Journal of Siberian Federal University. Humanities & Social Sciences 3 (2015 8) 533-541
~ ~ ~

УДК 81’44

A Typology of Linguistic Ambiguity

Marina A. Yuzhannikova*
Siberian Federal University
79 Svobodny, Krasnoyarsk, 660041, Russia

Received 21.12.2014, received in revised form 16.01.2015, accepted 30.01.2015

Although ambiguity is an integral part of any natural language up to now its study was focused mainly on ambiguity resolution. However, there is a growing interest to various linguistic typologies for development of new tools for automatic language analysis. This article focuses on the development of a typology for dealing with linguistic ambiguity. The first section of this paper discusses the problem of ambiguity classification and is illustrated with relevant examples. The second part of the article provides the complete typology of ambiguity based on Russian corpora, and it is presented within the table of complex ambiguity description. The typology consists of nine columns of extralinguistic information and twelve of linguistic description. As a result the typology allows to make a complete analysis of each example of ambiguity regardless of its intentionality.

Keywords: ambiguity, classification, typology, vagueness, polysemy, homonymy.

The article is written within PhD practice at the University of Barcelona, Spain (the ERANET MUNDUS project).

Research area: philology.

Introduction

Nowadays due to the advances of computer technologies there is a growing interest to empirical study of linguistic phenomena that were not previously available for the natural language analysis. To construct the really powerful tool for language analysis one need not only the developed software but also the highly elaborated linguistic theory. Therefore, the creation of linguistic typologies is one of the main tasks to carry out. There are some examples of such theories that are elaborating the most complex phenomena (Shutova 2011; Barrón-Cedeño 2013; Low 2010). One of the most complex subjects is a linguistic ambiguity, that, despite of a great number of references, is not well elaborated due to its complexity and diversity of its representations in the languages. There are only few typologies of ambiguity, which can be used for the automatic ambiguity recognition tasks, yet. Most of the works in this field are focusing on some varieties of ambiguity or some restricted aspects such as only lexical ambiguity, etc. (Franz 1996; Hirst 1992; Lexical ambiguity resolution 1988). This fact explains the timeliness of our investigation, which consist of an attempt of complex ambiguity description.

It should be marked in reference to every type of classification that natural language as every live organism is not discreet (Dawkins

* Corresponding author E-mail address: yuzhannikova@gmail.com

1993), especially at the high levels of its organizations, so it is impossible to divide all linguistic phenomena into some varieties exactly and inconsistently. However, within scientific description it is obligatory to seek to make the classification as precise as possible.

In the present work the ambiguity is defined as a presence of two or more senses in some phrase or its fragment, manifestating simultaneously or sequentially, which are determined by the combination of linguistic factors and appropriate context. On the pages below, it will be argued that such linguistic factors can be countered and described within the proposed framework.

Theoretical framework

Already some attempts to describe the ambiguity were made, few of them are really well-done (Wasow, Perfors & Beaver 2003; Lapteva 2007; Moskvin 2002; Tuggy 2006; Zaliznyak 2003), but so far, as we know, there is only one typology of ambiguity that could be used practically yet (Thomas Martin 2007). One of the most in-depth studies in that sense is the monograph of O. Lapteva "Textual homonymy (tekstovaya omonimia)", where the author describes the types of ambiguity basing on the thousands of examples. Although this work presents profoundly the variety of ambiguity examples, in the total, it is more the list of types than complete typology. Other works cannot offer such number of facts, what probably provokes deficiency of the generalization. And so the most of the ambiguity investigations focus at some limited aspect of ambiguity or they have it as a minor question.

Linguistic classifications of ambiguity

Nowadays one of the most significant discussions on ambiguity classification arises in the sphere of the language levels. Up to now there has been little agreement on what major classes ambiguity could be divided first of all. The overwhelming majority of researchers agree on setting the lexical type (ambiguity based on the change of meaning of a single word) and structural or syntactic one (where the ambiguity is provoked by some alternative syntactic analysis) (Alcaraz Varo, Martinez Linares 1997: 42; The Oxford Handbook of Cognitive Linguistics 2007: 160), but even in this case there are some discussions about its content. For example, lexical ambiguity is sometimes confused with the polysemy (word property to have several meanings). But polysemy is a phenomenon of language as a system, based on paradigmatic, while the ambiguity is a syntagmatic, discursive fact (or fact of "langage" according to the Saussure scheme) (L'ambigüité et la paraphrase 1987: 51; Bunt 2003; Moskvin 2002).

