A Corpus-based Analysis of *vvodnye slova*: Pedagogical Applications

Joan F. Chevalier

*United States Naval Academy*, chevalie@usna.edu

Recommended Citation

DOI: [https://doi.org/10.70163/0036-0252.1332](https://doi.org/10.70163/0036-0252.1332)

Available at: [https://scholarsarchive.byu.edu/rlj/vol73/iss1/4](https://scholarsarchive.byu.edu/rlj/vol73/iss1/4)
A Corpus-based Analysis of \textit{vvodnye slova}: Pedagogical Applications


d\textsc{joan f. chevalier}

\textbf{1. Introduction}

\textit{Vvodnye slova}, “introductory words,” (henceforth VS) commonly known as parentheticals in English, are a linguistic category that has received little attention until recently. There are several reasons why this category has remained on the periphery. The generally accepted definition of VS is quite broad, resulting in a considerable lack of consensus about which words and expressions are VS. In fact, no two lists of VS, even from authoritative linguistic sources, are the same. Pedagogical materials produced for Russian language as a L2 typically lack any discussion of VS as a category. Textbooks that do include VS introduce them as connectors without explanation of their larger discourse functions. The central questions posed by this study is whether VS should be included in Russian language curricula, if so why, and which ones. Based on the evidence presented below, this study will argue that VS are important for text comprehension and for the development of written expression, especially for those seeking to reach higher levels of proficiency.

Although a number of labels have been used to describe words and phrases function as VS, including, discourse markers, pragmatic markers, and parentheticals, the Russian term \textit{vvodnye slova} has been retained throughout the paper. As will be discussed below, although some VS can function as discourse markers and, in some cases, as pragmatic markers, both of categories, have “fuzzy” often overlapping definitions that are broader than the established lists of \textit{vvodnye slova}. This paper focuses specifically on the category defined in the Russian linguistic tradition as VS in an effort to clarify the use and frequency of these expressions.

The study begins with a description of the linguistic characteristics of VS, summarizing recent efforts to characterize and classify them. The second half of the paper presents the results of a quantitative analysis of data from the Russian National Corpus, pertaining to the frequency
of *vvodnye slova* and specific VS types in discourse. The purpose of this
corpus-based inquiry is to establish which types of VS are used most
frequently in Russian discourse. Based on this data recommendations
are made about which VS should be included in L2 instruction and
when. The data about which forms are used more frequently can inform
curriculum designers about which VS forms would be most useful to
learners.

2. Parentheticals and Vvodnye slova: Definitions and Types
Parentheticals are a “motley crew” (Dehé & Kavalova, 2007, p. 1); a
heterogenous linguistic category that can take a variety of forms
including but not limited to: sentences, adverbs, sentential adverbs,
reporting verbs, comment clauses, nominal appositions, nonrestrictive
relative clauses, question tags, prepositional phrases, noun phrases,
interjections, adjectival and adverbial phrases, and clauses both with
and without connectors (Kaltenböck, 2007). They can occur in spoken
utterances and written text. As interest in this elusive category has grown in
recent years, so too have the array of monikers applied to them, including:
Extra Clausal Constituents (Dik, 1997), non-clausal units (Biber, 1999),
supplements (Huddleston & Pullum (2002), and theticals (Kaltenböck
et al., 2016; Heine et al., 2013), to name a few. All of these various
approaches agree on two central defining facts about parentheticals: 1)
that they are syntactically independent – they are syntactically “disjunct”
from the host clause (Kaltenböck et al., 2016, 4; Burton-Roberts et al. 2006;
Dehé & Kavalova 2007), but 2) their meaning is derived from the context
in which they appear (Kaltenböck et al. 2016, p. 10). As Sonnenhauser
(2020) observes, parentheticals are “linearly embedded in, or ‘hosted’
by, another syntactic sequence without being structurally anchored to
it.” The syntactic independence of parentheticals has been established
in generative literature by means of linguistic tests demonstrating that
parentheticals are not part of the constituent structure of their host.
The tests (sentences 1a-d below) show that parentheticals cannot be the
focus of a cleft sentence, that they can be omitted without affecting the
grammaticality of the sentences, and that they are syntactically outside of
the scope of negative operators in the host clause (Kaltenböck et al., 2016,
p. 4-5). These properties are demonstrated by use of the English adverb
frankly in the examples below (1a-d).
1) \textit{Frankly, those reasons are not good enough.}

a. \textit{*It is \textbf{frankly} that those reasons are not good enough.} \\
[no focus of \textit{it}-cleft]

b. How is that? \textit{*Frankly.} \\
[no questioning]

c. Those reasons are not good enough. \\
[omissibility]

d. \textit{Frankly, those reasons are good enough.} \\
[unaffected by negation of the host]

Examples taken from Kaltenböck et al., (2016, p. 4-5).

Studies of parentheticals in spoken discourse have pointed out that their syntactic independence is also signaled prosodically; they are often set apart by “comma” intonation – pauses – as well as shifts in pitch (Crystal, 1969; Dehé & Kavalova, 2007). Kaltenböck et al. (2016), however, show that parentheticals do not in fact display homogenous prosodic features. Likewise, Grenoble (2004, p. 1954) argues that a large number of parentheticals in written discourse are not separated by commas.

Semantically as well as syntactically, parentheticals are best described negatively – by what they are not (Kaltenböck, 2016; Kaltenböck, 2007). They are semantically unrestrictive (Huddleston & Pullum 2002; Kaltenböck et al., 2016). The meaning of parentheticals is “related to but not part of the main message” (Biber 1999). In example 1c above, we see that the parenthetical \textit{frankly} can be omitted “without affecting the rest of that structure or its meaning” (Biber, 1999, p. 1067). Although parentheticals are semantically disjunct from the main discourse, they make vital contributions to the semantics of the utterance. They serve a metacommunicative function, commenting about the text in particular ways (Grenoble, 2004). The speaker can use them to express their subjective evaluations of the context, conveying their personal feelings, opinions or attitude about the host text. They can also supply supplementary or background information about the context and refer to the source of the information (Grenoble, 2004). As Urmson explains,
parentheticals provide “stage-directions” for the interlocutor, to “prime the hearer to see the emotional significance, the logical relevance, and the reliability of the statement” (1952, p. 484).

