

**MSC 2020****International Scientific and Practical Conference «MAN. SOCIETY.  
COMMUNICATION»****PRAGMATIC MARKERS OF RUSSIAN ORAL SPEECH:  
STRUCTURAL AND FUNCTIONAL ASPECT**

Natalia Bogdanova-Beglarian (a)\*

\*Corresponding author

(a) Saint Petersburg State University, Saint Petersburg, Russian Federation  
N.bogdanova@spbu.ru**Abstract**

The main object of this article is an integral element of oral discourse – pragmatic markers (PM). Such functional units are practically devoid of their lexical, and often grammatical meaning, however, they perform well-defined functions in speech: hesitative (*kak jego (jejo, ikh), eto, eto samoe*), metacommunicative (*znaesh', slushaj, (ja) ne znaju*), demarcative (*znachit, nu vot*), deictic (*vot tak vot, vot takoj vot*), rhythm-forming markers (*vot, tam, koroche*), approximators (*kak by, tipa*), xenomarker (*tipa (togo chto), takoj*), substitute markers (*blah-blah-blah, i vse dela, i vs'o takoe, tuda-s'uda*), reflexives (*skazhem (tak), ili kak ego? ili chto?*) and markers of self-correction (*eto, eto samoe*). The study is based on two oral corpora created at St. Petersburg State University: the corpus of everyday Russian speech “One Day of Speech” (ORD) (dialogs / polylogues) and “Balanced Annotated Text Library” (SAT) (monologues of various types). The article provides a description of the different structural types of basic PM: single-word – non-single-word; the original immutable lemma (*vot, tam*) – the word form (*predstav', koroche*); variable form – partially variable form – constant form; etc. This data can make the description of the PM of Russian spoken language more complex and suitable in different applied purposes, such natural language processing, machine translation or teaching Russian as a foreign language.

2357-1330 © 2021 Published by European Publisher.

*Keywords:* Functional aspect, oral speech, pragmatic marker, speech corpus, structural aspect

## 1. Introduction

Essential elements of any oral discourse, no matter what language is used, are such specific functional units, which have no direct connection with the content of speech, but necessary for the speaker during the speech under a time crunch, when speaker have to think and speak simultaneously. These units help speaker to build / streamline the text and overcome difficulties related to the spontaneous speech generation (hitches, slips of the tongue, etc.). Using them speaker establish and then maintain contact with the interlocutor and respond to their speech activity: reflect, comment, evaluate, cf.: Rus. *eto samoe, kak jego, znaesh'te*; Engl. *y'know, I mean, .well, look*; Germ. *wie war das doch, weißt du, siehst du*; Pol. *wie pan, widzisz*; Span. *digamos, esto, y tal, y todo*. Such units are appropriate to be called pragmatic markers (PMs) (Bogdanova-Beglarian et al., 2019) and must be distinguished from discourse markers (DMs) that are well described in linguistics (Baranov et al., 1993; Kiseleva & Pajar, 1998; Kiseleva & Pajar, 2003; Lenk, 1998; Shiffrin, 1996; Shourup, 1999).

DMs are ordinary linguistic units that potentially passed the grammaticalization process, as a result of which, for example, verb forms are separated from the general paradigm and begin to be used as introductory words and collocations (*skazhem, mozhet byt', kazhets'a, kak govorits'a*). PMs also created from the usual significant units of the language (words / word forms, collocations or even sentences), but passed in the oral discourse through the process of pragmaticalization (sometimes it is also preceded by grammaticalization – these two processes are closely related in the language) and, as a result, have lost their lexical meaning. In such cases, researchers speak about “washing out” the semantics of a unit, or about word’s “bleaching” (Bybee, 2003), or about its “semantic emasculation” (Shmelev, 2004). In this case, the grammatical meaning of a unit also weakens and sometimes remains only at the level of “atavisms”: preserved, for example, variability by gender, number and case (*eto samoe*), by gender and number (*kak jego (jejo, ikh), takoj*), gender and case (*p'atoe-des'atoe*), only by number (contact verbs) or only by case (*vse dela*). In addition, the PMs, unlike the DMs, is generated unconsciously by the speaker, they used exclusively in spoken language or its stylizations in the written text and usually outside the lexicographic fixation and, accordingly, beyond the scope of linguodidactics and other applied aspects of linguistics. The latter circumstance makes the study and description of PMs especially relevant.

