Types of information for the multimedia dictionary of Russian discourse markers

I. M. Kobozeva, L. M. Zakharov
Moscow State Lomonosov University, Russia

The thorough study of discourse markers (DM), i.e. words and phrases bearing mainly discourse-pragmatic information (sentence adverbs, particles, interjections and the like) have started in seventies and continues at present. In semantics, pragmatics and discourse analysis detailed descriptions of many of DMs based on different theories have been proposed. Special dictionaries of DMs were published with word-lists ranging from dozens to hundreds. Dictionaries of Russian DMs give information on their phonetic, grammatical, syntactic and pragmatic properties. However this information is insufficient for such purposes as learning/teaching Russian as a foreign language and for automated text or speech processing. Generally there is not enough phonetic (prosodic), syntactic and paralinguistic information needed for correct speech production (synthesis) and speech understanding (analysis).

We argue that the most natural way to proceed is to make a computer multimedia dictionary that would supply information concerning every aspect of DM that has to be taken into account. We propose a format for such a dictionary and discuss relevant types of information, concentrating on those that are poorly presented in existing paper dictionaries.

General and specific information on DM.

DM is identified in the dictionary by its standard graphic form that serves as an input to the hierarchically organized lexical entry, covering various uses of a given form. Such uses differ significantly in their phonetic, grammatical and semantic properties. The diversity of these uses raises the problem, that A. I. Smirnitsky called "the problem of the word identity": are these uses contextual modifications of one and the same lexical unit or are they homonyms? We believe (together with many other researchers) that this problem is irrelevant for DMs because the existing criteria for the distinction of polysemy and homonymy are inapplicable to words of this kind. On the one hand criteria that determine lexicogrammatical category, or part of speech of a functional word are far from clear (at least for Russian).

This fact is reflected in lexicographical practice, where there are numerous cases of controversial category ascriptions (See e.g. DM voobsche. In MAS and [1] all its uses are classified as adverbs, in [2] one of them is characterized as a parenthetical word and a particle at the same time, and in [3] two of its uses are considered as particles. One and the same use of DM da in [3] is marked as a particle, while in [1] it is marked as interjection.) of one and the same DM. On the other hand the use of a purely semantic criterion (the degree of semantic similarity) is ineffective too, because semantic 'atoms' into which meanings of DMs are decomposed are rather abstract and in a way they are all similar for they belong to the same semantic domain including elements of communicative situation and relations between them. In such circumstances it seems rational to treat diverse functional uses of the same word form as variants of one and the same word and describe them in one complex entry, just as it is done in [4]. The number and the nature of these variants as well as relations between them are presented at the beginning of an entry in a form of a synopsis, as proposed by Ju. D. Apresian for lexical entries of the integrated linguistic description (see [5], [6: 485–537]). In the synopsis each use of DM is characterized by its grammatical category label (or labels, if there is disagreement among specialists), simplified formulation of its meaning expressing the main idea underlying this kind of DM usage and a short typical example of such a use. Thus the synopsis for DM voobsche is given in (1):

1. 'in general; ignoring individual characteristics': perevod voobsche i xudozhestvennyj perevod v osobennosti "translation in general and literary translation in particular"

2. 'marker of a general statement in the presence of a particular deviation': Ja, voobsche, zanjet, no esti nadao, ja priedu. "As a matter of fact I am busy but I shall come if needed."

3. 'marker introducing generalization after mentioning some particular case(s)': On brosil uchebu i voobsche vedet sebja kak-to stranno. "He gave up his studies and on the whole exhibits strange behavior."

4. 'in any circumstances': Po vyxoynym oni voobsche ne vykljuchajut televizor. "On weekends they do not turn off TV at all"

5. 'marker of the ultimate degree in some hierarchy' Reka ot doma nedaleko, a prud voobsche v pjiut minutax xodu. "The river is not far from the house and the pond is even closer: in five minutes walk from there."

6. 'expression of the emotional reaction (positive or negative) to the observed or discussed situation, that the speaker considers to be the extreme case of its kind' Voobsche! "That beats me!"

