

Danko Šipka

## ON THE CLASSIFICATION OF HOMONYMS\*

### 0. Introduction

The issue of homonyms is an ever present subject of interest in linguistics. As can be seen in Ivić (1990), homonymy has been used as an argument in the discussion between so-called anomalists and analogists, as early as in ancient Greek tradition. One way or another, homonyms were used in various linguistic approaches, e.g. linguistic geography, glossematics, generative theories, and others. Furthermore, the monographs and dictionaries of homonyms, such as Philoponi's dictionary (Philoponi 1983), dating probably from 6 a.d, as well as the monograph by Sosnecki (1874) on Russian homonyms, reveal early and continued presence of this issue in linguistics. White's (1981) agricultural paper clearly illustrates the interest among non-linguists.

Careful analyses of the relevant literature (cf. Šipka 1991) evidenced nine central problems in the research of homonymy, as follows:

1. definition of homonymy,
2. homonymy vs. polysemy,
3. lexical homonymy vs. homonymy at other levels,
4. classification of homonymy,
5. causes of homonymy,
6. differentiation of the homonymic lexemes in text,
7. use of homonyms (e.g. in poetry),
8. role of homonymy within the lexicon,
9. language reactions to homonymic conflict.

---

\* This paper is based upon a part of my doctoral dissertation "Lexical homonymy in contemporary standard Serbo-Croatian" (defended at the University of Belgrade in May 1989). I would like to express my gratitude both to the Alexander von Humboldt Foundation, which made the writing of the paper possible, and Prof. Leonhard Lipka, of the University of Munich, my adviser during the Humboldt Post-Doctoral Program, for his support and various useful comments on a former version of the manuscript.

In this paper we will restrict our discussion to just one of the above problems - the classification of homonymy. For our purposes, therefore, even a very loose, and thus inaccurate, definition is sufficient that homonymy is a lexical relation between two lexemes having (partially) identical form and different meaning.

### 1. Existing classifications

Two major approaches to the classification of lexical homonyms can be identified in the relevant literature. The first and dominant one is based upon systemic features of the homonymic lexemes; the second, however, concentrates on their other properties or classifies them according to the specific theory developed by the scholar in question.

Within the first approach the following categories can be distinguished:

- a) full homonymy,
- b) homomorphy,
- c) homography,
- d) homophony.

Depending on whether they distinguish all four, or only some of the categories stated above, scholars can be divided into the following four groups:

- 1) all four categories are distinguished,
- 2) only full homonymy vs. homomorphy,
- 3) full homonymy, homography, and homophony,
- 4) full homonymy (or full homonymy and homophony) vs. homography.

The first group can be illustrated by the following passage from Shanskii (1969:16f):

"Homonyms are words which sound identical but which have completely different meanings... Homomorphs, in contrast to homonyms, may or may not be the same parts of speech; they coincide in sound, but always in isolated forms only... Homophones resemble homonyms and homomorphs in that they too are words and forms of different meanings which are pronounced identically, but differ from these in that they are written differently... Homographs, which are words and forms which are different in meaning but written alike, must be distinguished from homonyms, homomorphs, and homophones, which are words and forms which are different in meaning but sound alike."

Homonymy is treated in this manner in Galkina-Fedoruk (1954), Winkel (1958), Mel'nikova (1974), Fomina (1978), Singh (1980), Dešić (1982), Gortan-Premk (1984), Kaliszan (1997) and by many other scholars. This approach is doubtless the dominant one.

The treatment formulated in 2 can be found in Trnka (1931), and Hervey (1978:80):

"Homonymy may be total or partial, depending on whether all the morphs of sign x have their corresponding homomorphs among the morphs of sign y, or whether some of the morphs of sign x are not homomorphs of any of the morphs of the sign y, or vice versa."

The third position is advocated by Tyšler (1963), Franklyn (1966), and Truby (1966:191f):

"Words, parts of words, and/or combinations of words which sound alike and are spelled alike and differ as to semantic content (where this last is relevant) are HOMONYMS... Words, parts of words, and/or combinations of words which sound alike but are spelled differently and differ as to semantic content (where this last is relevant) are HOMOPHONES... Words, parts of words and/or combinations of words which are spelled alike but do not sound alike and which differ as to semantic content (where this last is relevant) are HOMOGRAPHS."

