

Anita Bartulović · Linda Mijić · Max Silberztein
Editors

Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities

17th International Conference, NooJ 2023
Zadar, Croatia, May 31 – June 2, 2023
Revised Selected Papers

Editors

Anita Bartulović
University of Zadar
Zadar, Croatia

Linda Mijić
University of Zadar
Zadar, Croatia

Max Silberztein 
Université de Franche-Comté
Besançon, France

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-56645-5 ISBN 978-3-031-56646-2 (eBook)
<https://doi.org/10.1007/978-3-031-56646-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2024

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Contents

Morphological and Lexical Resources

SANTI-Network Prototype of an Indonesian Multi-level Tagger	3
<i>Prihantoro</i>	
Deciphering the Nomenclature of Chemical Compounds in NooJ	16
<i>Kristina Kocijan, Krešimir Šojat, and Tomislav Portada</i>	
NooJ Dictionary for Rromani: Importing of an Editorial Dictionary to the NooJ System	28
<i>Masako Watabe</i>	
Latin Pronouns, Numerals and Prepositions in the NooJ Tool	39
<i>Anita Bartulović and Linda Mijić</i>	
Towards a Linguistic Annotation of Arabic Legal Texts: A Multilingual Electronic Dictionary for Arabic	51
<i>Khadija Ait ElFqih, Maria Pia Di Buono, and Johanna Monti</i>	
Recognition of Frozen Expressions in Belarusian NooJ Module	64
<i>Yauheniya Zianouka, David Latyshevich, Mikita Suprunchuk, and Yuras Hetsevich</i>	

Syntactic and Semantic Resources

A Proposal for the Processing of the Nucleus Verb Phrase of Pronominal (SVNPr) Verbs in Spanish	79
<i>Andrea Fernanda Rodrigo, Rodolfo Bonino, and Silvia Susana Reyes</i>	
German <i>selbst</i> -Compounds: Towards a NooJ Grammar	90
<i>Marco Angster</i>	
Disambiguation Grammars for the Ukrainian Module	102
<i>Olena Saint-Joanis</i>	
Automatic Disambiguation of the Belarusian-Russian Legal Parallel Corpus in NooJ	114
<i>Valery Varanovich, Mikita Suprunchuk, Yauheniya Zianouka, and Yuras Hetsevich</i>	

Automatic Disambiguation of the Belarusian-Russian Legal Parallel Corpus in NooJ

Valery Varanovich¹✉, Mikita Suprunchuk², Yauheniya Zianouka³,
and Yuras Hetsevich³

¹ Belarusian State University, Minsk, Belarus
ssrlab221@gmail.com

² Minsk State Linguistic University, Minsk, Belarus

³ United Institute of Informatics Problems of the National Academy of Sciences of Belarus,
Minsk, Belarus

Abstract. The article deals with the problem of lexical ambiguity in the legislative texts of the Belarusian legal codes. Specialists of the Speech Synthesis and Recognition Laboratory in Minsk translate law codes from Russian into Belarusian to ensure equal access to them for speakers of both official languages of our country. First, the texts are translated automatically and checked and edited by translators, linguists, and lawyers. Over 20 of the 26 codes have already been prepared. The article examines types of homonyms, their semi-automatic extraction and classification using the NooJ development environment. There are two dominant classes of ambiguity resolution mechanisms: automatic, implying a completely computerized solution to this problem and interactive (semi-automatic), supposing a joint solution by a person and a computer. It means the user has a set of alternatives from which he should choose one option. About 17% of words in the experimental corpus have homonyms. The article examines their types, and presents several syntactic NooJ grammars, which allow for removing different types of homonymy.

Keywords: Ambiguity · Automatic Translation · Disambiguation · Homonymy · Natural Language Processing · NooJ · Text Corpus · Text Tagging

1 Introduction

Homonymy (and ambiguity) is still the central problem of automatic text processing at the lexical level. The problem of the automatic (rarely semi-automatic) solving of lexical ambiguity was first formulated for creating Machine Translation systems. To date, this is a critical problem for improving the quality of systems for various branches of computational linguistics.

There are two dominant classes of ambiguity resolution mechanisms: (1) an automatic one, implying a fully computerized solution to this problem; (2) an interactive (dialogic, semi-automatic) one, supposing a joint solution by a person and a computer [1]. It means the user has a set of alternatives from which they should choose one option.

We are creating a legal text corpus in Belarusian and Russian for various software products, such as speech synthesis, machine translation, spell checkers, etc. As part of the project, we developed a trilingual dictionary of legal terms in Belarusian, Russian and English. Also, in forming the corpus, we compiled contextual dictionaries, which became the basis for dictionaries in the NooJ system. The dictionaries reflect contexts that show the preferred translation of a given term from Russian into Belarusian [2].