In spite of numerous phonetic, grammatical and semantic distinctions different variants of the same DM form a unity that is reflected in their common properties and characteristics. In order to capture this unity of a word general information that remains constant through all its uses is distinguished from specific information, related only to one or more variants. That is why information on the same aspect of DM (e.g. phonetic or semantic) may appear on two levels: general and specific.
Structure of the lexical entry

On both levels — general and specific — various properties of DM are presented. The information about these properties is divided into several zones according to the linguistic aspect that it belongs to. At present we the data on DM are packaged into the following zones:

(2) 1. Graphic information
   1.1. Spelling
   1.2. Punctuation
   2. Phonetic information
   2.1. Transcription
   2.2. Prosodic information
   2.2.1. Word prosody
   2.2.2. Phrasal prosody
   2.2.3. Sound files with their visualization confirming the given phonetic characteristics
   3. Syntactic information:
   3.1. Linear position (illustrated by well formed and ungrammatical sentences)
   3.2. Possibility of independent use
   3.3. Argument structure with restrictions on arguments
   3.4. Regularly co-occurrence with other DMs:
   3.4.1. free, e.g. ved’ + zhe, (in entries of both DMs)
   3.4.2. idiomatic, e.g. vot eshche, edva li (in entries of the first elementary DM)
   4. Semantic information
   4.1. Explication (definition) of meaning
   4.2. Paradigmatic relations
   4.2.1. Synonyms and analogs (e.g. kak takovoj for voobsche 1, voobsche-to for voobsche 2 and sobsem for voobsche 4)
   4.2.2. Correlative expressions (e.g. v osobennosti, v chastnosti for voobsche 1)
   4.2.3. Antonyms (e.g. imeno for voobsche 1)
   5. Communicative information:
   5.1. Relation to topic/focus opposition
   5.2. Relation to given/new opposition
   5.3. Relation to contrast, emphasis etc.
   6. Pragmatic information
   6.1. Style markers, e.g., colloquial, bookish etc. (neutral by default)
   6.2. Restrictions on illocutionary force of the utterance
   6.3. Meaning modifications within specific kinds of illocutions
   7. Paralinguistic information
   7.1. Accompanying facial expressions
   7.2. Accompanying gestures
   8. Derivational information (here words derived from the given DM are presented, e.g. “diminutive” forms for one variant of A — Ajuški! and Ain’ki!, verb poddakvat’ for confirmative variant of Da).
   9. References to selected bibliography

The structure of the entry can be expanded as new types or kinds of relevant properties emerge during the study of DM linguistic behavior. In what follows we address the kinds of information concerning the surface form of the DM, because it this kind of information is under-specified in present dictionaries [7].

Information on the surface form of the DM (“significant”)

The dictionary should supply data about all graphic and phonetic forms of a given DM for each meaning listed in its synopsis.

1. Graphic information.

The zone of graphic information consists of spelling and punctuation.

1.1. Spelling. In case the spelling of a DM is constant it belongs to general information zone of its entry and in case this spelling is unique it serves as an input to the entry and not repeated in graphic zone. But often a given DM has spelling variants specific to one or more of its functions. In such a case the standard (“input”) variant is repeated in the general graphic zone, if it is possible for all semantic functions of a given DM and in specific graphic zones of those functions for which it is appropriate while function specific spelling variants are given in the graphic zone of the corresponding functions. E. g. DM A has a standard variant appropriate for all its meanings and two function specific variants. Information about the standard variant is given as general information of the entry, and variants such as A-a and A-a-a are assigned only to such functions as ‘reaction of understanding’ and ‘exclamation at seeing (and being able to seize) what one was after’.

1.2. Punctuation. This zone is relevant for the DMs that are associated with particular punctuation in all or some of their uses. In this zone general or specific punctuation norms are stated and illustrated. In future we hope to come to a punctuation classification of DMs that will be given in an introductory part of a dictionary. Then we shall only have to mark the corresponding class of an item in question. Thus da in all its uses as a positive communicative response to a number of speech acts obey general punctuation rules for main sentences, i. e. it must either be followed by a sentence final punctuation sign (full stop, exclamation mark) or isolated from the rest of the utterance by a comma (cf. Da. Da! Da, ja gotov). Some other DMs (or their variants) that may function as independent utterances (e.g. net, vot, tak, a etc) have the same punctuation pattern. All such DMs can be assigned to one punctuation class coded as 1 or STATEMENT and marked as such in the dictionary.