The group 4 can be observed with Knežević (1970), and Salomon (1966:9):

"Words that are pronounced alike, regardless of how spelled, are called homophones; those that are spelled alike, regardless of how pronounced, are called homographs..."

The second approach, having other underlying criteria, rather than systemic properties of homonymic lexemes can be subdivided into two groups, the first using other lexemic properties, for example etymology, the second using a specific theory as their classifying criterion.

The first group is advocated in Kuznecova (1982), and Akhmanova (1974), the second by Maslov (1963), and Belokrinickaja (1960).

The approaches outlined above, can be summarized as in diagram (1):

(1)

approach	1 systemic features	2 other
group	<ul style="list-style-type: none"> <li>— 1.1 all four categories</li> <li>— 1.2 full homonymy vs. homomorphy</li> <li>— 1.3 full homonymy, homography, homophony</li> <li>— 1.4 full homonymy vs. homography</li> </ul>	<ul style="list-style-type: none"> <li>— 2.1 other features</li> <li>— 2.2 specific theories</li> </ul>

As already mentioned, the classification 1.1 is the most commonly used. It will thus be the primary subject of our critical remarks.

It can be observed that (although the Slavic literature prevails) the classifications discussed above pertain to various languages, and therefore cannot be considered language-specific.

## 2. Critical remarks

Existing classifications are a natural result of the somewhat indistinct position of lexicology towards lexicography, as it was till the early 1970s. The categories distinguished were directly caused by the lexicographic treatment of homonymy. After Zgusta's (1971) trail-blazing *Manual of Lexicography*, the two disciplines managed to obtain their autonomous positions (with continued cooperation, of course), as it can be clearly observed with the recent lexicologic crescendo, Lipka's (1990) *Outline of English Lexicology*. Therefore, a new, more "lexicologic" classification is needed.

From the lexicologic standpoint, the following, purely theoretical, remarks can be made on the existing lexicologic classifications.

1. They do not distinguish between the homonymy within the *lexicon*, as a part of Saussurian *langue*, or Chomskian *competence*, and homonymy within a *text*, as its concrete realization, as a part of Saussurian *parole*, or Chomskian *performance*.

2. They, furthermore, do not stress the difference between really existing homonymy in the language and its realizations on one side, and technical treatment of homonymic lexemes in dictionaries (as idealized, one-sided projections of the lexicon) on the other.

3. They are, finally, one-dimensional in the sense that three different and level-specific criteria (i.e. paradigmatic, graphic, phonetic features) are used as if they were one-level criteria, which results in the discernment of only four, rather than 12 categories, as it will be shown further on, (i.e. full homonymy, homomorphy, homography, homophony).

This theoretical inadequacy leads into two major practical disadvantages:

a) a number of instances of homonymy remains unaccounted (for example the lexemes coinciding in some paradigmatic form other than in the so-called canonical dictionary form), or only partially accounted (for example the lexemes that are homomorphs and homophones at the same time);

b) the existing classifications are of little use for the textual solution of homonymic clashes; the information, for example, that two lexemes are homomorphs, does not help to discern the two coinciding forms, for they coincide, irrespective of whether the lexemes share the other paradigmatic forms or not.

### **3. A proposal of classification**

In order to overcome the disadvantages summarized above, we will propose a new, more complex, but hopefully more universal and useful classification.

We will use mostly Serbo-Croatian examples. The difference between universal and language-specific properties of the classification will be appropriately stressed.

The term *homonymy* denotes at least three phenomena:

- a) a state or situation,
- b) a relation between two or more lexemes,
- c) a property of a lexeme.

This holds for all types of homonymy to be classified as follows.

At the first level three types of *homonymy* should be distinguished:

- 1) *homonymity* (i.e. homonymy within the lexicon),
- 2) *homonymicity* (i.e. homonymy in a text),
- 3) *dictionary homonymy*.