Our plan includes the creation of high-priority dictionaries for each of the 26 law codes in Belarus. We assume that many diagnostic contexts will be consistent across different codes (настоящий 'this' = гэты 'this' (code), not настоящий 'this' = сапраўдны 'real' (code)). Still, we hypothesize that in some cases, the contexts for the same values will be different. It is also possible that one polysemous word (or a homonym) will have different meanings in several codes.

The following examples of words with multiple meanings were identified during our work: дaнный – 1) гэты, 'this'; 2) дадзены, 'given'; отпуск – 1) выдача (тавараў), 'issuance', 'supply'; 2) адпачынак (даць), 'holiday', 'vacation'. These meanings (translations) can be selected based on their neighboring words. Given that NooJ is an effective tool for resolving word ambiguity, we intend to utilize it for constructing syntactic grammars [3]. They will search for homonyms by analyzing the context (the sequence of words) and form a list of lexical units for different domains. This will aid in identifying terms within various thematic domains, a crucial step in compiling specialized vocabularies for the mentioned fields.

2 Linguistic Analysis of the Text Corpus

To handle the question of translating legislative documents into Belarusian, the Speech Synthesis and Recognition Laboratory of UIIP NASB (<https://ssrlab.by/en/>), in cooperation with specialists from Belarusian State University, has translated all 26 codes of the Republic of Belarus (*The Civil Code, The Labor Code, The Education Code, The Banks and Banking Code*, etc.) into the Belarusian using the automatic services of the website *corpus.by*. We constructed a corpus of texts on medical and social domains, and integrated it into NooJ to extend the research results to other fields of knowledge. Belarusian code links, the translation of each legal code into Belarusian and the dictionary of Regular substitutions are posted here: <https://ssrlab.by/7804>.

To determine the types of homonymy, we performed an experimental corpus of legal texts (Fig. 1). It consists of 5 first pages from 23 codes. The collection contains 4,572 text units (TUs), 52,463 tokens, including 40,647 word forms, 1,796 digits and 10,020 delimiters. The text comprises 95,184 annotations (11,891 different).

Thus, standard tools (the basic Belarusian module for NooJ) analyze legislative texts quite well, but not all forms receive an unambiguous grammatical characteristic. We suggest several ways to improve the results of automatic disambiguation.

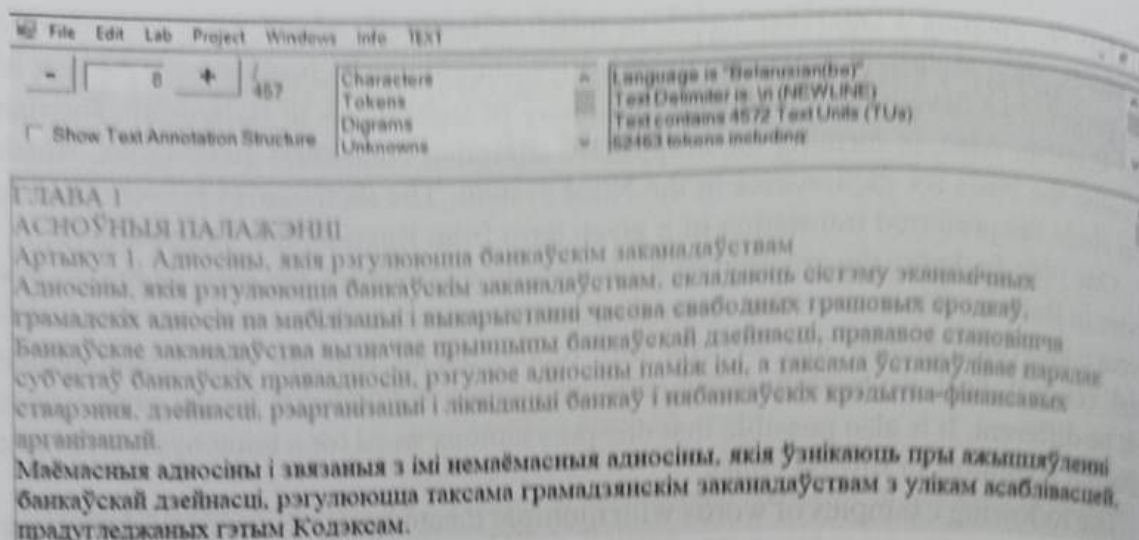


Fig. 1. A fragment of the corpus of the legal domain.

3 Linguistic Analysis of the Text Corpus: Homonymy Type Determination

In the following research stage, we exported the list of ambiguities computed by the NooJ linguistic analyzer. It comprises 3,172 examples of homonyms (Fig. 2).