2. Phonetic information

Phonetic specifications include transcription of all the standard pronunciation variants of the word and prosodic information.

2.1. Phonetic transcription. For many DMs (e. g. ved’ [v’et’], dazhe [dazh+], vose [vöö+s]) transcription belongs to general information section of the entry. But some DMs have pronunciation variants specific to one or more of their meanings. Thus DM A in its use as an initial particle, opening a turn (generally reactive) in a dialogue is normally pronounced [a], and
In its use as a response to "What's the answer when it's so late? - I'll be back soon," the answer is normally pronounced [eɪ].

**4. a. Oohk (look). b. Sox.**

"a. ooh, look at that! b. oh, okay."

2.1 Prosody

Here such prosodic properties are given that are fixed for DM as a lexeme or for its lexeme-like semantic readings in the context of the sentence. This includes some marked stress, some marked intonation, and some marked stress. In the context of the sentence, this includes some marked stress, some marked intonation, and some marked stress.

The lexeme DM as a whole or its semantic variants are generally associated with one or more intonational labels. T. M. Nikolaeva (1978) notes that for such intonational labels, such as DM as a whole or its semantic variants, it is common to mark these features with higher stress and lower frequency of change. 

TheTheme on (in) (in) constitutes a theme in the prototypical noun phrase and has a falling tone. In the Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone. In the Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone. In the Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone.

For example, if a DM is marked for high frequency, it is often marked for high frequency, but not for low frequency. However, if a DM is marked for low frequency, it is often marked for low frequency, but not for high frequency. Therefore, it is important to consider the prosodic properties of DMs when determining their intonational labels.

2.2 Phrasal stress

The Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone. In the Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone. In the Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone. In the Theme on (in) (in) constitutes a Theme in the prototypical noun phrase and has a falling tone.

For example, if a DM is marked for high frequency, it is often marked for high frequency, but not for low frequency. However, if a DM is marked for low frequency, it is often marked for low frequency, but not for high frequency. Therefore, it is important to consider the prosodic properties of DMs when determining their intonational labels.

In summary, the prosodic properties of DMs are important in determining their intonational labels. These properties can include stress, intonation, and frequency, and they vary depending on the context in which the DM is used. Understanding these properties can help language learners to better comprehend and produce natural-sounding speech.

---

**Excerpt from a linguistic text:**

It is well established in the field of linguistics that Russian DMs have special categories with [a]. Prosodic features, such as stress, duration, and pitch, play a crucial role in distinguishing different DMs. One way to approach this is to study the distribution of prosodic features across different contexts.

One such context is the distribution of prosodic features within the sentence. For example, the beginning of a sentence often has a falling tone, while the end of a sentence often has a rising tone. This distribution is consistent across different types of DMs, such as personal pronouns, interrogatives, and declaratives.

Another context is the distribution of prosodic features across different parts of the sentence. For example, the beginning of a noun phrase often has a higher stress than the end of a noun phrase. This distribution is consistent across different types of DMs, such as personal pronouns, interrogatives, and declaratives.

By studying the distribution of prosodic features across different contexts, we can gain a better understanding of how Russian DMs are used in natural speech. This knowledge can be applied to language instruction, language testing, and other areas of linguistic research.

---

**Additional notes:**

1. Prosodic features, such as stress, duration, and pitch, play a crucial role in distinguishing different DMs. One way to approach this is to study the distribution of prosodic features across different contexts.

2. The distribution of prosodic features within the sentence is consistent across different types of DMs, such as personal pronouns, interrogatives, and declaratives.

3. The distribution of prosodic features across different parts of the sentence is consistent across different types of DMs, such as personal pronouns, interrogatives, and declaratives.

By studying the distribution of prosodic features across different contexts, we can gain a better understanding of how Russian DMs are used in natural speech. This knowledge can be applied to language instruction, language testing, and other areas of linguistic research.