Semantically, parentheticals can be divided into two basic types: those that have a “conceptual meaning” and those that have “procedural meaning” (Grenoble, 2004, p. 1955). Parentheticals with conceptual meaning affect the truth conditions of the utterance and are a component of its propositional content. Procedural parentheticals do not affect the truth conditions of an utterance; they “carry information about how conceptual meaning is to be processed” (Grenoble, p. 1955). The original distinction between procedural and conceptual meaning was developed by Grice (1989), who provided sentence 1 below as an example of procedural meaning. In example 1, the phrase “on the other hand” does not affect the truth content of the sentence and therefore does not have conceptual meaning. In sentence 2, the parenthetical phrase enclosed in parentheses adds essential information about the subject of the sentence. The omission of the parenthetical clause in sentence 2 changes the propositional content of the sentence. The omission of “on the other hand” in sentence 1, however, does not affect propositional content.

1) My brother-in-law lives on a peak in Darien; his great aunt, on the other hand, was a nurse in World War I. (Grice 1989, p. 362; qtd in Grenoble, 2004, p. 1955)

2) The library (the six-storied building on Main Street) is open twenty-four hours a day.

This study focuses exclusively on procedural parentheticals in Russian, referred to in the Russian grammatical tradition as vvodnye slova, “introductory phrases.” The commonly accepted English translation for the term vvodnye slova is “parentheticals.” Although this translation is not incorrect, it can be misleading since in English the term “parenthetical” refers both to procedural and conceptual parentheticals. This paper focus specifically on VS, that is procedural parentheticals that are identified in the Russian grammatical tradition as vvodnye slova.

The Russian examples below (sentences 3 to 9) provide examples of the diversity of forms that can be used as parentheticals and the wide
range of meanings they can express. For example, in examples 3 and 4 below, the prepositional phrases к сожалению “unfortunately,” and к счастью, “fortunately,” express the speaker’s attitude about context conveyed in the text. In example 5 the speaker uses the verb кажется, “[it] seems,” to express some doubt about the authenticity of statement. The phrase честно говоря, “frankly speaking,” in example 6 indicates that the speaker believes he is being candid and confiding in the listener. In 7, the epistemic adverb очевидно, “evidently,” indicates that the information is not based on fact, but rather on supposition. In example 8 the speaker is addressing the interlocutor directly using the verb phrase поверить ли, “if [you] can believe [it],” in an attempt to dispel any doubt about the veracity of events described. The adverbs во-первых, “firstly,” во-вторых, “secondly” and в-третьих, “thirdly,” are used in example 9 to order the descriptive adjectives and to provide a semantic link between them. In all of the examples below, the parentheticals contribute meaning critical to communication as a whole, and, more specifically, to the interpretation of the context even though they are not part of the “main message” of the host sentence.

3) К сожалению, не ездила. Но очень хочется.\(^1\)

4) Но действительность, к счастью, сложней и интересней любой самой убедительной схемы.

5) По специальности он, кажется, не работал ни дня.

6) Я, честно говоря, не уверен, что сейчас все это происходит не во сне.

7) Те же результаты, очевидно, может дать и массовый социологический опрос.

8) Поверить ли, бабушка, весь народ на ярмарке в пляс пустился.

\(^1\) All Russian examples are from the Russian National Corpus unless otherwise noted.
Morphologically, VS are a diverse group. VS can take the form of single lexical items: nouns (правда, словом), adjectives (главное), adverbs (вероятно, несомненно, кстати) and verbs (знаешь/знаете). Multi-word VS are morphologically quite diverse ranging from prepositional phrases (к счастью, в общем, между прочим), noun phrases (главное дело, иними словами, and verb phrases (скажу тебе, видишь ты, бог даст, иначе говоря, как мне кажется). Verbal VS are a heterogeneous group comprised of finite verb forms (знаешь, видишь) and verb phrases with verbal adverbs or infinitives: собственно говоря, коротко говоря, так сказать, стало быть. Although most VS are fixed phrases (одним словом, словом), some are semi-fixed, containing an inflected form (на чей-либо взгляд). A small number of VS are morphologically and lexically unique; they only function as VS (впрочем, итак, следовательно) (Valgina et al., 2002). Many lexical items that appear as VS also have a non-VS use. Sentences 10-11 show examples of word forms that have both VS use and non-VS use. The A sentences illustrate parenthetical use of VS (VS are bolded). In the B sentences these same words are not parenthetical; they function syntactically and semantically as full sentential constituents.

10A) ВООБЩЕ, хотелось бы знать что произошло.

10B) Он вообще выглядело чудаком.

11A) КАСАЛОСЬ, он был не рад.

11B) Небо казалось тяжёлым.

It should be pointed out that despite the moniker vvodnye slova, “introductory words,” VS can occur in any position within a sentence. Although as a rule VS are more frequent in sentence initial position, each VS has particular distribution patterns. For example, sentences 12-14 below show the distribution patterns of the VS, откровенно говоря, “frankly speaking.” Data from the Russian National Corpus (RNC) establish that of the three positions exemplified below, откровенно
говоря most frequently appears in sentence initial position and is least likely to appear in sentence final position.

12) Откровенно говоря, я даже не слишком удивился.

13) Раньше это был, откровенно говоря, притвор.

14) Красавица, откровенно говоря.