We classified the selected homonyms with a frequency of ten or more. Sometimes, one-word form belongs to different types of homonymy. Such a case was marked as, for example, 6+1 гэты 'this' (Fig. 3).

For example, ambiguous forms of the word асобы, which can be a plural noun in the genitive case, a single noun in the genitive (a person) or a single adjective in the nominative or accusative case (special) according to the Belarusian NooJ dictionary:

```
<асобы, ADJECTIVE+FLX=AA3ICHЫ+Case=Accusative+Animation=Inanimate+Gender=Masculine+Meaning=Qualitative_Relative>
<асобы, ADJECTIVE+FLX=AA3ICHЫ+Gender=Masculine+Case=Nominative+Meaning=Qualitative_Relative>
<асоба, NOUN+FLX=АБАЛОНА+Meaning=Common+Gender=Feminine+Case=Genitive+Animation=Inanimate>
<асоба, NOUN+FLX=АБАЛОНА+Case=Accusative+Meaning=Common+Gender=Feminine+Animation=Inanimate+Number=Plural>
<асоба, NOUN+FLX=АБАЛОНА+Meaning=Common+Gender=Feminine+Animation=Inanimate+Case=Nominative+Number=Plural>
<асоба, NOUN+FLX=АБАТЫСА+Animation=Animate+Meaning=Common+Gender=Feminine+Case=Genitive>
<асоба, NOUN+FLX=АБАТЫСА+Animation=Animate+Meaning=Common+Gender=Feminine+Case=Nominative+Number=Plural>
```

Homonymy is widespread in Belarusian texts: almost every token has several variants of grammatical annotation (Fig. 3). Other examples of homonyms are presented in Fig. 4.

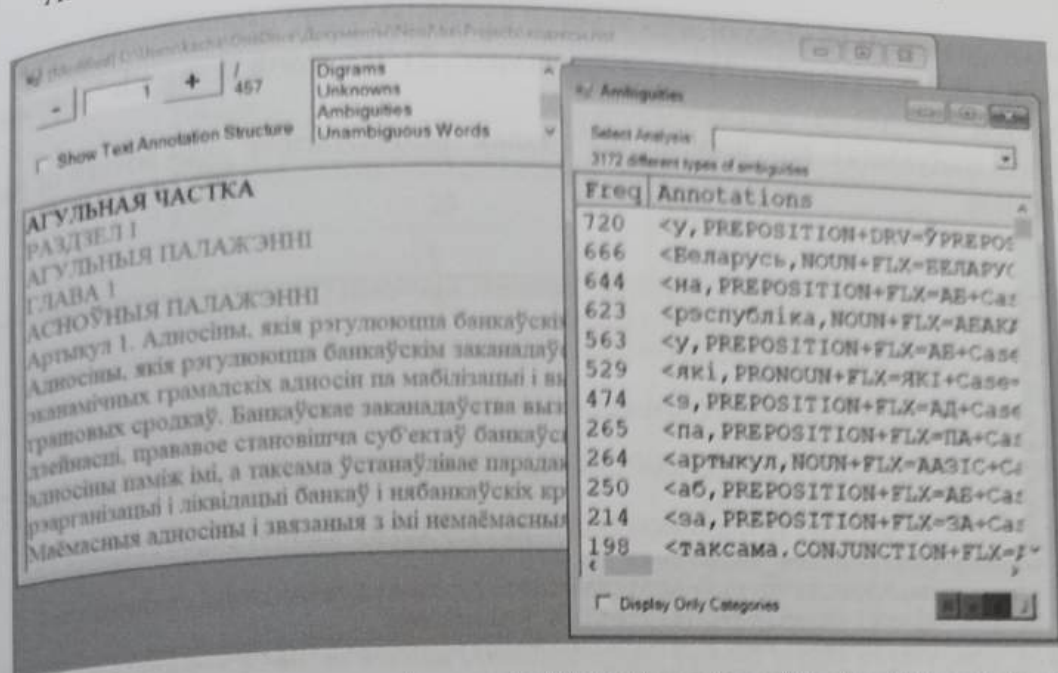


Fig. 2. The list of ambiguous words exported from the corpus of the legal domain in NooJ.