---

**Cross-references:**

- **[a]**
- **[b]**
- **[c]**
- **[d]**
- **[e]**
- **[f]**
- **[g]**
- **[h]**
- **[i]**
- **[j]**
- **[k]**
- **[l]**
- **[m]**
- **[n]**
- **[o]**
- **[p]**
- **[q]**
- **[r]**
- **[s]**
- **[t]**
- **[u]**
- **[v]**
- **[w]**
- **[x]**
- **[y]**
- **[z]**

---

**Conclusion:**

In summary, the prosodic properties of DMs are important in determining their intonational labels. These properties can include stress, intonation, and frequency, and they vary depending on the context in which the DM is used. Understanding these properties can help language learners to better comprehend and produce natural-sounding speech.
of DMs in the context of typical utterances. In addition such a dictionary could visualize the corresponding acoustic properties, what is important for research purposes.

Our experience in describing Russian DMs within the proposed dictionary format immediately showed that the attention to DM phonetic characteristics, besides its evident necessity for a full picture of its linguistic behavior, also helps to achieve deeper understanding of complex interrelations between semantic and formal surface properties of these polyfunctional words. Here we confine ourselves to just one characteristic example - DM voobshche that has been an object of fine semantic analysis in [4] and [12]. In [7] and [13] we have two attempts to interrelate semantic and prosodic variation of this DM which gave incompatible results.

It should be noted that prosodic metalanguages of the two papers differ significantly. Ju. D. Apresian uses an integral parameter of the type of phrasal accent that has 4 values: syntagmatic, main, contrastive and emphatic, while S. V. Kodzasov operates in terms of more elementary and concrete acoustic parameters. This circumstance makes the comparison more difficult but still not impossible. Analyzing the two accounts we found out the disagreement in treatment of examples of the type (4), subsumed under the meaning 3 of the synopsis given in (1):

(4) Ona poobshchala zaxodit’ k nemu i voobshche zabotit’sja o nem.

"She promised to visit him and on the whole take care of him"

According to Apresian in cases like (4) voobshche can either have syntagmatic accent (that in a syntagm-initial position should be the rising one) or have no phrasal accent at all. According to Kodzasov in such cases as (4) voobshche always bears rising phrasal accent of big amplitude as shown in (4’):

(4’) Ona poobshchala zaxodit’ k nemu i ‘voobshche’(‘con’) ‘zabotit’sja’(‘con’) o nem.

Even without instrumental acoustic analysis it is easy to establish that the pronunciation of (4) without any phrasal accent on voobshche sounds strange and unnatural. After analyzing instrumentally samples of recorded material containing standard variants of pronunciation of the DM in question we discovered that the accent in (4’) shows only one of the two possibilities of accentuating voobshche in context of the type (4). The second possibility is the falling contrastive accent on voobshche, as shown in (4’):

(4”) Ona poobshchala zaxodit’ k nemu i ’voobshche’(‘con’) ‘zabotit’sja’(‘con’) o nem.

It is worth noting that the second prosodic possibility is excluded for voobshche in contexts like (5) that formally are very close to (4) and are treated in [7] as identical to (4) in terms of prosody:

(5) a. Sam on pil dorogoe vino, no voobshche(‘/’)/ | projavljal strashnuju skarednost’.

b. "Sam on pil dorogoe vino, no voobshche(‘con’) projavljal strashnuju skarednost’.

This difference in prosodic form of voobshche in the two types of context that has been thus discovered testifies against the unification of the two corresponding semantic variants of this DM in one lexeme that has one semantic definition with alternating components, as was proposed in [7]:

(6) P i voobshche <no voobshche> Q = ‘P has place and Q has place and P is a special case of Q’

On the contrary our prosodic data confirm the purely semantic analysis of voobshche given in [4] according to which the use of this DM in contexts like (4) is treated as an exponent of the semantic scenario “From particular to general”, its use in contexts like (5) is associated with different scenario - “Rules and exceptions”.

So the detailed account of the different types of information on DM made possible by computer technology is needed not only for its storage in the convenient form of the electronic dictionary but as an instrument of further investigation of this complex phenomenon.

References