Many вводные слова have become formulaic fixed phrases. These formulaic phrases have undergone a process of grammaticalization whereby the semantic link to constituent lexicon is weakened (Brinton, 1996, 2008; Traugott, 2010). As semantic bleaching progresses, some VS and other similarly grammaticalized connectors become formulaic and take on expanded discourse functions such as, topics shift markers, focus markers, hesitation markers, and so on. The process of grammaticalization can produce slova parazity, “parasite words” that seem to have little or no semantic value and are used with a high degree of frequency in speech. Contrary to popular perception, these “parasite words” are not semantically empty. As a growing body of research has established, these widely used phrases can take on highly nuanced pragmatic functions within spoken discourse (Bogdanova-Beglarian, 2021; Birzer, 2012).

The Russian grammatical tradition, beginning with Vinogradov (2021 [1947]; Vinogradov, 1960), divides the linguistic category of parentheticals into two classes: вводные слова, “introductory words,” and вставные слова, “inserted words,” defined primarily in terms of their function. As Sonnenhauser (2020) points out, the understanding of вводност’, “parentheticality,” in the Russian grammatical tradition is rooted in function and is not an “inherent grammatical feature.” In the 1960 Academy Grammar Vinogradov defines вводные слова as words and phrases that are syntactically independent from the host and serve a metacommunicative or discoursive function. A speaker or writer uses вводные слова to express opinions or evaluations of the utterance or

---

2 For examples see Khachaturyan’s (2010) analysis of the meanings of та́к сказа́ть and Kolyaseva’s (2018, 2022) discussions about the discourse functions of ти́па.
context, to identify the source of the information, and to organize the text. Vinogradov identifies VS as “modal” – they express the writer’s or speaker’s point of view. VS give clues to the interlocutor as to how to interpret the utterance or text (Valgina et al., 2002). They can also serve a pragmatic function, expressing the speaker’s subjective evaluation of the context in which they appear. Grenoble (2004) also points out that the semantics of vvodnye slova are derived not from their lexical meaning, but rather from “their pragmatic interpretation through a formulaic usage, a kind of codification or normalization” of their non-restrictive semantics (p. 1959). Grenoble’s assertion that VS are best defined by function rather than morphological form is confirmed by the lack of consensus among grammarians about what words and word strings are VS. “Comprehensive” lists of VS vary a great deal in content. The functional discoursive functions of VS are discussed in more detail in section 3 below.

Vstavnye slova, unlike vvodnye slova, are not codified and can occur in an “infinite number of possible phrases and clauses” (Grenoble 2004, 1957). Vstavnye slova are described in the Academy Grammar as “insertions in the sentence of a variety of additional information, side commentaries, and so on” (p. 165). As Grenoble (2004) points out, vstavnye slova have conceptual meaning; they affect the truth conditions of the utterance contributing to its propositional content. In written texts they are typically enclosed in parentheses or dashes unlike VS, which are usually marked by commas. Like parenthetical comments in English, vstavnye slova can occur in infinite combinations of words and phrases. Unlike vvodnye slova, vstavnye slova are not formulaic; their meaning is derived directly from the lexical semantics of the constituent words, but they also have a metatextual function. For example, the parenthetical phrases in 15 and 16, separated from the rest of the sentences by parentheses in 15 and by dashes in 16, augment the meaning of the matrix sentence in a fundamental way. These two examples demonstrate the inherent lexical flexibility and heterogeneity of vstavnye slova.

15) Солдаты (их было трое) ели, не обращая внимания не Пьера. (L. N. Tolstoy, War and Peace, as cited in Vinogradov, 1960, II, p. 167)
16) Он встал и, прихрамывая – он был на протезе – подошёл к окну. (Kaverin, Two Captains, as cited in Vinogradov, 1960, II, 166)

The category of *vstavnye slova* is omitted from more recent Russian grammars and pedagogical materials. As Sonnenhausen points out (2020), this trend indicates an “increasing convergence with the general linguistic tradition in more recent descriptions.” *Vstavnye slova* have come to be understood as parentheticals in a general linguistic sense (see above description) and after 1960 are not described in Academy Russian language grammars.

VS are typically categorized into eight semantic groups, described and exemplified below (Vingradov, 1960). The semantic categorization presented below is used in this study to analyze the corpus data.

I) Express emotional assessment or reaction to the context:

16) **К сожалению, в Англии или США не была.**

17) Когда он подплыл ближе, я заметил, к удивлению, плывшего перед лодкой лебедя.

II) Modal meanings: Express certainty or uncertainty about the context:

18) **Очевидно, так оно и было.**

19) Мне, несомненно, повезло с третьим браком

III) Provide information about the source of the information:

20) Страна, по-моему, уже может позволить себе идти на такие реформы.

21) С точки зрения безопасности будет лучше, если это уменьшение будет минимально.
IV) Are used to structure and sequence information, indicating the relationships between elements of the utterance:

22) Он долго ждал, он терпел и, наконец, его прорвало.

23) Это с одной стороны плохо, а с другой хорошо.

V) Are used to express the relationship of the speaker to the utterance itself, or to the manner of the utterance, signaling a shift in style of speech:

24) Иными словами, глупо тратить жизнь, расходуясь на негатив, когда можно наслаждаться позитивом.

25) Короче говоря, он работал, работал и умер.

VI) Used to enhance the expressive nature of the utterance:

26) Честно говоря, я не совсем понимаю, зачем это.

27) Глупый человек, между нами говоря.

VII) Are used to attract the attention of the interlocutor/reader:

28) Всё может измениться, поверьте.

29) Понимаете, в чём дело: Путина окружают разные люди.

VIII) Assert the frequency of an event:

30) Как правило, открыты они с 12 до 2 часа.

31) Бывало, Андрюша заходил к нам после занятий.

