

**Maria Khokhlova**  
(Saint-Petersburg State University)

**The Russian language of the XVIII century: presentation of texts written by  
M.V. Lomonosov**

The Russian language of the XVIII century is heterogeneous one both in lexical and grammatical aspects. There was no academic grammar until 1757, when M.V. Lomonosov, the great Russian scientist and literary man, published his famous “Russian grammar”, which became the first reference book. It was the first attempt to fix standards of usage, to abolish obsolete word forms and syntactic constructions. It isn't surprising that the idea of the compiling of an explanatory dictionary which would include all M.V. Lomonosov's works is rather old (the first attempt dates back to 1863). Such a dictionary would contribute to the comprehension which elements of the language of M.V. Lomonosov and his epoch reflect modern language and which ones are obsolete (for instance, the usage of archaic forms connected with the Church Slavonic language due to the long history of Russian). Hence we could trace changes in the vocabulary of the Russian language over several centuries.

The aim of the paper is to demonstrate linguistic phenomena typical for the XVIIIth century language and particularly for the individual language of M.V. Lomonosov. Lomonosov's tragedy “Tamira and Selim” was chosen as the starting point for the work. It is common knowledge that poetic texts do not conform to grammatical rules in order to keep rhyme.

On the first stage of the work the text was tagged by the “Dialing” tagger. Its morphological dictionary is created on the basis of A. Zaliznyak's “Russian Grammar Dictionary” that represents the Russian grammatical system of the second half of the XXth century. When working with texts of previous centuries a scholar has to face a number of phenomena which are not inherent in the present morphological system. The “Dialing” tagger has a prediction algorithm for unknown words which searches in the dictionary for the most conterminous word form to an entrance word. When tagging, a “predictor” assigns to an unknown token such a set of possible tags which is determined by the similar dictionary entry. Nevertheless on the basis of the 18-19th century vocabulary this algorithm appears to be not so efficient (about 70%). In the output the processed text is stored in XML-file (see an example below):

```
<w> Не
      <ana lemma="не" pos="ЧАСТ" gram="|" />
</w>
<w> должно
      <ana lemma="должен" pos="П" gram="|но,од,кр,ср,ед," />
      <ana_del lemma="должно" pos="ПРЕДК" gram="|нст," />
      <ana_del lemma="должный" pos="П" gram="|но,од,кр,ср,ед," />
</w>
<w> царское
      <ana lemma="царский" pos="П" gram="|но,од,ср,вн,ед," />
      <ana lemma="царский" pos="П" gram="|но,од,ср,им,ед," />
</w>
<w> Царю
      <ana lemma="царь" pos="С" gram="|од,|мр,дт,ед," />
      <ana_del lemma="царить" pos="Г" gram="|нп,нс,|дст,1л,нст,ед," />
</w>
<w> быть
      <ana lemma="быть" pos="ИНФИНИТИВ" gram="|нп,нс,|дст," />
</w>
<w> слово
      <ana lemma="слово" pos="С" gram="|но,|ср,им,ед," />
      <ana_del lemma="слово" pos="С" gram="|но,|ср,вн,ед," />
</w>
```

```
<w> ложно
      <ana lemma="ложный" pos="П" gram="кач,|но,од,кр,ср,ед," />
      <ana_del lemma="ложно" pos="Н" gram="|" />
</w>
```

On the second stage the procedure of manual morphological disambiguation takes place. At the same time several lists are compiled. They contain information about:

- morphological ambiguity;
- “predicted” lexemes.

On this basis we form list of cases in which the program failed to attribute proper lexemes and grammatical characteristics to a token such as gender, case, and number for nouns; aspect, voice, and tense for verbs and so on. This register will be added to the dictionary of the tagger for its improving.

The following examples illustrate such “predicted” lexemes:

1) Селим противу стен стоит наперед (Selim is standing opposite the walls).

Incorrect:

```
<w> противу
      <ana lemma="против?" pos="С" gram="|мр,дт,ед," />
</w>
```

Correct:

```
<w> противу
      <ana lemma="противу" pos="ПРЕДЛ" gram="|" />
</w>
```

The word “противу” is a preposition and not a noun as it was attributed.