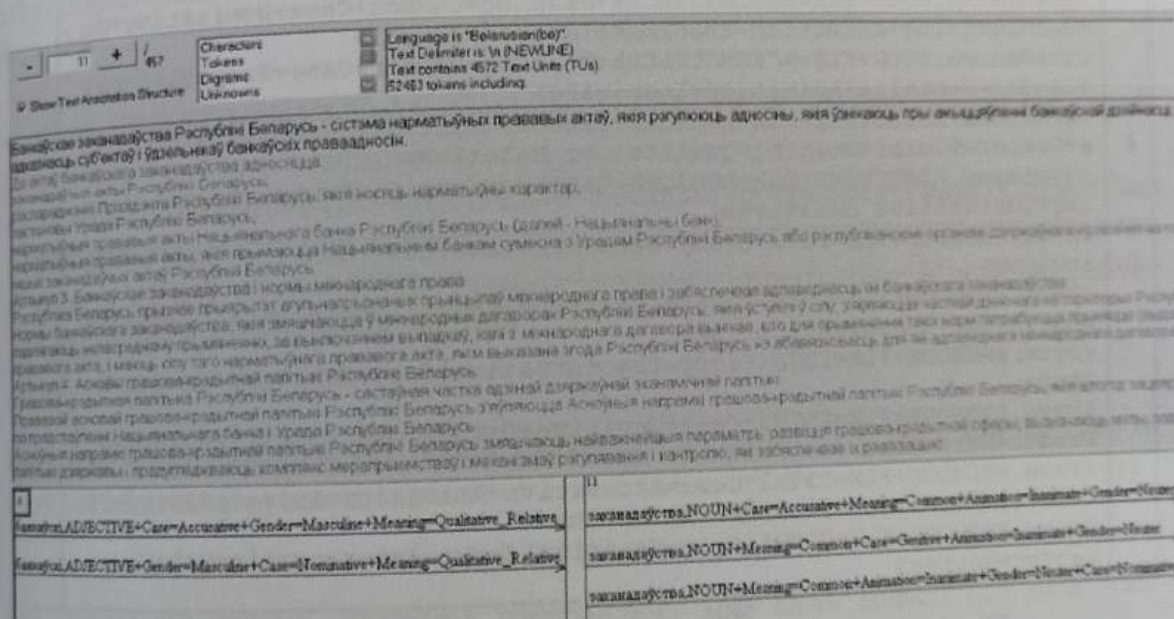


Fig. 3. Examples of ambiguous word annotations in NooJ.

Within the research, we identified the following types of homonyms:

- grammatical (coincidence in case only or case and number): Беларусь 'Belarus' (nom. case = acc. case);
- grammatical (coincidence in animacy: animate or inanimate): асоба 'a person' (animate) or 'a legal person' (inanimate);
- preposition (different case management): у, на 'in, on', etc.;
- partial homonyms (coincidence in parts of speech): рабочы – 1) 'employee, worker', 2) 'working'; таксама – 'also', 'too';

- e. lexico-grammatical (homoforms): які – 1) 'which', 2) 'yak (animal)'; дзкікі – 1) 'wild', 2) 'boar'; варты – 1) 'security, service', 2) 'it's worth doing sth.'; правы – 1) 'not left', 2) 'liberty' or 'permission';
- f. lexical (polysemy): яго 'him' or 'his'; далей 'farther/further' – an adverb of time or direction.

7	<права, NOUN+FLX=ПРАВА+Meaning=Common+Case=Genitive+Animation=Inanimate+Gender=Neuter+Number=Plural>	5
9	<праўіць, VERB+FLX=БАВІЦЬ+Person=2+Mood=Imperative+Aspect=Imperfective+Number=Singular>	
	<праці, VERB+FLX=КЛАСЦІ+Aspect=Imperfective+Mood=Indicative+Gender=Masculine+Tense=Past>	
7	<зямельны, ADJECTIVE+FLX=ААЗІЧНЫ+Case=Genitive+Number=Plural+Meaning=Qualitative_Relative>	1
5	<зямельны, ADJECTIVE+FLX=ААЗІЧНЫ+Number=Plural+Case=Prepositional+Meaning=Qualitative_Relative>	
	<зямельны, ADJECTIVE+FLX=ААЗІЧНЫ+Case=Accusative+Animation=Animate+Number=Plural+Meaning=Qualitative_Relative>	
7	<дзеясць, NOUN+FLX=АСКОМІТАЦЬ+Meaning=Common+Case=Dative+Gender=Feminine+Animation=Inanimate>	1
4	<дзеясць, NOUN+FLX=АСКОМІТАЦЬ+Meaning=Common+Case=Genitive+Gender=Feminine+Animation=Inanimate>	
	<дзеясць, NOUN+FLX=АСКОМІТАЦЬ+Meaning=Common+Case=Nominative+Gender=Feminine+Animation=Inanimate+Number=Plural>	
	<дзеясць, NOUN+FLX=АСКОМІТАЦЬ+Meaning=Common+Case=Prepositional+Gender=Feminine+Animation=Inanimate>	
7	<зямельны, ADJECTIVE+FLX=ААЗІЧНЫ+Case=Accusative+Animation=Inanimate+Number=Plural+Meaning=Qualitative_Relative>	1
4	<зямельны, ADJECTIVE+FLX=ААЗІЧНЫ+Case=Nominative+Number=Plural+Meaning=Qualitative_Relative>	
7	<гэты, PRONOUN+FLX=ГЭТЫ+Group=Demonstrative+Case=Genitive+Gender=Masculine>	6
1	<гэты, PRONOUN+FLX=ГЭТЫ+Case=Accusative+Animation=Animate+Group=Demonstrative+Gender=Masculine>	1
	<гэты, PRONOUN+FLX=ГЭТЫ+Group=Demonstrative+Case=Genitive+Gender=Neuter>	
	<гэты, PRONOUN+FLX=ГЭТЫ2+Case=Genitive+Group=InterrogativeRelative+Gender=Masculine>	
	<гэты, PRONOUN+FLX=ГЭТЫ2+Case=Accusative+Animation=Animate+Group=InterrogativeRelative+Gender=Masculine>	
	<гэты, PRONOUN+FLX=ГЭТЫ2+Case=Genitive+Group=InterrogativeRelative+Gender=Neuter>	