In many ways these eight semantic groups determine how VS function in discourse.
3. Vvodnye slova in Discourse

Vvodnye slova are best defined functionally – by the ways they are used by the writer or speaker to interact with speakers and readers. The functions of VS are best captured by the term “metadiscourse,” a term widely used in applied linguistic research examining English academic written discourse (Hyland, 1996; 2005; Hyland & Jiang, 2020). Writing and speaking are fundamentally interactive; an author or speaker uses elements of metadiscourse to establish a relationship with the readers and listeners. Metadiscoursive elements establish the interactional meaning of a text. This interaction can take a variety of lexical, morphological, and semantic forms from direct appeals directed at readers, often in the form of imperatives, to adverbs expressing emotion and prepositional phrases used to structure the text. The common thread holding together the diverse sets of meanings conveyed by metadiscourse is that they communicate meaning outside of the propositional context of the text. As Hyland (2005) points out, metadiscourse has more of a rhetorical and pragmatic function than a primarily semantic one. Writers use it to position themselves in reference to the text and its message and to “shape and construct a text that meets the needs of particular readers” (p. 49).

Discussions of parentheticals often use the terms discourse marker and pragmatic marker to describe many of the functions that VS assume in discourse. There is, however, an ongoing terminological debate about these two difficult to define categories. As Jucker and Ziv (1998) point out, there is no “generally agreed upon definition of the term “discourse marker”” (p. 1). A variety of terms have been used to describe elements that can take the form of procedural parentheticals, these include discourse marker (Shiffrin, 1987), pragmatic marker (Fraser, 1990, 1996; Brinton, 1996), discourse particle (Abraham, 1991), and connective (Blakemore, 2009). For the purposes of this study, I have adopted the definition of procedural parentheticals (VS) provided by Grenoble (2004). According to Grenoble, VS are discourse markers which have a “metacommunicative” function (p. 1968). They signal shifts in the discourse from propositional content to authorial comment, expressing the speaker’s or writer’s attitude towards what is said or written. All of the basic meanings of VS can be described as discoursive in that they are by definition parenthetical, and by the fact that authors use them for two main purposes: 1) to
express their relationship to the propositional content of the text, and 2) to organize the text. Authors use VS semantic types I and II (above) to express attitudes towards or evaluation of the text. Type I VS (examples 16, 17) are used to signal the author’s emotional response to text. Type II VS with modal meanings (examples 18, 19) “highlight the subjectivity of a position,” signaling “the writer’s decision to recognize alternative voices and viewpoints” and “withhold the writer’s commitment to the proposition” (Hyland 2005, p. 52). For example, in sentence 32 the author uses the adverb возможно parenthetically to signal uncertainty about the facts presented.

32) На территории КНДР, возможно, произошли два взрыва.

VS semantic types VI and VII elevate the expressive nature of the text by issuing direct appeals to readers. For example, in sentence 27 VS между нами говоря, the prepositional phrase with the first-person pronominal complement, attracts the attention of the reader by addressing him/her directly. With Type VII VS, typically 2nd person imperatives, the author issues a direct appeal to the readers (examples 28 and 29). All of these VS are “interactional”: the author uses them to comment on the propositional content, conveying his/her reaction to the text (Hyland, 2005, p. 171).

The second main discourse function of VS is to “manage the flow of information” (Hyland, 2005, p. 169-171). VS Groups IV and V (above) are used primarily to structure the informational content of texts. VS from group IV, such as: во-первых, во-вторых are used to organize argument structure, signaling the introduction of individual elements of argumentation. Group IV VS, such as итак, следовательно, наконец, таким образом, are used to establish transitions between the information presented, establishing relationships between parts of the text. They are used to structure arguments and to provide cohesion to the text.

Contemporary models of text comprehension (both L1 and L2) posit that text comprehension is a complex process whereby readers create mental representations of text (Kintsch, 2012). In order to produce a mental representation, readers must establish coherence
relations between chunks of text. The term “coherence” refers to the “ways in which the various parts of what a text says is linked together” (Bailin & Grafstein, 2016, p. 131). VS and other connectors provide cues about how texts are organized. These linkers lend text coherence by conveying underlying semantic relationships between parts of a text, setting up logical relationships such as sequence, relative importance of text elements, and exemplification. There is a substantial body of research indicating that connectives play a critical role in discourse comprehension enabling both L1 and L2 readers to establish coherence within a text. A number of studies investigating text comprehension in L1 and L2 learners indicate that connectors (including rhetorical organizers like VS) foster faster processing times and more coherent mental representations of written texts. Kintsch and Yarbrough’s (1982) study of L1 English-speaking university students found that textual organizers improved subjects’ ability to comprehend the rhetorical structure of texts. Cain and Nash’s 2011 study of 8–10-year-old L1 English readers provided evidence that connectives aid text processing in “typically developing” L1 readers. They found that the presence of connectors in text enhanced reading speeds. Silfhout et al.’s 2014 study of Dutch L1 learners found that the presence of textual coherence makers improved comprehension and processing time across genres. Mills and Just (1994) found that the use of connectives to connect two clauses in written texts improved text comprehension in the second clause. In Degand & Sander’s (2002) study of L1 French students learning Dutch and L1 Dutch speakers learning French, text connectors improved text comprehension in both L1 and L2 learners. Although equivalent comprehension studies of L1 and L2 learners of Russian have yet to be done, there is substantial cross linguistic evidence about the importance of connectors, including procedural parentheticals, for the comprehension of written discourse. VS are an important category of text connector providing critical cues about rhetorical structure, the sequence of text elements, and authorial attitudes towards propositional content. Learners who fail to identify VS are less likely to be able to identify supporting elements of argumentation, and they are more likely to misunderstand authorial intent and to misconstrue the text’s message. Given the important role they play in discourse, VS should be included in L2 curricula.
4. Vvodnye slova in L2 Russian Pedagogy: Why
Given the fact that vvodnye slova typically contribute meaning that is typically outside of propositional content, one could simply deem them peripheral and exclude them from L2 pedagogy. Russian language L2 pedagogical texts as a rule do not discuss vvodnye slova as a category, but rather introduce them as individual word combinations— as useful gambits that can be used in speech and writing. Given the lexical and semantic heterogeneity of VS and the challenge of making generalizations about them, it is not surprising that designers of L2 pedagogical materials neglect them.