Concerning syntactic class of ambiguity, some scientists believe, in accord to the generativism, that only superficial structures as opposed to underlying structures can be ambiguous (Slobin 1971; Danesh 1964; Alcaraz Varo, Martinez Linares 1997: 42). However, others divide cases of syntactic ambiguity into superficial and underlying (L'ambigüité et la paraphrase 1987), understanding these terms out of generativism theory. That corresponds to the O. Lapteva division of ambiguity into spatial (positional) and inherent (essential) (Lapteva 2009: 227).

In other classes of ambiguity, corresponding to the main language levels, the situation of agreement is even worse. Another independent class which is frequently revealed is a semantic one, but since it is not clear where is the difference between lexical and semantic classes, some investigators avoid this class (Stanford Encyclopedia of Philosophy 2011; Zaliznyak 2003). For instance, A. Zaliznyak divides all the ambiguity cases only into two classes: lexical and syntactic. The lexical one she delimits

according to the sense distance between the word meanings to three types: coexistence of two independent meaning; creating the new one on the base of two initial meanings; and shifting between meanings (Zaliznyak 2003). However, other investigators, on the contrary, mark out the semantic type instead of syntactic: "At a low (e.g. speech recognition) level, a signal can be ambiguous between various utterances; at a higher (semantic) level, a fully recognized utterance can be used to express various different propositions; and at an even higher (pragmatic) level, a proposition may be used for various purposes" (Van Deemter 1998). In addition, this cite is illustrating the vagueness of class definitions, which is the typical problem for most of classifications.

In addition, exists the alternative classification tree of ambiguity where the lexical and syntactic types are parts of general semantic class. But here syntactic class has a non-common sense, it is comprehends as the part of speech ambiguity opposite to structural ambiguity, which is associated with the different syntactic trees of one sentence. In regard to semantic lexical class it is separated into polysemic and homonymic (Lexical ambiguity resolution 1988: 5). The authors of "Ambiguity in psycholinguistics" also put the ambiguity of parts of speech into the lexical class (Ambiguity in Psycholinguistics 1981: 101).

Besides, exists the point of view that integrates the phonetic, lexical, semantic and syntactic types of ambiguity into one general class of linguistic ambiguity as opposed to the pragmatic class (Gutiérrez-Ordóñez 1995: 29), which has even more discrepancy in the scientific literature than other ones. For instance, some authors name the anaphora based ambiguity as pragmatic (or discursive) (Thiollent 1987: 214; Thomas Martin 2007: 60), while others consider it as a type of the structural ambiguity (www.wikilengua.org/index.php/Ambigüedad; Gutiérrez-Ordóñez 1995).

The most exhaustive and well-structured classifications of ambiguity in terms of its linguistic levels (phonetic, lexical, semantic, syntactic and pragmatic or discursive) are (Fuchs 2009; Wasow, Perfors & Beaver 2003; Stanford Encyclopedia of Philosophy 2011; Thomas Martin 2007; Bunt 2007), that we were generally used during developing the linguistic ambiguity typology.

Pragmatic (extralinguistic) classifications of ambiguity

There are few classifications besides the linguistic that can help to analyze the ambiguity more completely and are called pragmatic. So far there has been little discussion about pragmatic classification due to lack of works devoted to the question and the variety of noncontradictory grounds of such classification presented in them.

First of all, ambiguity is divided accordingly to its intention characteristics into intentional and unintentional (Larrauri, Monteagudo; Benitez Soto; Kim 2006; Poesio 1996). Besides, L. Kim describes two grounds of ambiguity: the first represents ambiguity actualized (explicit) and non-actualized (implicit); and the second – realized and unrealized. She names the ambiguity realized when it was not resolute by the recipient (for several reasons: did not noticed, did not knew what meaning to prefer), the unrealized type, on the contrary, is usually simply resolved by the recipient in accordance with the intention of communicator.

