatorial Dictionary for BCS Learners. In *Bosnisch. Kroatisch. Serbisch. B.K.S. als Fremdsprachen an den Universitäten der Welt*, eds. Biljana Golubović, and Jochen Raecke, 155-168. München: Verlag Otto Sagner.

Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. "Introduction to WordNet: An Online Lexical Database." *Five papers on WordNet*. ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps (accessed December 1, 2008).

Nakamoto, Kyohei. 1994. Monolingual or Bilingual, that is NOT the Question: the "Bilingualised" Dictionary. *Iwasaki Linguistic Circle* 24,

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. New York: Newbury House Publishers.

Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge; New York: Cambridge University Press.

Obradović, Ivan, and Ranka Stanković. 2008. Software Tools for Serbian Lexical Resources. *Infotheca* IX, no. 1-2: 43a-57a.

Ordan, Noam, and Shuly Wintner. 2005. Representing Natural Gender in Multilingual Databases. *International Journal of Lexicography* 18, 357-370.

Partington, Alan. 1998. *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Studies in corpus linguistics. Amsterdam: Benjamins.

Pujol, Dídac, Montse Corrius, and Joan Masnou. 2006. Print Deferred Bilingualised Dictionaries and their Implications for Effective Language Learning: A New Approach to Pedagogical Lexicography. *International Journal of Lexicography* 19, no. 2: 197-215.

Reif, J. A. 1987. The Development of a Dictionary Concept: an English Learners' Dictionary and an Exotic Alphabet. In *The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1-3 April 1985*, ed. Anthony Paul Cowie, Lexicographica. Series maior, 17, 146-158. Tübingen: Niemeyer.

Schmitt, Norbert. 2000. *Vocabulary in Language Teaching*. New York: Cambridge University Press.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Describing English language. Oxford: Oxford University Press.

Sinclair, John. 2000. Lexical Grammar. *Darbai ir dienos* 24,

Sinclair, John McHardy, and Ronald Allan Carter. 2004. *Trust the Text Language, Corpus and Discourse*. London: Routledge.

Sinclair, John McHardy, and Anna Mauranen. 2006. *Linear Unit Grammar: Integrating Speech and Writing*. Studies in corpus linguistics. Amsterdam: Benjamins.

Singleton, David. 1999. *Exploring the Second Language Mental Lexicon*. Cambridge; New York: Cambridge University Press.

Stamou, S, K Oflazer, K Pala, D Christoudoulakis, D Cristea, D Tufis, S Koeva, G Totkov, D Dutoit, and M Grigoriadou. 2002. BALKANET A Multilingual Semantic Network for the Balkan Languages. *Proceedings of the International Wordnet Conference, Mysore, India* 21-25.

Stein, G. 1990. From the Bilingual to the Monolingual Dictionary. In *BudaLEX '88 Proceedings: Papers from the 3rd International EURALEX Congress, Budapest, 4-9 September 1988*, eds. Tamás Magay, and J Zigány, 401-407. Budapest: Akadémiai Kiadó.

Thumb, Jenny. 2004. *Dictionary Look-Up Strategies and the Bilingualised Learner's Dictionary: A Think-Aloud Study*. Tübingen: M. Niemeyer.

Tufis, D., D. Cristea, and S. Stamou. 2004. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. *SCIENCE AND TECHNOLOGY* 7, no. 1-2: 9-43.

van Ek, Jan Ate, and J. L. M. Trim. 1991. *Threshold Level 1990*. Strasbourg: Council of Europe Press.

Vasić, Vera, Vladislava Ružić, and Ljiljana Subotić. 2008. Serbian Language Threshold Level: Concept and Design. In *Bosnisch. Kroatisch. Serbisch. B.K.S. als Fremdsprachen an den Universitäten der Welt*, eds. Biljana Golubović, and Jochen Raecke, 195-202. München: Verlag Otto Sagner.

Vitas, Duško, Cvetana Krstev, Ivan Obradović, and Gordana Pavlović-Lažetić. 2003. Processing Serbian Written Texts: An Overview of Resources and Basic Tools. In *Workshop on Balkan Language Resources and Tools, 21 November 2003*, eds. S. Piperidis, and V. Karkaletis, Thessaloniki, Greece:.

Vitas, Duško, Cvetana Krstev, Gordana Pavlović-Lažetić, and Goran Nenadić. 1998. *Recent Results in Serbian Computational Lexicography*. Belgrade: Faculty of Mathematics, University of Belgrade.

Vossen, Piek, ed. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht [The Netherlands]; Boston: Kluwer Academic.

Whitcut, Janet. 1985. Usage Notes in Dictionaries: The Needs of the Learner and the Native Speaker. In *Dictionaries, Lexicography, and Language Learning*, ed. Robert Ilson, 75-80. Oxford; New York: Pergamon Press in association with the British Council.

Wiegand, Herbert Ernst. 1985. Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzung einsprachiger Wörterbu̇cher. In *Lexikographie und Grammatik: Akten d. Essener Kolloquiums zur Grammatik im Wörterbuch 28. - 30.6.1984*, ed. Henning Bergenholtz, Tubingen: Niemeyer.

Wierzbicka, Anna. 1992. *Semantics, Culture and Cognition: Universal Human Concepts in Culture-Specific Configurations*. New York: Oxford University Press.

Апресян, Ю. Д. 1988. Прагматическая информация для толкового словаря. In *Прагматика и проблемы интенсиональности (Сб. науч. тр.)*, ed. Нина Давидовна Арутюнова, Москва: Институт языкознания АН СССР.

Дешић, Милорад, ed. 2007. *Српски као страни језик у теорији и пракси: зборник радова*. Београд: Министарство просвете и спорта Републике Србије; Филолошки факултет, Центар за српски као страни језик.

Крстев, Цветана. 2004. Специфични концепти Балкана у семантичкој мрежи Wordnet. In *Сусрети култура*, ed. Љиљана et al. Суботић, 275-285. Нови Сад: Универзитет у Новом Саду, Филозофски факултет.

МС. 1967. *Речник српскохрватскога књижевног језика*. Нови Сад: Матица српска.

МС. 1997. *Речник српскога језика*. Нови Сад: Матица српска.

САНУ. 1959. *Речник српскохрватског књижевног и народног језика*. Београд: Институт за српскохрватски језик.

Суботић, Љиљана. 2004. Праг знања и речник (историјат проблема). *Научни састанак слависта у Вукове дане: Граматички опис српског језика* 33, no. 1: 167-177.

Шведова, Наталья. 1988. Парадоксы словарной сатьи. In *Национальная специфика языка и ее отражение в нормативном словаре. Сборник статьей*, ed. Юрий Николаевич Караулов, 6-11. Москва: Наука.