There are a number of compelling arguments for the inclusion of the category of VS as a category in L2 curricula. First and foremost is the critical role that procedural parentheticals functioning as comment clauses to facilitate text cohesion and to improve text comprehension and reading speed. Procedural parentheticals are generally omitted from most Russian language pedagogical material. When they are included, they are typically presented in the form of lists as expressions to be memorized. The problem with limiting L2 learners’ exposure to VS and other comment clauses to vocabulary lists for rote memorization is that this approach ignores the important systemic role they play as text and rhetorical organizers.

There is another perhaps more basic argument to be made for including VS in L2 curricula. L2 learners can draw direct parallels to procedural parenthetics from their L1. While the meaning of VS type connectives from L1 to L2 may not completely overlap, their overall function as cohesive devices and rhetorical organizers is the same. In learning and practicing the use of VS in Russian L2 learners can draw from their understanding of how equivalent parenthetical connectors function in their native language. The notion of text cohesion and organization is rarely presented in pedagogical material for L2 Russian learners and when it is it is limited to more advanced textbooks specifically focused on the development of writing skills. It is assumed that any discussion of text cohesion should be postponed until learners have reached advanced stages of learning. However, since L2 learners already have some innate knowledge of how to use VS-type connectors in L1, introducing this category early would allow them to begin using them at lower levels of proficiency.

---

3 For example, see the presentation of VS and other connectors in pedagogical texts see Kashper et al. (2006).
The third argument for including VS in L2 Russian language curricula is that there is substantial evidence that explicit instruction about connectors improves both reading comprehension and reading speed of written discourse for both L1 and L2 learners (Kintsch & Yarbrough, 1982; Silfhout et al., (2014); Millis & Just, 1994; Cain & Nash, 2011). The role of connectors in comprehension of spoken language has yet to be investigated.

Given these strong arguments for the inclusion of VS in L2 Russian language curricula, the next question is which VS should be included? The corpus-based study examining the relative frequency of different types of VS presented in the next section provides some answers to this question.

5. Corpus-based Study
This study uses data from the Russian National Corpus to establish relative frequency counts for *vvodnye slova*. The goal of this quantitative study is to ascertain which semantic and lexical types of VS are used most often. The study shows that VS made up of a single word are used more frequently than VS composed of more than one word, including VS consisting of fixed phrases. Data was also gathered across semantic groups of VS. This data should be useful for curricula designers and L2 teachers, enabling them to decide which VS should be introduced to learners and at what level of instruction.

5.1. Methodology
In order carry out this corpus-based frequency study, a number of problems had to be solved. The first issue was how to select VS for inclusion in the study, given the fact that virtually no two lists of VS, no matter how comprehensive they claim to be, are alike. In order to maximize the scope of focus, this study was based on two “comprehensive” lists of VS 1) the list of VS provided in the Russian National Corpus with 217 VS and 2) the list of VS in Morkovkin (1997) with 208 VS. The final combined list of VS totaled 316 tokens. The RNC was used to compare the frequency, composition, semantics, and uses of the 316 *vvodnye slova*.4

4 The “old” version of the RNC used for this study was not set up to separate written texts from the RNC as a whole. For this reason, no comparison of frequency of VS in written texts to oral discourse was possible for this study. The data from the RNC as whole is included in this study as well as frequency data from the oral subcorpus.
In order to analyze VS, several challenges needed to be overcome. Many VS, as illustrated in sentence 10-11 above, particularly those consisting of a single lexical item, can overlap in form with uses of that word which are not parenthetical. It was critical for this study to distinguish VS uses from non-VS, particularly for tokens of high frequency. Due to the sheer volume of data, the only practical solution was to rely on the Russian National Corpus to mark and identify which words and expressions are VS and which were not. Unfortunately, although the RNC consistently identifies single word adverbial VS, single word finite verbal VS are rarely marked as VS. For this reason, single word verbal VS, except for a few exceptions which were marked as such in the RNC, are not included in this study and are omitted from the frequency list.\(^5\)

There are currently two versions of the RNC available on the Internet. The default or “old” version has a comprehensive list of VS linked to the RNC search engine. This study used the “old” version of RNC, containing texts up to 2014. The “new” version, at least at the time this paper was written, does not contain a list of frequency statistics for VS. One of the goals of this study was to update the frequency list of VS in the RNC, which at present is based on data up to 2014.

5.2. Vvodnye slova: Frequency Data
This study examined frequency data for 316 vvodnye slova obtained from the Russian National Corpus (RNC).\(^6\) The RNC was used to establish normed frequency counts for all VS across all registers and all genres, and to identify the 150 most frequently used VS.\(^7\) Normed frequency counts in this study are based on the frequency of use per 10,000 words. Normed frequency counts (per 10,000 words) corroborate that VS are used more frequently in oral speech. VS in the oral database occur with a frequency of 19 tokens per 10,000 words, while in the RNC as a whole

\(^5\) The RNC does not mark any finite verbal VS. 14 finite verbal VS listed in Morkovkin were omitted, including: бывало, говорят, знаешь, знаете, может, etc.
\(^6\) It is important to note that the frequency data listed in RNC (old version) lists frequency statistics for texts added up to 2014.
\(^7\) Normalized text frequency is a way to adjust frequency counts from texts of different lengths. In order to obtain normed frequency counts per 10,000 words for this study frequency data was divided by the total number of words in the corpus and multiplied by 10,000.
VS average 10 uses per 10,000 words. Although the reasons for the more widespread use of VS in speech require further study, a number of possible explanations can be postulated based on the data in this study. A high percentage of VS, especially of the most frequently used VS in both written and oral registers, are modals. Biber’s corpus-based study of English (1999) established that modals (not necessarily parenthetical modals) are generally used more frequently in speech than in writing. If this is true for Russian as well, it could suggest that the higher use of VS in speech simply mirrors the higher frequency of modals in speech as a whole.