In connection with previous classification the functions of intentional ambiguity can be mentioned. A. Zaliznyak suggests the two types of ambiguity according to its function: the calambour kind (mainly the entertaining examples) and non-calambour one (examples from the poetic texts) (Zaliznyak 2003). There

are some more elaborated classifications, but this question stay to be indefinite. For instance, G.V. Bobrovskaya sorts out the list of ambiguity functions: formative, sense-generative, persuasive, informative, which is divided into some subclasses. V. Moskvin marks in general the same functions, though with other names. The outward non-similarity of functions lists depends frequently on the difference of their names, while their internal content could be similar.

We suppose that ambiguity is represented in the language communication through the system of stylistic devices of ambiguity. Our classification of stylistic devices of ambiguity functions is based both on the proposals stated above and on the functions of universal system of stylistic devices (Kopnina 2009) with some concretizations.

Next sort of classification is related to the differences of the meaning occurrence. O. Lapteva sorts out three types of additional meaning perception: 1) such meaning become apparent entirely; 2) it become apparent partially; 3) stay conceal. We think that in the last case she means that ambiguity stays unresolved, because it does not exist if it was not noticed. Exist more common standpoint about two general sorts of ambiguity (Kiklevich 2007; Lapteva 2007; Pertsova 1988): the first sort is resolved in the narrow or broad context (we call it the partial ambiguity) and the second one do not resolved, at least, by the language means (complete ambiguity).

The model of connector-shifter ubication is another ground of classification. A connector is a keaword that makes it possible to change the sense of sentence. A shifter is an element switching the meanings of the connector. There are some different names for the same terms: connector (S. Attardo), functor (N.L. Myshkina), embrayeur (P. Charaudeau), connecteur (Greimas); and for shifter – disjuncteur (A.J. Greimas), script-switch trigger (J. Raskin), désembrayer (P. Charaudeau). However, the types of disposition of these elements are standard for the absolute majority of the authors.

The last classification ground that needs to be mentioned is the ambiguity intensity scale. In some sources it is possible to find out the scale that points out the movement from the homonymy to polysemy and the metaphorical meaning (Lexical ambiguity resolution 1988: 7), or from the insignificant perturbation of predictability to the complete ambiguity (Kess, Hope 1981).

Typology proposal

A new typology was elaborated as a result of some critical discussing of the special literature and preliminary analysis of language facts, extracted from 2000 text fragments in Russian language collected manually. For the analysis facility the typology was implemented to the language facts using a table (see the figure 1) with the possibility of multiple selection in each column (notice, please, that for the publishing reasons the table columns are presented as a lines). There are 21 columns in the table, developed in the base of data in Microsoft Access 2007, one part of them is extralinguistic (or pragmatic) characteristics, and another one is, in fact, the linguistic typology of ambiguity. The consolidation of these two parts permits to analyze the ambiguity facts completely and to retrieve more data for the phenomenon description not only from the point of view of its structure or inherent properties, but also from its use.

The table starts from the extralinguistic part, which describes the functional aspect of ambiguity. It can be used for the analysis of distribution of ambiguity in Russian language according to the sphere of its use, text type or its position in the text structure and so on.

The first column is the numerical one, which enumerates the ambiguity examples. The second contains directly the analyzed context with the ambiguity's example. The third and the fourth columns – source data and the date, respectively. The next section (fifth) characterizes the text's sphere of use. There are 11 items: Fiction, Folklore, Scientific literature, Mass media newspaper, Mass media journal, Mass media web page, Mass media radio, Mass media TV, Retranslated spoken language, Spoken language and other. Nearly all description points have the position "other" as the last variant for the possible fact unknown yet, that could have the characteristics different to the proposed.

In the sixth column it is necessary to specify the genre of text, from which the context was retrieved. The last two sections are used to establish some (at least, general) correlations between the ambiguity example and the sphere and genre of text in which it occurs.

The next, seventh column is devoted to the description of position of the extracted context in the source text, that permits us to make the conclusions about ambiguity's frequency in correlation with their place in the text.

The eighth – is a free-choice column, which can indicate the theme of the source text. This point is not obligatory and is filled up in the case if a text belong to the some specific type of discourse as political, for instance, where the theme itself can influence on the peculiarities of the used stylistic devices.