The differences between these three types of homonymy can be summarized as in the diagram (2):

(2)

type/attribute	1	2	3	4
homonymity	+	-	-	-
homonymicity	-	+	+	-
dictionary homonymy	-	-	+	+

+/- presence or absence of an attribute,

1 = all paradigmatic forms are involved,

2 = two elements can exist in one form,

3 = limitations due to the scope,

4 = limitations due to the structural features.

The lexicon contains all paradigmatic forms of a lexeme, so there is a possibility for all of them to be involved in homonymic relations. In a text and a dictionary, on the contrary, only the forms actually present in it can be homonyms of each other. The attribute 1, therefore, can be assigned only to homonymity.

Only homonymicity can be assigned the attribute 2, i.e. the possibility that two or more lexemes are present in just one physical form, as in the car commercial slogan: *Designed for the human race* (*race* 'a group of humans, sharing certain bodily features' and *race* 'a competition in speed'). This is impossible both in the lexicon and the dictionary, where each lexeme has its own separate forms.

There is no dictionary or text which can contain all lexemes existing within the lexicon. They are, thus, limited in their scope, and therefore can be assigned the attribute 3.

Besides being limited in their scope, dictionaries can also be limited by their structural characteristics. Dictionaries normally contain only the so-called "canonical forms", and thus exclude (as a possibility, as well) all homonymic pairs coinciding in a non-canonical form (such as in Eng. *rose* 'a flower' : *rose*, the verb *rise* in Past Tense, as a non-canonical form). The attribute 4, therefore, can be observed only with dictionary homonymy.

It should be quite clear that the three homonymy types distinguished differ both in qualitative and quantitative manner, and thus should be understood and treated as separate phenomena.

Further classification of dictionary homonymy does not substantially differ from the one applied to homonymy. We will, therefore, provide a comprehensive insight into sub-classes of homonymy, specifying the differences towards dictionary homonymy at the appropriate places.

### 3.1 Homonymy

As already pointed out, *homonymy* pertains to the lexemes within the lexicon. In our further discussion, we will use two different criteria, resulting in two possible, compatible classifications. Homonymy can be classified according to:

- 1) the inherent systemic features, i.e. which homonymic relations do the lexemes develop within the lexicon, and
- 2) their likelihood of forming homonymic relations in a text, i.e. their systemic features with respect to their possible usage.

The first classification can be summarized in diagram (3).

(All examples are Serbo-Croatian. We present, and in further text mention homonymic pairs, since each group consisting of more than two homonymic lexemes can be broken into the binary relations - A:B:C = A:B, A:C, B:C.)

(3)

#### 1 *absolute homonymy*

##### 1.1 *full absolute homonymy*

- *ātlās* (textile) and *ǎtlās* (book of maps)
- *bāza* (base) and *bāza* (elder)
- *vāl* (wave) and *vāl* (veil)

##### 1.2 *homographity*

- *lūk* (onion) and *lūk* (bow)
- *jārica* (young female goat) and *jārica* (early wheat)
- *bēdrenica* (cattle disease) and *bedrēnica* (sabre)

##### 1.3 *homotony*

- *Ātlās* (mountain) and *ǎtlās* (book of maps)
- *Mārica* (proper name) and *mārica* (marijuana)
- *jāt* (ancient Slavic vowel) and *JĀT* (Yugoslav Airlines)

##### 1.4 *homophonity*

- *Āzija* (continent) and *āzija* (letter)
- *NĪN* (Yugoslav weekly) and *Nĭn* (town in Croatia)
- *Līka* (region in Croatia) and *līka* (bast)

## 2 partial homonymy

### 2.1 homoformity

#### 2.1.1 full homoformity

- *stô, oia* m (table) and *stô* (hundred)
- *pût, a* m (road) and *pût, i* f (carnation)
- *ôstar, tra, tro* (sharp) and *ôstar, a, o* (rather old)

#### 2.1.2 homographity

- *bésan, sna, sno* (rabid) and *bésan, a, o* (insomnic)
- *sád, a* m (orchard) and *sād* (now)
- *jâ* (I) and *jă* (yes)