Another possible explanation for slightly higher frequency of VS in speech is the tendency of some VS to develop into слова паразиты, “parasite words.” Some VS that appear high up on the frequency list for oral texts, such as значит, так сказать and короче, have undergone a process of grammaticalization. As semantic ties to these lexical items have weakened, they have assumed pragmatic functions taking on broader roles in the discourse such as topic markers, markers of topic shift, and marking speaker’s distance from their interlocutors, and other functions (Bogdanova-Beglarian, 2021; Khachaturyan, 2010; Kolyaseva, 2018, 2022).

Table 1 contains data pertaining to VS in the entire RNC, presenting information about the frequency of single and multi-word VS. For the purposes of this study, multi-word VS are identified by the linguistic definition of “multi-word items” provided by Moon (1997, p. 43): they are vocabulary items consisting of a sequence of two or more words that “semantically and/or syntactically form a meaningful and inseparable unit.” Multi-word VS are typically fixed phrases. Formulaic multi-word VS are distinct from collocations, which are lexical word associations that co-occur with a few other words to produce such frequently occurring lexical phrases as “blond hair,” “strong tea.” Multi-word VS are also distinguished from idioms in that the latter consist of lexical strings which are semantically non-compositional – their semantic meaning is not a sum of their parts (buy the farm). Multi-word VS are compositional in that their meaning is derived from lexical items that comprise them (к счастью). A small number of VS are semi-fixed expressions, allowing replacement of one or more of their components such as for example на (чей-л.) взгляд.
Data in Table 1 indicate that in the RNC multi-word VS are more frequent than single word VS. However, when we examine the list of the most frequently used VS, single word VS are predominant. As Table 1 below shows, 72% of the twenty-five most frequently used VS consist of a single lexical item, while only 28% are made up of more than one word. Out of the full list of 316 VS used for this study, only 14% of all VS are made up of a single word. Data collected for this study indicate that 70% of the fifty most frequent single word VS are adverbs. Most of them are either epistemic modals or are adverbs used to structure and sequence information, indicating the relationships between elements of the utterance (наконец, например).

Table 1: Frequency of Single Word/Multi-Word VS by Percentage in the RNC

<table>
<thead>
<tr>
<th>Frequency of Multi-Word/Single Word VS By Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 Most Frequent VS</td>
</tr>
<tr>
<td>50 Most Frequent VS</td>
</tr>
<tr>
<td>100 Most Frequent VS</td>
</tr>
<tr>
<td>150 Most Frequent VS</td>
</tr>
<tr>
<td>ALL VS IN RNC: 316</td>
</tr>
</tbody>
</table>

Table 2 presents the frequency of single and multi-word VS in RNC oral texts. The oral data shows several similarities with that of the RNC as a whole. As in the RNC, multi-word VS are much more frequent than single word VS in the oral database. However, single word VS (single word VS with modal meanings in particular) predominate among the 25 most frequently used VS in the oral database. The 100-150 most frequently used VS in oral texts are more likely to be multi-word in composition. A possible explanation for this tendency is the phenomenon of grammaticalization and concomitant pragmatic widening. As multi-phrase fixed phrase VS are used more frequently, the semantic link to the lexical items that make up the phrase weakens, and the phrase takes on pragmatic meanings and in turn becomes even more frequently used.
Table 2: Frequency of Multi-Word/Single Word VS by Percentage in Oral RNC Texts

<table>
<thead>
<tr>
<th>Frequency of Multi-Word/Single Word VS By Percentage in RNC Oral Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 Most Frequent VS</td>
</tr>
<tr>
<td>50 Most Frequent VS</td>
</tr>
<tr>
<td>100 Most Frequent VS</td>
</tr>
<tr>
<td>150 Most Frequent VS</td>
</tr>
<tr>
<td>ALL VS IN DATABASE: 316</td>
</tr>
</tbody>
</table>

The lists of the 25 most frequently used VS in the RNC and in the oral texts of the RNC in Table 3 exhibit several important features. While a more comprehensive review of semantic types of VS represented in the database follows below, Table 3 reveals the predominance of both modal and single-word VS in both lists. What is particularly striking about Table 3 is the high degree of overlap between the two data sets: half of the same VS appear in the top ten of both lists. The list of VS in oral texts not surprisingly contains a number of word “parasites.” These are VS that have taken on expanded pragmatic meaning as the semantic link to their constituent lexical items has weakened, these include: значит, в общем, так сказать and короче. Both lists contain a number of VS used to organize arguments. The RNC list, however, contains more text organizers (наконец, например, главное, таким образом, прежде всего and кроме того).

Many of the same trends evident in Table 3 are also apparent in Table 4 showing the distribution of meaning types in the RNC. Tables 1 and 4 show that the most frequently used VS in the RNC are single word in composition, typically adverbs expressing epistemic modality or adverbs used to sequence information or argument elements. Research on the authorial expression of evaluation in English may provide a possible explanation for the high frequency of epistemic adverbs used as VS (Grey & Biber, 2012). Adverbs denoting probability, possibility, and certainty, are typically used in non-fiction English texts to express hedges. Hedging is a rhetorical strategy, “by which a speaker, using a linguistic device, can signal a lack of commitment to either the full semantic membership
of an expression or the full commitment to the force of the speech act being conveyed” (Fraser, 2010, p. 22). Authors use hedges to imply that a statement is presented as based on opinion or, in some cases, as plausible reasoning rather than fact, providing readers with the opportunity to interpret the text in their own way. Hedges allow the author to appear open to other points of view (Frazer, 2010).