The ninth column is marked up if a context contains some ambiguity or not (with the three labels: ambiguity, non-ambiguity and controversial). From the one side, it seems logically to have the entire examples ambiguous if we are analyzing ambiguity. But, from the other side, there are many cases of use of the same linguistic mechanisms (as polysemy, homonymy, irony, ellipsis and so on) which are not ambiguous, so the demonstrating of these cases permits us to illustrate the difference between the investigated phenomenon and its possible structural foundations. A line "controversial" used for mark up the facts, which we are not able to put definitely into one or another category.

The tenth section is devoted to the ambiguity resolution type, where the choice is carry out from the six possibilities: Complete; Partial context explicit; Partial context implicit; Partial author; Partial interlocutor; Garden path.

The next one, eleventh, is used to label the type of connector-shifter configuration (this classification is valid, in general, only for the semantic class of ambiguity).

The second part of the sections is devoted to the linguistic characteristics of analyzed fragments, which we will describe more detailed.

One of the most important characteristics of every ambiguity example (number twelfth) is the class of ambiguity. It is the basic part of our structural (or substantial) typology, which consists of 6 items (not include the "Other" line): two principal types for each general class (lexical, semantic and pragmatic level). In our opinion there is no reason to sort out the morphological and discursive ones, because they can go into the lexical and pragmatic classes respectively.

In the next, thirteenth, column the class of ambiguity is concretized with the choice of one of its formative mechanisms. It permits to make the description of each example the most precise to avoid the unification of the heterogeneous elements, but at the same time not to lose the necessary level of generalization and not to come down into the individual descriptions.

The gradation of the ambiguity structure (fourteenth) is refer to the formal characteristics and is not obligatory.

The fifteenth column provides the selection between the ambiguity devices' models. The list of stylistic devices based (completely or partly) on ambiguity was formed on the basis of numerous lists of rhetorical (stylistic) devices from antiquity up to date. There is a choice between amphiboly, antiphrasis, antanaclasis (ploca), antanaclasis (diafora), literalization of metaphor, dilogy or zeugma. It is another essential part of the table, because the results of classification according to columns 12, 13 and 15 permit us to perform one of the principal tasks of our investigation – to retrace the correlation between the linguistic (formative) grounds of ambiguity and their possible use as a stylistic devices.

In the sixteenth column the choice of formal markers is presented, which are not so common for the ambiguity in general, but very frequent in the case of some particular devices (as antiphrasis, for instance).

Other essential characteristic is situated in the seventeenth column, that we call the "status" column. There is a choice between two items: if the analyzed example is a device or an error. The third item "controversial" permits to put into it the complex examples. It is important to note that even erroneous example, when it reused with some intention (often for entertain), is converted into the stylistic device (so the functionality is relevant in that case). This section entails the next one, with the choice of the function for each ambiguity context (in case when it is not an error). The functions usually are correlated closely with the sphere and genre of the source text and depend on their author's intentions, but in general their identification stay mostly intuitive.

The following characteristics are specialized and are non-common for all the instances and, therefore, are facultative at some point, though can help in the some particular descriptions of ambiguity.

It is, in the first place, the degree of meaning change (column number 19), which is per se rather arbitrary division of the meaning degree into the three grades: strong (when one of the meanings is predominating), medium (when the meanings are nearly equal but rather independent) and week (when the meanings or, to be more exact, the meaning nuances, are very close). This section is based on the hypothesis of the graduality of ambiguity manifestation.

The penultimate column characterizes the direction of meaning change (literal-figurative) or the order of appearance (initial-derivative) and applies only to the semantic class of ambiguity.

The last one is the free-choice section assigned for the some description, for instance, to make it possible to describe some special features of any example or to explain its labeling one or another category that can be sometimes rather unobvious.

Conclusion

The presented scheme characterizes the ambiguity from different points of view and allows providing more detailed qualitative and quantitative analysis not only of linguistic grounds of its origin, but also its pragmatic attributes. Besides, the elaborated typology could be applied to the annotation of both a special corpus, such as corpora of colloquial speech or mass media texts, as well as the National Corpus of Russian Language. Hypothetically, this typology can be implemented for the annotation of the ambiguous contexts for other natural languages and present article in this case could serve for the development of annotators' guide.