#### 2.1.3 homotony

- *sîv, a, o* (gray) and *SîV* (Federal Executive Council)
- *Sîk, a* m (Indian ethnic group) and *sic* [sîk] (sic)
- *îris* f (proper name) and *îris, a* m (iris)

#### 2.1.4 homophony

- *sîv* (gray) and *SîV* (Federal Executive Council)
- *Gr̥k, a* m (Greek) and *gr̥k, a, o* (bitter)
- *òno* (civil defence) and *òno* (it)

### 2.2 homomorphity

#### 2.2.1 full homomorphity

- *bája* f (bug) and *bája* m (brother)
- *bríca* f (razor) and *bríca* m (barber)
- *păša* f (pasture) and *păša* m (pasha)

#### 2.2.2 homographity

- *băka* (soldier) and *băka* (grandmother)

#### 2.2.3 homotony

- *jóva* f (sallow) and *Jóva* m (proper name)
- *íva* f (sallow) and *íva* m (proper name)
- *O* [ô] (oxygen) and *ô* (oh)

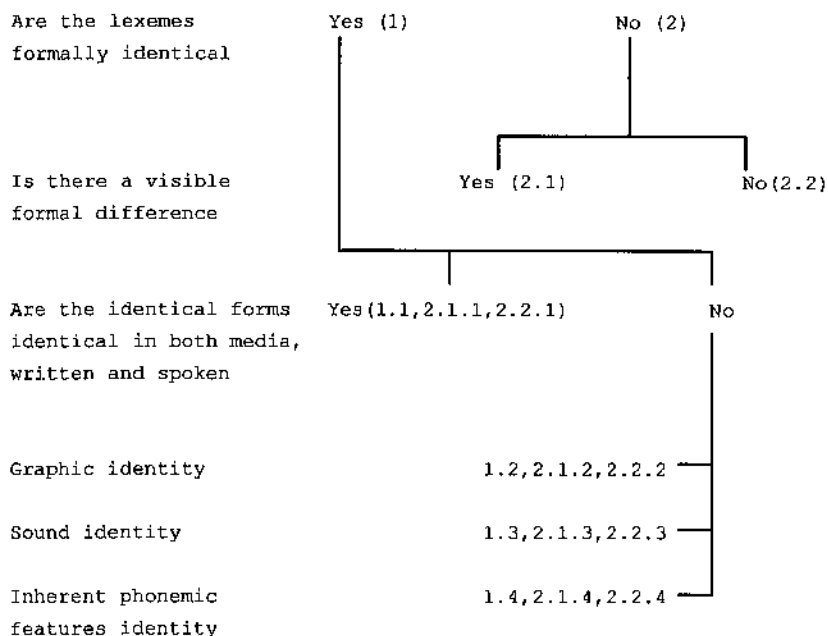
#### 2.2.4 homophony

- *íva* f (tree) and *Íva* m (proper name)
- *K* [kâ] (kilobyte) and *ka* (toward)
- *O* [ô] (oxygen) and *o* (about)

The classification has the underlying criteria, as described in (4).



(4)



Basic split accounts for the fact that certain homonymic pairs are not formally identical. These pairs are traditionally called *homomorphs*. Here, however, we distinguish two types of formally non-identical homonymic pairs, or what we here call *partial homonymity*: *homoformity*, where the pairs differ in respect of paradigmatic forms (such as noun cases, or verb tenses), and *homomorphity*, where the pairs coincide in all paradigmatic forms, but differ as to their other grammatical properties (such as noun gender, which causes the difference in congruence: NOM *brica*, GEN *brice*, DAT *brici*... 'razor': NOM *brica*, GEN *brice*, DAT *brici*... 'barber', but: *oštra brica* 'sharp razor': *oštar brica* 'sharp barber').

Further improvement in our classification when compared with the existing ones is that both the *absolute* and the two types of *partial homonymity* can be further discerned according to their identity in written and spoken form, which is traditionally called *homography* and *homophony*. Here we can account for the cases where the lexemes coincide both in written and spoken form (*full absolute homonymity*, *full homoformity*, *full homomorphity*), cases where they coincide only in their written form (*homography*), only in their spoken form (*homotony*), or when they share only inherent phonemic features (*homophony*).