## 2. Problem Statement

Traditionally, for such units, practically devoid of both lexical and grammatical meanings, in linguistics, terms with a very negative connotation are used: “parasite words” (Daragan, 2000; Razlogova, 2003; Shmelev, 2004), “superfluous words” (Sirotnina, 1974), “empty lexemes / particles” (Rožanova, 1983), etc., however, “in all such cases, the semantic emptiness of the linguistic expression is imaginary in some sense, because it is filled with rich pragmatic content” (Shmelev, 2018); meaning is replaced by a function; as a result, such words become an integral part of oral speech, ensuring the success of communication. Studies have shown that without these units the listener perceives the text as artificial (Riekhakainen, 2016), and their usage suggests that they are a kind of regulators of communication, in one way or another influencing the perception of the statement by the interlocutor.

### 3. Research Questions

The importance of pragmatic markers for both the speaker and the success of the process of oral communication makes their description relevant and necessary for a variety of purposes. Including their description from the structural side, to which this article is devoted.

### 4. Purpose of the Study

The purpose of this study is to describe the different structural types of basic PM, taking into account the functions that they perform in oral discourse. The presence of such data can make the description of the pragmatic markers of Russian spoken language more complete and suitable for use for various applied purposes, such as automatic processing of sounding speech, the practice of translating or teaching Russian as a foreign language.

### 5. Research Methods

The main source of the material for the study was two oral corpora created at St. Petersburg State University: the corpus of everyday Russian speech “One Day of Speech” (ORD), compiled with the method of long-lasting oral speech monitoring (dialogs / polylogues, more than 1250 hours of sound, 130 informants and more than 1000 of their communicants, 1 million word forms in transcripts) (see recent works on it: (Bogdanova-Beglarian, 2016; Bogdanova-Beglarian et al., 2016) and “Balanced Annotated Text Library” (SAT) (800 monologues of various types, about 50 hours of sound) (Bogdanova-Beglarian, 2013).

The speech corpus, by definition, should contain not only a certain array of texts, but also their annotation (Gries, & Berez, 2017; Plungjan, 2008; Zakharov, 2005). Both corpora used in the work are partially annotated. The volume of annotated sub-corpora is 21 504 tokens for the ORD corpus and 50 128 tokens for SAT. In particular, pragmatic markers were annotated in both cases, making up a significant proportion of their structural elements: 2.77 % in the material as a whole, 2.83 and 2.57 % separately in dialogue (ORD) and monologue (SAT) (Bogdanova-Beglarian et al., 2019).

### 6. Findings

An annotation of pragmatic markers in the corpus material allowed us to create the typology of PMs models (their basic options, which are opposed by contextual variants with various extensions: *skazhem – tak skazhem, skazhem tam, vot skazhem tak; tipa – tipa togo – tipa togo chto; znaesh’ – ty znaesh’*; etc.) and get some quantitative data separately for dialogue and monologue.

In total, 59 basic PMs were revealed in the annotated sub-corpora based on the ORD, and 28 PMs in the SAT sub-corpora. Markers in both cases are repeated (28 units from SAT are completely included in the PMs vocabulary from the ORD-corpus), and the “top” of the frequency dictionaries of the implemented PMs variants in both cases looks similar: *vot, tam, da, tak, kak by, govorit (grit), znaesh’, slushaj, znacjit, eto* – in ORD; *vot, znacjit, tam, nu vot, kak by, tak, da, vs’o, skazhem tak, v obshchem-to* – in SAT.

First of all, PMs can have single-word form and multi-word form. In the ORD-corpus they are distributed approximately equally (50.9 and 49.2 %, respectively); in the SAT-corpus, there were slightly more single-word ones (57.1 and 42.9 %).

Single-word PMs were divided into three types:

1) initial unchanging forms (which came to the PMs from adverbs or particles, the result of pragmaticalization only): *von, vot, tam, tak, da, voobshche, shchas* (namely in a reduced form), *vrode*;

2) frozen and also unchanging forms (former names and verbs, the result of grammaticalization first, and then pragmaticalization processes);

3) partially mutable forms (also former names and verbs that have retained the “truncated” grammatical paradigm, the result of pragmaticalization only): *znaesh’/te, eto/eta/etim...*, *govor’u/govorit...* (usually in a reduced form – *gr’u / grit...*), *takoj/takaja/takie, predstav’/te* and so on.