Fig. 4. The fragment of ambiguous homonyms of different types.

Table 1 shows which type of homonym is the most frequent among the word forms of the text (2 and 3 columns), as well as statistics on unique forms (4 and 5 columns) (cf. [4], as well). Grammatical homonymy is highly frequent and covers the maximum number of different lexemes.

4 Syntactic Grammars for Removing Grammatical Ambiguity

4.1 A Subsection Sample

It is known that the Slavic languages have well-developed inflection. The endings of verbs indicate gender, number, person, time, etc. For example, Belarusian and Russian nouns and adjectives have declension, i.e., six cases and two numbers. Due to this

Table 1. Statistical data on different types of ambiguity.

Type	Tokens		Other words (= "types")	
	Units	%	Units	%
1	2303	25	423	78
2	803	9	41	8
3	3548	37	20	4
4	702	8	19	3
5	1429	15	32	6
6	536	6	9	1
All	9321	100	544	100

paradigmatic abundance, some words coincide in all forms, while others only coincide in specific forms. In other words, it is observed that they are not complete homonyms but rather homoforms. The following examples with the word *мае* 1) 'has'; 2) 'my; mine' show this:

- **<мой, PRONOUN+FLX=МОЙ+Case=Accusative+Animation=Inanimate+Number=Plural+Group=Possessive>**
- **<мой, PRONOUN+FLX=МОЙ+Case=Nominative+Number=Plural+Group=Possessive>**
- **<мець, VERB+FLX=МЕЦЬ+Person=3+Aspect=Imperfective+Mood=Indicative+Tense=Present+Number=Singular>**

Банк **мае** права ажыццяўляць іншыя банкаўскія аперацыі. 'The bank **has** the right to carry out other banking operations.'

(1)

(...) кожны выбаршчык, удзельнік рэферэндуму **мае** адзін голас. '(...) each voter participant in the referendum **has** one vote.'

(2)

We have implemented a syntactic grammar to remove this kind of ambiguity and assigned the correct verb tags (Fig. 5). All 48 ambiguous cases of the word *мае* were cleared out with this grammar.

The case when several forms of a word coincide within the same paradigm is more common. Consider the example of a noun's genitive and dative cases that coincide. Since the forms of the dative and the genitive cases of the words *бюджэт* 'budget' and *фонд* 'fund' are the same (G. sg. = D. sg. = *бюджэту*, G. sg. = D. sg. = *фонду*), then the automatic annotation, of course, attributed to them two tags of these cases.

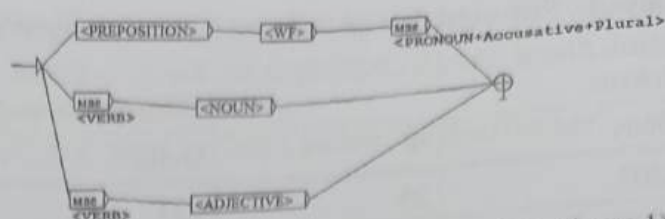


Fig. 5. Syntactic grammar for removing ambiguous forms of word має 1) 'has'; 2) 'my; mine'.

Бюджетны кредит – бюджетныя сродкі, якія даюцца з вышэйстаячага бюджэту ніжэйстаячаму бюджэту, а таксама з рэспубліканскага бюджэту, бюджэту дзяржаўнага пазабюджэтнага фонду або з бюджэту дзяржаўнага пазабюджэтнага фонду рэспубліканскаму бюджэту на зваротнай аснове. (3)

'Budget credit (budget loan) – budget funds provided from the higher budget to the lower budget, as well as from the republican budget to the budget of the state extrabudgetary fund or from the budget of the state extrabudgetary fund to the republican budget on a reciprocal basis'.

Fig. 6. Syntactic grammar for extracting masculine nouns of dative and genitive cases.