The last category is another innovation. The identity of inherent phonemic features (with the difference in prosody, as well as graphical form), gives the lexemes in question the ability to form a homonymic clash in some special instances, for example in uppercase printed text or at the beginning of a sentence if one lexeme has an uppercase initial, and there are no further graphemic differences. They thus form a specific relation of homonymy in the lexicon.

The second classification of homonymy, the one which has in mind the ability of homonymic pairs to create homonymic clashes in a text, can be represented in diagram (5).

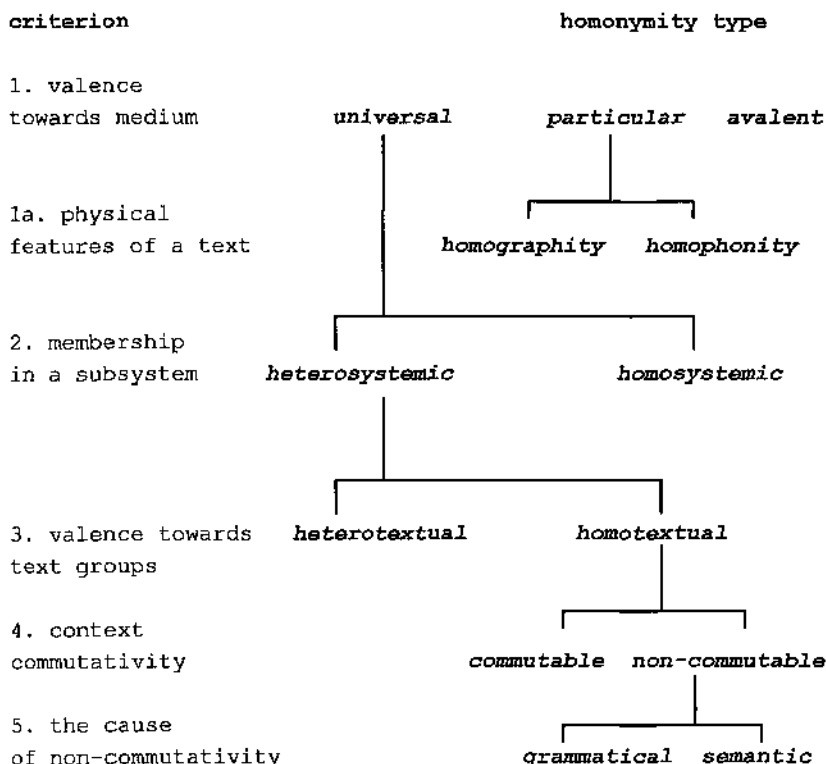
1) *Universal homonymy* can be realized in both media, spoken or written (e.g. S-Cr *lûg* /ashes/: *lûg* /grove/), *particular homonymy* can be realized only in the spoken medium (*homophony*, e.g. S-Cr *Âzija* /Asia/: *âzija* /badge/) or in the written medium (*homography*, e.g. S-Cr *lûk* /onions/ and *lûk* /bow/). *Avalent homonymy* cannot be realized, except in some specific cases, for instance S-Cr *sânjâ* /3.p.sg.Prae. of the verb *sanjati* - to dream/ and *Sânja* /a proper name/ can be realized only at the beginning of the sentence in written text, e.g. SANJA SANJA SEDMICU /Sanja is dreaming seven (correct numbers in lotto)/.

2) Lexemes can belong to the same subsystem (*homosystemic lexemes*, e.g. *lûg* : *lûg* - both lexemes are contemporary), or they can be members of two different subsystems (*heterosystemic type*, e.g. *čêst* /'frequent' - contemporary/ : *čêst* /'honor' - obsolete/).

3) The above mentioned pair *čest* is *heterotextual*, since the latter lexeme is avalent towards newspapers, textbooks, etc. *Homotextual* homonymy is rather rare. Such a pair is, for example, *bîti* /to be/: *bîti* /to beat/.