In the monologue, the single-word PMs of the second type clearly prevails (50.0 % of the total number), in the dialogue (although not so explicit) the third-type PMs (43.3 vs. 26.7 and 30.0 %), mainly metacommunicative (*znaesh’/te, predstav’/te, smotri/te*) and xenomarker (*grit, takoj*).

Multi-word PMs are divided into groups differently than single-word PMs:

1) the original multi-word units, both mutable (*to-s’o, p’atoe-des’atoe*) and unchanging (*kak by, tuda-s’uda*);

2) unchanging combinations-contaminations, often reduplicated: *vot (...) vot, tak i tak, to-to..., te-te..., na-na..., blah-blah...*;

3) the original immutable collocation: *koroche govor’a, sobstvenno govor’a, tak dalee, kak srazat’, tak skazat’*;

4) frozen and already unchanging collocation or elliptic predicative units: *na samom dele, kak eto, kak govorits’a, kak nazyyaets’a, ili tam*;

5) partially mutable collocation or elliptic predicative units: *eto samoe, kak jego (jejo, ikh), ili kak jego (jejo, ikh), vs’o takoe, vse dela, vs’akoe takoe*.

Most often, the last two types (4-5) (24.1 %) met in the ORD-corpus; the last type (5) (33.3 %) in the SAT-corpus.

The difference between the dialogue and the monologue is associated more with the functional rather than the structural characteristics of the PMs: there is a predominance of metacommunicatives and xenomarkers in the dialogue, where the interaction of communicants plays a large role, and differentiating (starting, navigation and final) PMs in the monologue, where it is important for the speaker somehow structure the text, mark its borders.

Particularly noteworthy are also ways to expand the basic options for PMs: cf.: VOT – *i vot, da vot*; ZNAESH’ – *ty znaesh’, vot znaesh’, nu znajete*, etc.; GOVORIT – *govor’u, govorish’, govorim*, etc.; VRODE – *nu vrode, vrode kak, vrode by*. For 59 basic PMs in the ORD, 315 of their specific realizations were identified, for 28 basic PM in the SAT – 133, and there are statistics of these realizations in both types of speech. This, however, is beyond the scope of this article.

## 7. Conclusion

The article presents the results of a structural-functional description of pragmatic markers obtained on corpus material, both dialogical (ORD corpus) and monological (SAT corpus). This data can make the description of the PM of Russian spoken language more complex and suitable in different applied purposes, such natural language processing, machine translation or teaching Russian as a foreign language.

## Acknowledgments

The presented research was supported by the Russian Science Foundation, project #18-18-00242 “Pragmatic Markers in Russian Everyday Speech”.