2) по повелению отца своего перешед через Натолию ([he] having crossed Natolia by his father’s command)

Incorrect:

```
<w> перешед
      <ana lemma="перешед?" pos="С"
      gram="|ср,жр,мр,пр,тв,вн,дт,рд,им,ед,мн," />
</w>
```

Correct:

```
<w> перешед
      <ana lemma="перейти" pos="ДЕЕПРИЧАСТИЕ" gram="пе,св,|дст,прш," />
</w>
```

The word “перешед” is an adverbial participle and not a noun as it was attributed

There are no such words in present Russian, but they existed in Lomonosov’s idiolect.

The cases when the “predictor” fails to assign an appropriate lexeme could be subdivided into following groups:

- 1) names and place-names: Клеона, Нарсим, Натолия etc;
- 2) adjectives with endings -аго, -ья, -ия: высокия, гордаго, котораго, общаго, светлыя etc;
- 3) particles written jointly with principal words: емуль, ежель, естлиб, ктоже, лутчели etc;
- 4) other cases.

The instances of morphological ambiguity which are of our main interest add up to ones of coinciding of verb forms or verboids with forms of nouns or pronoun adjectives:

Селим

```
<ana lemma="селить" pos="ПРИЧАСТИЕ" gram="пе,нс,|стр,но,од,нст,кр,мр,ед," />
<ana_del lemma="селить" pos="Г" gram="пе,нс,|дст,1л,нст,мн," />
```

строй

```
<ana lemma="строй" pos="С" gram="но,|мр,вн,ед," />  
<ana lemma="строй" pos="С" gram="но,|мр,им,ед," />  
<ana_del lemma="строить" pos="Г" gram="пе,нс,|дст,пвл,2л,ед," />
```

Царя

```
<ana lemma="царь" pos="С" gram="од,|мр,вн,ед," />  
<ana lemma="царь" pos="С" gram="од,|мр,рд,ед," />  
<ana lemma="царить" pos="ДЕЕПРИЧАСТИЕ" gram="нп,нс,|дст,нст," />
```

день

```
<ana lemma="день" pos="С" gram="но,|мр,им,ед," />  
<ana_del lemma="день" pos="С" gram="но,|мр,вн,ед," />  
<ana_del lemma="деть" pos="Г" gram="пе,св,|дст,пвл,2л,ед," />
```

сей

```
<ana lemma="сей" pos="МС-П" gram="|но,од,жр,пр,ед," />  
<ana_del lemma="сей" pos="МС-П" gram="|но,мр,вн,ед," />  
<ana_del lemma="сей" pos="МС-П" gram="|но,од,жр,дт,ед," />  
<ana_del lemma="сей" pos="МС-П" gram="|но,од,жр,рд,ед," />  
<ana_del lemma="сей" pos="МС-П" gram="|но,од,жр,тв,ед," />  
<ana_del lemma="сей" pos="МС-П" gram="|но,од,мр,им,ед," />  
<ana_del lemma="сеять" pos="Г" gram="пе,нс,|дст,пвл,2л,ед," />
```

мой

```
<ana lemma="мой" pos="МС-П" gram="|но,мр,вн,ед," />  
<ana lemma="мой" pos="МС-П" gram="|но,од,мр,им,ед," />  
<ana lemma="мыть" pos="Г" gram="пе,нс,|дст,пвл,2л,ед," />
```

чаю

```
<ana lemma="чай" pos="С" gram="но,|2,мр,пр,ед," />  
<ana lemma="чай" pos="С" gram="но,|2,мр,рд,ед," />  
<ana lemma="чай" pos="С" gram="но,|мр,дт,ед," />  
<ana lemma="чаять" pos="Г" gram="пе,нс,|дст,1л,нст,ед," />
```

The language of M.V. Lomonosov bears traces of German syntax and sometimes the text is difficult to interpret because of this. We mean to show these interrelations in the text using technical labels. It is also intended to introduce special tags to label phenomena lacking in “Russian Grammar Dictionary”, for instance obsolete words. Also we plan to work out system of tags according to Lomonosov’s “Russian Grammar”. This new approach will contribute to understanding the individual writer’s language which cannot be explained within the framework of the XXth century grammar.

On the next stage this information from XML-file is converted into MySQL database. And a user can search through it by two types of parameters. The former is one that was attributed by the tagger. The latter is the system of “Lomonosov’s” tags which are attributed manually. So the text will be interpreted from two different points of view. And comparing them a scholar will be able to observe language phenomena which are typical for the XVIIIth century texts, and also correlations between items.

We hope that this quite a new approach will be of benefit for investigations in the field of historical linguistics, and will help to trace the development of Russian grammatical system.