4,5) If the pair is homotextual, the lexemes can be *commutable*, such as *òdvesti* /to give a lift/ and *òdvesti* /to show the way/ in the sentence *Ja ću vas odvesti do pošte* /I'll give you a lift / I'll take you to the post office/ or *non-commutable* - semantically (e.g. *kòsa* /hair/: *kòsa* /scythe/ in the sentence *Ispustio sam kosu* /I dropped a scythe, I dropped hair is nonsense/) or grammatically (like the above-mentioned pair *bîti* in the sentence *On će biti predsjednik* /He is going to be a president/, the noun *predsjednik* is in the nominative case, with the verb *biti* /to beat/ the form would be accusative - *On će biti predsjednika*).

(5)



The main advantage of this classification is that it can take into account the relations between the lexemes and their usage, which was not the case with the existing models. It can, therefore, also be used as an underlying pattern for the automatic disambiguation based on usage variations of lexemes, to be discussed later on.

### 3.2 Dictionary homonymy

Dictionary homonymy pertains to dictionary relations. It has the same types as systemic homonymity, but it is applied only to the canonical forms. For example, the above-mentioned forms *kosa* : *kosa*, are treated as full homonymy even though their stress differs in the accusative singular; the forms *sam* /I am/

and *sâm* /alone/, however, are not registered in a dictionary since their canonical forms differ - *biti* : *sâm*.

### 3.3 Homonymicity

Homonymicity pertains to the usage of lexemes. Lexemes can be realized **monoformically** (lexemes are condensed in one form, e.g. *Početak najdužeg leta* /Beginning of the longest flight or summer, *leta* is a condensed form of both 'flight' and 'summer'/) or **heteroformically**, e.g. the above-mentioned pair SANJA SANJA SEDMICU /Sanja is dreaming seven, both lexemes are in their separate forms./

### 4. Advantages and applications

Using the classification described in 3, both the theoretical inadequacies and practical disadvantages outlined in 2, can be avoided. The proposed classification has in mind the difference between the lexicon and text, as well as between the lexicon and dictionary. It is, furthermore, multidimensional. Consequently, it can classify all borderline cases and has a much higher degree of applicability.

Furthermore, this classification could throw some more light on our understanding of the mental lexicon. There are many more links (in the sense of Aitchison 1987) between homonymic lexemes in the mental lexicon, than one should think based on the classifications described in 1. Moreover, the proposed classification stresses some important factors in one of the processes within the mental lexicon - the process of semantic interpretation.

The most direct and obvious application can be found within computational linguistics, and the process of disambiguation. Homonymicity, namely, can be viewed as a subclass of ambiguity.

The distinction between homonymity and homonymicity stresses that in the construction of a lexicon (for example in form of a database) one has to take into account quite different factors, when compared with those needed in text-analysis.

The classification of homonymity according to their systemic features reveals the fact that the range of situations and lexemic features that are to be taken into account varies significantly. Thus, for example, if we construct the database to be used in disambiguation of written text, we can disregard all pairs which do not coincide in their written forms. In speech to text conversion, however, one has to take care of all existing pairs. The category of homomorphity can be used as the basis for the contextual disambiguation, if for example we have a noun that can be detected using its adjectival modifier (e.g. in the above-mentioned

S-Cr homomorphic pair *bríca* 'barber' and *bríca* 'razor', using adjectives *dobar* to recognize the former, and *dobra* to recognize the latter member).

The second classification of homonymity might be most useful. It can be used as the underlying pattern of a disambiguating algorithm which, matching lexemic with textual usage labels discerns the homotextual from heterotextual pairs, so that the latter (most frequent ones) are disambiguated without searching for the contextual clues, while the former have to be solved according to the contextual commutativity. This would bring some advancement in the process of disambiguation, which is normally carried out with respect to contextual factors only.

Finally, the two types of homonymicity reveal that besides disambiguation, there is another situation to be taken care of in the process of semantic interpretation, namely, the heteroformic homonymicity.