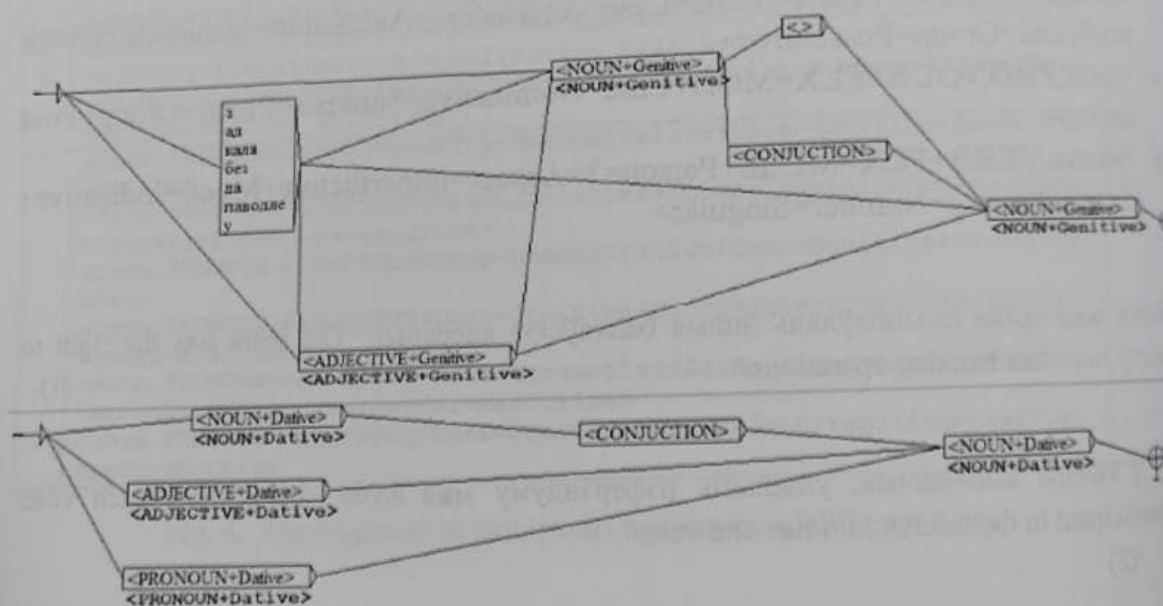


Fig. 6. Syntactic grammar for extracting masculine nouns of dative and genitive cases.

To remove this type of homonymy, we developed two grammars (Fig. 6). They analyze the left and the right contexts and allow users to define the genitive and dative cases, respectively.

4.2 Identification of Partial Homonyms

One more group of homonyms is partial homonyms, i.e., the same word is associated with multiple parts of speech categories. Parts of speech homonymous to each other

have the same spelling; that is, they represent one homocomplex (term). This type of homonym may be extracted by analyzing left and right contexts.

We identified ten groups of partial homonymy, which are:

- a. conjunction – particle (таксама, калі ‘also’, ‘when’),
- b. conjunction – adverb – verb (imperative) (калі, калоць ‘when to split’),
- c. conjunction – particle (ці, толькі ‘if’, ‘only’),
- d. adverb – preposition (пасля ‘after’),
- e. short adjective – adverb (юрыдычна, спецыяльна, пастаянна ‘legal, special, permanent’ vs. ‘legally, specifically, permanently’),
- f. noun – verb (справа, спрасці, справіць ‘law case’, ‘to span, to fulfill’),
- g. parenthesis – verb (можа, магчы ‘maybe’ vs ‘can’),
- h. adjective – participle (населены ‘populated’, ‘with population’),
- i. verb (gerund) – preposition (уключаючы ‘switching’ vs. ‘including’),
- j. noun – adjective (правы ‘the right’, ‘right’).

The group of short adjectives and adverbs exhibited the most frequent type of partial homonymy. Let us consider the coincidence between the short feminine adjective ‘legal’ and the adverb ‘legally’ that have the same graphics in Belarusian (юрыдычна). For example,

- <юрыдычны, ADJECTIVE+FLX=ААЗІЧНЫ+Gender=Feminine+Case=Nominative+Meaning=Qualitative_Relative+Form=Short>

Банк – юрыдычная асоба, якая мае выключнае права ажыццяўляць у сукупнасці наступныя банкаўскія аперацы (...). ‘A bank is a legal entity that has the exclusive right to carry out the following banking operations (...).’ (4)

- <юрыдычна, ADVERB+FLX=АБАВЯЗАЦЕЛЬНА+Type=Quality_Manner>

Судовыя дакументы ва ўстаноўленым парадку ўручаюцца юрыдычна зацікаўленым у выходзе справы асобам на мове судаводства. ‘Court documents are duly handed over to persons legally interested in the outcome of the case in the language of the proceedings’. (5)

The wordform *спраў* coincides with the plural noun in the genitive case (*спраў* ‘law cases’) and singular verbs of the past tense (‘to span, to fulfill’). For the legal texts, the most typical usage is a plural noun in the genitive case, but still, the following syntactic grammar helps to find out all cases of the word *спраў* in all texts of different styles (Fig. 7).