## References

- Baranov, A. N., Plungyan, V. A., & Rakhilina, E. V. (1993). *Putevoditel' po diskursivnym slovam russkogo jazyka* [Russian Discourse Words Guide]. Pomovsky i partner.
- Bogdanova-Beglarian, N. V. (Ed.). (2013). *Zvukovoj korpus kak material dl'a analiza russkoj rechi: kollektivnaja monografija. Chast' 1. Chtenie. Pereskaz. Opisanie* [Speech Corpus as a Base for Analysis of Russian Speech. Collective Monograph. Part 1. Reading. Retelling. Description]. Saint Petersburg State University Publishing House.
- Bogdanova-Beglarian, N. V. (Ed.). (2016). *Everyday Russian Language: Functioning Features in Different Social Groups. Collective Monograph* [Russkij jazyk povsednevnogo obshchenia: osobennosti funkcionirovania v raznykh social'nykh gruppakh. Kollektivnaja monografija]. LAYKA.
- Bogdanova-Beglarian, N., Blinova, O., Martynenko, G., Sherstinova, T., Zaides, K., & Popova, T. (2019). Annotirovanie pragmaticeskikh markerov v russkom rechevom korpuse: problemy, poiski, resheniya, rezul'taty [Pragmatic markers annotation in Russian speech corpus: research problem, approaches and results]. In V. P. Selegey (Ed.), *Computational Linguistics and Intellectual Technologies* [Kompjuternaja lingvistika i intellektual'nye tekhnologii] (pp. 72-85). RSUH Publ.
- Bogdanova-Beglarian, N., Sherstinova, T., Blinova, O., Baeva, E., Martynenko, G., & Ryko, A. (2016). Sociolinguistic Extension of the ORD Corpus of Russian Everyday Speech. *Lecture Notes in Artificial Intelligence, 9811 LNCS*, 659-666. [https://doi.org/10.1007/978-3-319-43958-7\\_80](https://doi.org/10.1007/978-3-319-43958-7_80)
- Bybee, J. (2003). Mechanisms of Change in Grammaticization: The Role of Frequency. In B. D. Joseph, & J. Janda (Eds.) *The Handbook of Historical Linguistics* (pp. 602-623). Oxford: Blackwell.
- Daragan, Yu. V. (2000). Funkcii slov-«parazitov» v russkoj spontannoj rechi [Functions of Parasite Words in Russian Spontaneous Speech]. In A. S. Nariniani (Ed.), *Proceedings of the International Seminar “Dialogue-2000” on Computational Linguistics and Their Applications. Theoretical Problems* (pp. 67-73). Protvino, RF: RGGU.
- Gries, S. T., & Berez, A. L. (2017). Linguistic Annotation in/for Corpus Linguistics In N. Ide, & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 379-409). Springer.
- Kiseleva, K. L., & Pajar, D. (Eds.). (1998). *Diskursivnye slova russkogo jazyka: Opyt kontekstno-semanticheskogo opisania* [Russian Discourse Words: An Experience of Contextual-Semantic Description]. Metatext.
- Kiseleva, K. L., & Pajar, D. (Eds.). (2003). *Russian Discourse Words: Context Variation and Semantic Unity* [Diskursivnye slova russkogo jazyka: Kontekstnoe varjirovanie i semanticheskoe edinstvo]. Azbukovnik.
- Lenk, U. (1998). *Marking Discourse Coherence: Functions of Discourse Markers in Spoken English*. Narr.
- Plungjan, V. A. (2008). Korpus kak instrument i kak ideologia: o nekotorykh urokakh sovremennoj korpusnoj lingvistiki [Corpus as a Tool and as an Ideology: on Some Lessons of Modern Corpus Linguistics]. *Russian Language and Linguistic Theory*, 16(2), 7-20.

- Razlogova, Je. E. (2003). K voprosu o specificeskikh upotrebleniakh modal'nykh slov: slova-parazity v russkoj i francuzskoj rechi [On the Question of the Specific Uses of Modal Words: Words-Parasites of Russian and French Speech]. *Moscow State University Bulletin. Series 9. Philology*, 6, 152-169.
- Riekhakainen, E. I. (2016). *Perception of Russian Spoken Language: Context + Frequency* [Vosprijatie russkoj ustnoj rechi: kontekst + chastotnost']. Saint Petersburg State University Publishing House.
- Rozanova, N. N. (1983). Supersegmentnaja fonetika [Super-Segment Phonetics]. In E. A. Zemskaya (Ed.), *Russkaja razgovornaja rech': Fonetika. Morfologiya. Leksika. Zhest* [Russian Colloquial Speech: Phonetics. Morphology. Vocabulary. Gesture]. Nauka.
- Shiffrin, D. (1996). *Discourse markers*. Cambridge University Press.
- Shmelev, A. D. (2004). Zapolniteli pazv kak kommunikativnye markery [Placeholders as Communicative Markers]. In M. Leynonen (Ed.), *Interview Genre: Features of Russian Spoken Language in Finland and St. Petersburg* [Zhanr intervju: osobennosti russkoj ustnoj rechi v Finlyandii i Sankt-Peterburge] (pp. 205-222). Tampere univ. press.
- Shmelev, A. D. (2018). *Vekhi pragmatikalizacii* [Milestones of Pragmatization]. [https://www.slavistik.uni-muenchen.de/forschung/tagungen/tagungsarchiv/konferenz-pragmatik/abstract\\_aleksej-shmelev.pdf](https://www.slavistik.uni-muenchen.de/forschung/tagungen/tagungsarchiv/konferenz-pragmatik/abstract_aleksej-shmelev.pdf)
- Shourup, L. (1999). Discourse Markers. *Lingua*, 107, 227-265.
- Sirotnina, O. B. (1974). *Sovremennaja razgovornaja rech' i jejo osobennosti* [Modern Conversational Speech and Its Features]. Prosveshchenie.
- Zakharov, V. P. (2005). *Korpusnaja lingvistika: Uchebno-metodicheskoe posobie* [Corpus Linguistics: Tutorial]. Saint Petersburg State University Publishing House.