### Bibliography

- Aitchison, J. 1987, *Words in the Mind. An Introduction to the Mental Lexicon*, Oxford-New York: Basil Blackwell.
- Akhmanova, O.S. 1974, *Slovar omonimov russkogo jazyka*, Moskva: Izdatel'stvo Sovetskaja enciklopedija.
- Belokrinickaja, S.S. 1960, "Različnye tipy omonimii i sposoby ich razgraničenja pri mašinnom perevode (na materiale anglijskogo, nemeckogo, russkogo, kitajskogo i japonskogo jazykov)", *Voprosy jazykoznanija*, No 2, 97-101.
- Dešić, M. 1982, "Polisemija i homonimija u rječnicima savremenog srpskohrvatskog jezika", *Naš jezik* n.s. XXV/4-5, 231-241.
- Fomina, M.I. 1978, *Sovremennyj russkij jazyk. Leksikologija*, Moskva: Vysšaja škola.
- Franklyn, J. 1966, *Which witch? being a grouping of phonetically compatible words*, London: Hamich Hamilton Ltd.
- Galkina-Fedoruk, E.M. 1954, "K voprosu ob omonimach v russkom jazyke", *Russkij jazyk v škole*, No 3, 14-19.
- Gortan-Premk, D. 1984, "Polisemija i homonimija u srpskohrvatskom jeziku", *Južnoslovenski filolog*, XL, 11-19.
- Hervey, S. 1978, *Axiomatic Semantics. A theory of linguistic semantics*, Edinburgh: Scottish Academic Press.
- Ivić, M. 1990, *Pravci u lingvistici*, Beograd: Nolit.
- Kaliszan, J. 1997, *Omonimija proizvodnych slov v russkom jazyke*, Poznań: UAM.

- Knežević, A. 1970, *Homophone und Homogramme in der Schriftsprache der Kroaten und Serben*, Meisenheim am Glan: Verlag Anton Hain.
- Kuznecova, E.V. 1982, *Leksikologija ruskog jezika*, Moskva: Vysšaja škola.
- Lipka, L. 1990, *An Outline of English Lexicology. Lexical Structure, Word Semantics, and Word-Formation*, Tübingen: Max Niemeyer Verlag.
- Maslov, Ju.S. 1963, "Omonimy v slovarjach i omonimija v jazyke (k postanovke voprosa)", *Voprosy teorii i istorii jazyka*, Leningrad: Izdatel'stvo Leningradskogo Universiteta, 198-202.
- Meľnikova, A.M. 1974, "K voprosu o russkich omografach", *Russkij jazyk v škole*, No 4.
- Philophoni, I. 1983, *De vocabulis quae diversum significatum exhibent secundum differentiam accentus* (edited by Lloyd W. Daly), Philadelphia: American Philosophical Society.
- Salomon, L.B. 1966, *Semantics and Common Sense*, New York: Holt, Rinehart and Winston.
- Shanskii, M.N. 1969, *Russian Lexicology*, London: Pergamon Press Ltd.
- Singh, R.A. 1980, *An Introduction to Lexicography*, Mysore: Central Institute of Indian Languages.
- Šipka, D. 1990, *Leksička homonimija (na primjeru savremenog srpskohrvatskog standardnog jezika)*, Sarajevo: Institut za jezik i književnost.
- Sosnecki, M. 1874, *Opyt omonimov v russkom jazyke*, St. Peterburg.
- Trnka, B. 1931, "Bemerkungen zur Homonymie", *Travaux du Cercle Linguistique de Prague*, 4.
- Truby, H.M. 1966, "The Homoneme", *Word* (Journal of the Linguistic Circle of New York), vol. 22, numbers 1-2-3.
- Tyšler, I.S. 1963, *Slovar omonimov sovremennogo anglijskogo jazyka*, Saratov: Izdatel'stvo Saratovskogo universiteta.
- White, R.E. 1981, "Homonymy in word species group names of Criocerinae (Coleoptera: Chrysomelidae)", *United States Dept. of Agriculture Technical Bulletin*, number 1629.
- Winkel, H.J. 1958, *Über die Homophonie in der russischen Literatursprache*, Meisenheim am Glan: Verlag Anton Hain.
- Zgusta, L. 1971, *Manual of Lexicography*, The Hague: Mouton.