Determining the coincidence of auxiliary parts of speech is more challenging in this case, as it requires consideration of the underlying meaning in the sentence. We tried to distinguish conjunction and particle *таксама* ‘also’, ‘too’. To identify whether *таксама* is a conjunction or a particle, we used the following syntactic grammar (Fig. 8).

Another example of partial homonymy is the coincidence of the word *калі* (conjunction/adverb ‘when’; imperative ‘split, stab’). For legal texts, an adverb and a conjunction

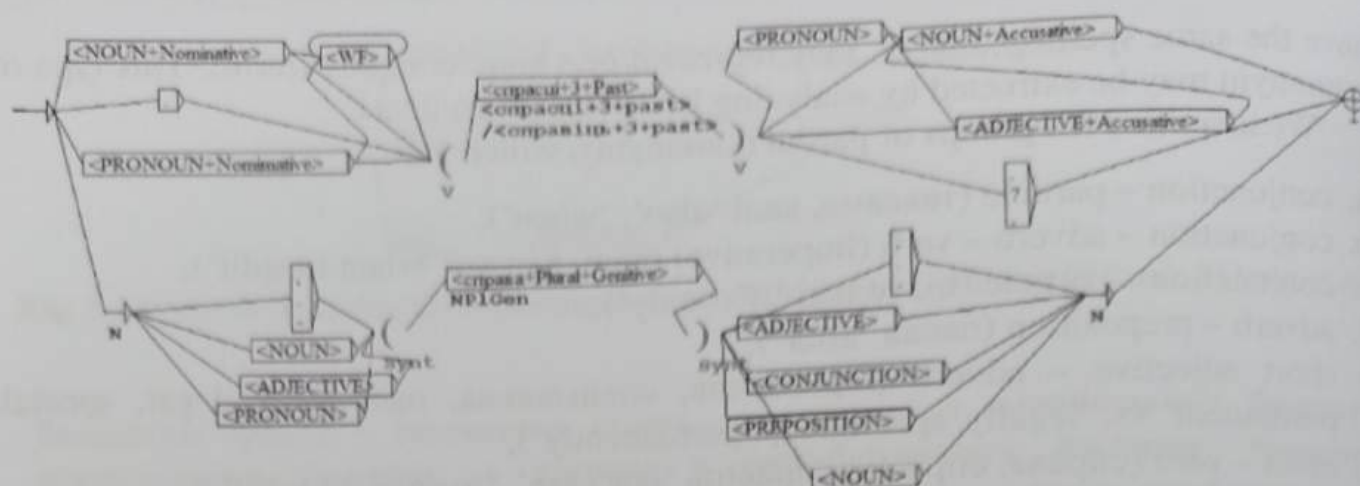


Fig. 7. Syntactic grammar for extracting the wordform cripaŭ in NooJ.

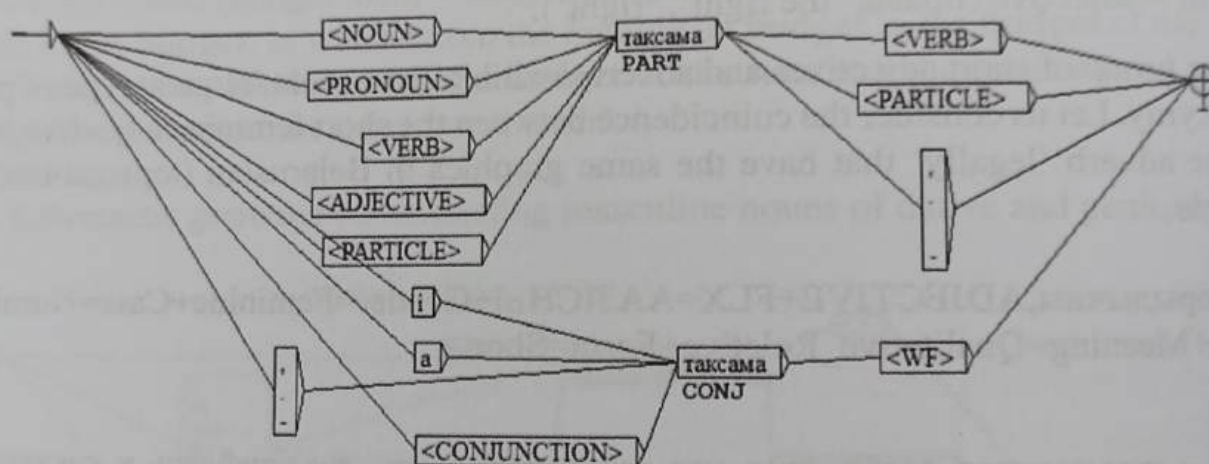


Fig. 8. Syntactic grammar for extracting conjunction or particle takcama 'also', 'too'.

are more typical. At the same time, the imperative form of the verb is used relatively rarely and primarily in literary texts (Fig. 9).