References

1. Alcaráz Varo E., Martínez Linares M.A. (1997). Diccionario de lingüística moderna (The dictionary of modern linguistics). Editorial Ariel, Barcelona, 1997. 643 p.

2. Barrón-Cedeño, A., Vila, M., Martí, M.A. & Rosso, P. (2013). To appear. Plagiarism meets paraphrasing: Insights for the next generation in automatic plagiarism detection. *Computational Linguistics*. DOI: 10.1162/COLI_a_00153.

3. Benitez Soto, M.V. Explicación y aplicación del concepto de "transmisión pseudo-ostensiva de información encubierta" a través de un caso de ambigüedad deliberada (Understanding and applying the concept of "pseudo-ostensive transmission of hidden information" through a case of deliberate ambiguity). *Interlingüística*, 14, 2003, pp. 141-150.

4. Bobrovskaya, G.V. (2011). *Kognitivno-elocutivnyj potentsial gazetnogo diskursa* (The cognitive-elocutive potential of newspaper discourse). Volgograd, 2011. 319 p.

5. Bunt, H. (2007), "Semantic Underspecification: Which Technique For What Purpose?", *Computing Meaning*. Vol. 83, pp. 55-85.

6. Danes, F. (1964). A three level approach to sintax. *Travaux Linguistiques de Prague*. 1, pp. 225-240.

7. Dawkins, R. (1993). Meet my couzin, the chimpanzee! *New Scientist*, 5 June, 138 (1876), pp. 36-38.

8. Franz, A. (1996). *Automatic Ambiguity Resolution in Natural Language Processing: An Empirical Approach*. Springer. 155 p.

9. Fuchs, C. L`ambigüité: du fait de langue aux stratégies interlocutives (The ambiguity: from the language to the interlocative strategies) *Travaux neuchatelois de linguistique*. 2009, 50, 3-16. Available at: www. archive-ouverts.fr

10. Gutiérrez-Ordóñez, S. Ambiguedades pragmaticas (Pragmatic ambiguities). *Ambiguïtés / Ambivalences. Actes du colloque de Rouen*, 13-15 mai 1994. Texts réunis par Anne-Marie Vanderlynden. Les cahiers du CRIAR: Centre de recherches d'etudes ibérique et ibero-americanes. №14. Publications de l'université de Rouen №203. Université de Rouen, 1995. Pp. 27-35.

11. Hirst, G. (1992). *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge University Press. 267 p.

12. Kess, J.K., Hoppe R.A. (1981). *Ambiguity in Psycholinguistics*. AMSTERDAM/ JOHN BENJAMINS B.V., Wilrijk, Belgium, 1981.

13. Kiklevich A. (2008). Parasemy (Parasemia). *Russkaya rech'* (Russian Speech), 1. Pp. 43-45.

14. Kim, L.G. Tipologia amfibolicheskich vyskazyivanij v svete kommunicativno-deyatel'nostnoj-kontseptsii yazyka. (A typology of amphibolic utterances considering a communicative-active conception of language). *Antropotekst-2*. Barnaul, 2006. pp. 135-143.

15. Kopnina G.A. (2009). *Ritoricheskie priemy sovremennogo russkogo literaturnogo iazyka: opyt sistemnogo opisaniia* (The rhetorical devices of modern Russian literary language: proposal of systematic description). M., 569 p.

16. *L'ambigüité et la paraphrase* (The ambiguity and the paraphrase). Opéracions linguistiques, processus cognitifs, traitements automatisés. Actes du colloque de Caen, 9-11 avril 1987, publiés sous la direction de Catherine Fuchs avec le concours du Centre National de la Recherche Scientifique. Université de Caen, CNRS UA 1234, 1987. 330 p.

17. Larrauri, L., Monteagudo, M.A. *La ambiguedad en el lenguaje jurídico: ¿amplitud o distorción semántica?* Available at: http://www.realiter.net/spip.php?article1651 (accessed 17.11.2011).

18. Lapteva, O. (2007). *Rechevie vozmozhnosti tekstovoj omonimii* (Speech possibilities of text homonymy). M., URSS, 413 p.

19. *Lexical ambiguity resolution*: perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence. Ed. S.L. Small, G.W. Cottrell, M.K. Tanenhaus. Morgan Kaufmann Publishers, Inc. San Mateo, California, 1988.