Table 3. 25 Most Frequently Used VS in RNS and in Oral Texts

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Most Frequent VS in entire RNC</th>
<th>*= use per 10,000</th>
<th>Most Frequent VS in Oral Texts</th>
<th>*= use per 10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>конечно</td>
<td>*****</td>
<td>конечно</td>
<td>*********</td>
</tr>
<tr>
<td>2.</td>
<td>может быть</td>
<td>****</td>
<td>значит</td>
<td>*********</td>
</tr>
<tr>
<td>3.</td>
<td>однако</td>
<td>***</td>
<td>может быть</td>
<td>*********</td>
</tr>
<tr>
<td>4.</td>
<td>наконец</td>
<td>***</td>
<td>с другой стороны</td>
<td>*********</td>
</tr>
<tr>
<td>5.</td>
<td>лучше</td>
<td>***</td>
<td>в общем</td>
<td>*****</td>
</tr>
<tr>
<td>6.</td>
<td>кажется</td>
<td>***</td>
<td>действительно</td>
<td>*****</td>
</tr>
<tr>
<td>7.</td>
<td>значит</td>
<td>***</td>
<td>например</td>
<td>*****</td>
</tr>
<tr>
<td>8.</td>
<td>например</td>
<td>***</td>
<td>лучше</td>
<td>*****</td>
</tr>
<tr>
<td>9.</td>
<td>действительно</td>
<td>**</td>
<td>наверное</td>
<td>****</td>
</tr>
<tr>
<td>10.</td>
<td>казалось</td>
<td>*</td>
<td>мне кажется</td>
<td>****</td>
</tr>
<tr>
<td>11.</td>
<td>возможно</td>
<td>*</td>
<td>по-моему</td>
<td>****</td>
</tr>
<tr>
<td>12.</td>
<td>главное</td>
<td>*</td>
<td>так сказать</td>
<td>****</td>
</tr>
<tr>
<td>13.</td>
<td>таким образом</td>
<td>*</td>
<td>понятно</td>
<td>****</td>
</tr>
<tr>
<td>14.</td>
<td>верно</td>
<td>*</td>
<td>наверное</td>
<td>***</td>
</tr>
<tr>
<td>15.</td>
<td>наверное</td>
<td>*</td>
<td>кстати</td>
<td>***</td>
</tr>
<tr>
<td>16.</td>
<td>вероятно</td>
<td>*</td>
<td>кажется</td>
<td>***</td>
</tr>
<tr>
<td>17.</td>
<td>прежде всего</td>
<td>*</td>
<td>в принципе</td>
<td>**</td>
</tr>
<tr>
<td>18.</td>
<td>кроме того</td>
<td>*</td>
<td>естественно</td>
<td>**</td>
</tr>
<tr>
<td>19.</td>
<td>кстати</td>
<td>*</td>
<td>во-первых</td>
<td>**</td>
</tr>
<tr>
<td>20.</td>
<td>словом</td>
<td>*</td>
<td>к сожалению</td>
<td>**</td>
</tr>
<tr>
<td>21.</td>
<td>видимо</td>
<td>*</td>
<td>главное</td>
<td>*</td>
</tr>
<tr>
<td>22.</td>
<td>в (таком, этом, данном) случае</td>
<td>*</td>
<td>короче</td>
<td>*</td>
</tr>
<tr>
<td>23.</td>
<td>пожалуй</td>
<td>*</td>
<td>видимо</td>
<td>*</td>
</tr>
<tr>
<td>24.</td>
<td>должно быть</td>
<td>*</td>
<td>возможно</td>
<td>*</td>
</tr>
<tr>
<td>25.</td>
<td>в общем</td>
<td>&lt;*&gt;</td>
<td>наконец</td>
<td>*</td>
</tr>
</tbody>
</table>
The prevalence of adverbial hedges in English academic prose has been established in a number of studies (Hyland 1994, 1996, 2005; Hyland & Jiang, 2020; Biber, 1999). Breitkopf-Seipman (2012) observes that authors of English academic prose use hedges to avoid criticism. The use of hedges in Russian non-fiction texts has not been fully investigated.8

Table 4: Distribution of VS Meanings in the RNC by percentage

<table>
<thead>
<tr>
<th>Meanings</th>
<th>Distribution In the RNC</th>
<th>Distribution in the 25 most frequent VS</th>
<th>Distribution in the 50 most frequent VS</th>
<th>Distribution in the 75 most frequent VS9</th>
<th>Distribution in the 100 most frequent VS10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modal (II)11</td>
<td>27%</td>
<td>52%</td>
<td>42%</td>
<td>36%</td>
<td>35%</td>
</tr>
<tr>
<td>Emotion (I)</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence Info (IV)</td>
<td>17%</td>
<td>44%</td>
<td>46%</td>
<td>40%</td>
<td>35%</td>
</tr>
<tr>
<td>Expressive (VI)</td>
<td>15%</td>
<td></td>
<td></td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>Style Shift (V)</td>
<td>7%</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
<td>8%</td>
</tr>
<tr>
<td>Attract Attention (VII)</td>
<td>8%</td>
<td></td>
<td></td>
<td>1%</td>
<td>4%</td>
</tr>
<tr>
<td>Source of Info (III)</td>
<td>6%</td>
<td></td>
<td></td>
<td>8%</td>
<td>6%</td>
</tr>
<tr>
<td>Frequency of Events (VIII)</td>
<td>&lt;1%</td>
<td></td>
<td></td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

8 For discussions of hedges in Russian academic prose see also Breitkopf (2005, 2006).
9 Due to rounding the total of this column does not equal 100.
10 Due to rounding the total of this column does not equal 100.
11 Roman numerals in the table refer to the semantic groupings of VS presented in part 2 of this paper.
The distribution data in Table 4 show that among the most frequently used VS, single and multi-word expressions used to establish the sequence of information and arguments (наконец, таким образом) are the second most widely used type. The high frequency of VS in written texts may be particularly marked in written texts with a rhetorical or academic purpose, although this remains to be explored.