20. Low, G., Todd, Z., Deignan, A., & Lynne Cameron (eds.) (2010). *Researching and Applying Metaphor in the Real World*. xii, 385 p.

21. Moskvin, V.P. (2002). Figury dvusmyslennoj rechi (The figures of ambiguous discource). *Russkij yazyk v schkole* (Russian Language in the school). 2. Pp. 86-90.

22. Pertsova, N.N. (1988). *Formalizatsia tolkovania slova* (The formalization of the word definition). M., 83 p.

23. Poesio M. (1996) Semantic ambiguity and Perceived Ambiguity. *Semantic Ambiguity and Underspecification*; ed. K. van Deemter and S. Peters. Ch. 8, p. 159-201. CSLI, Stanford, CA. Available at: cswww.essex.ac.uk/poesio/publications/vandeemter_book.pdf

24. Shutova, E. (2011). *Computational approaches to figurative language*. PhD thesis, Computer Laboratory, University of Cambridge, UK. 217 p.

25. Slobin, Dan I., *Psycholinguistics*, Glenview: Scott Foresman, 1971. 351 p.

26. *Stanford Encyclopedia of Phylosophy*. Principal Editor: Edward N. Zalta. Available at: http://plato.stanford.edu/

27. *The Oxford Handbook of Cognitive Linguistics* (2007). Edited by Dirk Geeraerts and Hubert Cuycens. Oxford University Press, 2007. 1334 p.

28. Thiollent, M. (1987). The pragmatic of the ambiguous representation (Pragmatique de la représentation ambigüe). *L'ambigüité et la paraphrase.* Opéracions linguistiques, processus cognitifs, traitements automatisés. Actes du colloque de Caen, 9-11 avril 1987, publiés sous la direction de Catherine Fuchs avec le concours du Centre National de la Recherche Scientifique. Université de Caen, CNRS UA 1234, 1987. Pp. 211-215.

29. Thomas Martin, B. (2007). PhD thesis. *A Typology of Ambiguity as it Relates to Natural Language Processing.* College of Engineering and Information Technology. University of South Carolina. 162 p.

30. Tuggy, D. Ambiguity, polysemy and vagueness. *Cognitive linguistics: basic readings* / ed. by Dirk Geeraerts (Cognitive linguistic research). 2006. Pp. 167–185.

31. Van Deemter, K. (1998) Ambiguity and Idiosyncratic Interpretation. *Journal of Semantics* 15 (1), pp. 5-36.

32. Wasow, T., Perfors, A. & Beaver, D. (2003). *The Puzzle of Ambiguity*. Stanford University, 2003. Available at: www.stanford.edu/~wasow/Lapointe.pdf

33. Zaliznyak, A.A. (2003). Neodnoznachnost', kalambur i nekalamburnoe sovmeschenie znachenij: k probleme predstavleniya mnogoznachnosti. (Underspecificity, calambour and non-calambour combination of meanings: towards the problem of representation of ambiguity). Available at: www.dialog-21.ru/Archive/2003/Zalizniak.htm

Типология языковой двусмысленности

М.А. Южанникова
Сибирский федеральный университет
Россия, 660041, Красноярск, пр. Свободный, 79

Несмотря на то что двусмысленность является неотъемлемой частью любого естественного языка, до сих пор ее исследование было сосредоточено в основном на проблеме ее разрешения. Сейчас же в связи с развитием компьютерных технологий автоматического анализа текста возрос интерес к построению лингвистических типологий. Данная статья посвящена разработке типологии речевой двусмысленности. В первой части работы рассматриваются теоретические проблемы классифицирования двусмысленности, во второй – предложена типология двусмысленности на основе русского языка, представленная в виде таблицы. Классификационная таблица состоит из восьми колонок, описывающих пример экстралингвистически, и тринадцати – представляющих собой лингвистическое описание. Данная типология позволяет описать любой пример двусмысленного высказывания вне зависимости от его коммуникативного статуса.

Ключевые слова: двусмысленность, классификация, типология, неопределенность, полисемия, омонимия.

Статья написана в рамках обучения в аспирантуре Университета Барселоны, Испания (проект ERANET MUNDUS).

Научная специальность: 10.00.00 – филологические науки.