While VS have a wide range of uses within the RNC as a whole, the distribution lists for the most frequently used VS included in this study can inform L2 curricula designers and language instructors about which types of VS to include in pedagogical materials and which should be prioritized. Of the 316 VS examined, adverbs expressing epistemic modality and adverbs used to sequence information are the most frequently used VS. Many of the epistemic adverbs used to express hedges have direct equivalents in English. L2 curriculum designers and instructors can draw from these parallels in targeting these highly used expressions.

Given the “fuzziness” of vvodnye slova as a category, it is not surprising that VS are mostly omitted from L2 textbooks for Russian as a Foreign Language. Not only are VS difficult to define, no two lists of VS are the same. One could argue that they are of secondary importance semantically since they introduce meanings that do not affect the propositional content of the sentences in which they appear. The data from this study establishes that the most frequently used types of VS are single word in composition, taking the form of epistemic and other adverbs, expressing either authorial evaluation or marking the information structure within a discourse. Learners that cannot identify VS are more likely to misinterpret authorial intent. Learners lacking mastery of VS functioning as text organizers are likely to have more difficulty in processing texts and in comprehending rhetorical arguments. For these reasons, VS should be included at all levels of Russian language instruction.

A strong argument can be made for including VS in curricula for advanced L2 learners. At more advanced levels, where learners are acquiring the lexical and syntactic tools to engage in persuasion in oral and written form, the inclusion of VS enable L2 readers and writers to develop more effective active and passive rhetorical skills. VS are a key tool for
organizing argumentation in discourse and are used extensively in writing and in speaking to express authorial stance and to organize arguments. Knowledge of VS is essential for the development of pragmatic competence. Lacking a working knowledge of the more frequently used VS, learners are more likely to miss the important clues that VS provide about the author’s relationship to the text, a critical element of text meaning. Mastery of VS gives L2 students the building blocks to construct effective argumentation in written and oral form. Studies in undergraduate writing in English show that the ability to express authorial stance is a crucial indicator of writing quality. Studies indicate that EFL students who receive instruction about how to use metadiscourse write “significantly better texts” (Cheng & Steffensen, 1996), and there is evidence that pragmatically procedural parentheticals are “indispensable” to interpreting academic discourse (Ifantidou, 2004, p. 1325). Few would argue with the importance of VS in academic prose. The introduction of VS in advanced L2 pedagogical material, particularly in pedagogical material targeted towards the development of writing skills, is generally accepted.

A strong case can also be made for including VS in lower-level instruction. Some of these arguments have already been presented. First, L2 learners can draw from their understanding of parenthetical connectors in their L1, many of which are similar in meaning and function. The second argument for introducing VS in lower-level instruction is rooted in the journey of L2 learners to develop language proficiency. That journey is a gradual one. The mastery of the syntactic, morphological, lexical, and cohesive devices needed to produce and comprehend advanced texts are not acquired overnight. Introducing the concept of parenthetical language to learners early in their study of Russian will provide them with powerful tools for expanding their semantic and syntactic repertoire.

The most frequently used VS, epistemic modal adverbs and adverbs used to organize argumentation, are undeniably important tools for rhetorical discourse. However, they need to be introduced at earlier stages of proficiency to foster active mastery. Beginning as early as intermediate level, according to the ACTFL guidelines, learners need to be able to formulate and understand comparisons. Even a Novice level student who has mastered the VS с одной стороны and с другой стороны can begin to express comparisons and thereby begin the long journey to acquire skills required for effective argumentation.
Another important issue is the methodology that should be followed in introducing VS to L2 students. Based on the data from this study three organizing principles for VS instruction are recommended. First, all VS uses and functions should be extensively modeled for learners. Modeling should include actual uses of VS in discourse, providing examples that they can learn from and emulate. Sentencelength examples cannot show the range of VS functions in the discourse as a whole. Students need to be provided with examples of VS in extended discourse in order to understand and master the critical cues VS as rhetorical organizers. Second, whenever possible L2 instruction should draw parallels with similar expressions in learners’ L1 and make these similarities implicit in instruction. In addition, those expressions that have direct semantic equivalents in L1 should be prioritized and introduced earlier. Pedagogical materials should provide lots of examples in larger chunks of discourse. Once VS have been modeled, initial phases of instruction should require students to identify VS and examine their function and meaning within a discourse. The third principle to be followed in introducing VS is that learners should be provided with ample opportunities to actively use VS in discourse.

One effective way of structuring the first stage of active use is to present students with texts that have been separated into multi-sentence parts and ask them to use VS to combine these parts into a complete discourse. These three components: extensive modeling, drawing parallels with L2, and structuring lessons to require active use of VS will provide a solid framework for the development of functional mastery of procedural parentheticals, important building blocks of effective argumentation.

7. Conclusion
This article reviews our current understanding of *vvodnye slova* and their main functions. VS have been typically deemed of peripheral importance for L2 Russian language learners because they are parentheticals, providing meaning outside of propositional structures. This review of the main functions of *vvodnye slova*, their meanings, and most frequent uses, suggests that VS play an important function in text cohesion and comprehension. A strong case has been made for including VS in L2 Russian language curricula and suggestions have been provided about effective methodologies for
introducing vvodnye slova to the Russian language classroom.

References


Fraser, B. (2010). Pragmatic competence: The case of hedging. In G. Kaltenböck, M. Mihatsch, & S. Schneider (Eds.), *New Approaches to Hedging* (pp. 16–34). Emerald.


A Corpus-based Analysis of *vvodnye slova*: Pedagogical Applications
JOAN F. CHEVALIER