4.3 Identification of Lexical Homonyms

Lexico-grammatical homonyms belong to different parts of speech. They have different lexical meanings, and may also have differences in grammatical categories: які 'which; yak', дзкі 'wild; boar', варта 'security service; it's worth doing sth.', правы 'not left; liberty, permission'.

The word form *правы* has the meanings of a noun and an adjective in different contexts. Linguistic analysis has shown that the noun is most often found in legal texts, which is natural since the word is a part of the legal vocabulary. Note that this pair of homonyms also have different accents, so this grammar can be used in text-to-speech synthesizers for correct voicing, depending on the context. For example,

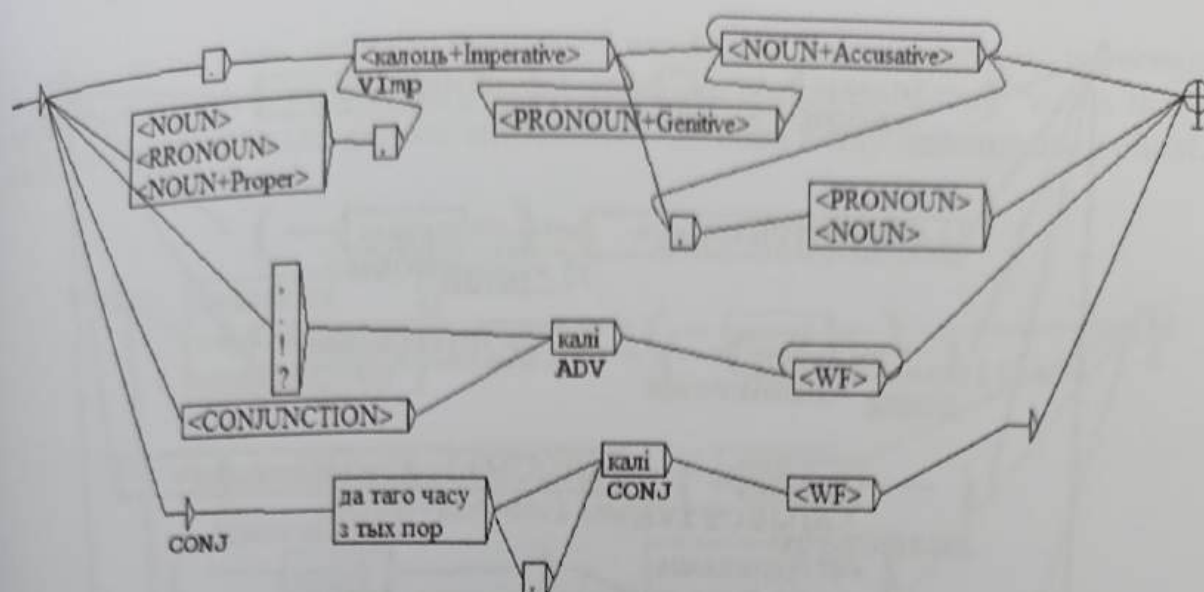


Fig. 9. Syntactic grammar for extracting калі (conjunction, adverb or imperative).

- <права, NOUN+FLX=ППАВА+Case=Accusative+Meaning=Common+Animation
=Inanimate+Gender=Neuter+Number=Plural>

Шлюбны дагавор, які прадугледжвае **правы** і абавязкі былых сужонкаў пасля спынення шлюбу, дзейнічае да іх выканання. 'The marriage contract, which provides for the rights and obligations of the former spouses after the termination of the marriage, is valid until their fulfillment.'

- <правы, ADJECTIVE+FLX=ААЗІЧНЫ+Gender=Masculine+Case=Nominative+
Meaning=Qualitative_Relative>

Кіраўскі мост праз Дзвіну знаходзіцца ў цэнтры Віцебска, злучае вуліцы Кірава (**правы** бераг) і Замкавую (левы бераг). 'The Kirov Bridge over the Dvina River is located in the center of Vitebsk, connecting the streets of Kirov (**right** bank) and Zamkavaya (left bank).'

The developed grammar helps to determine the part-of-speech homonyms of the word *правы*, as well as its grammatical characteristics (Fig. 10).

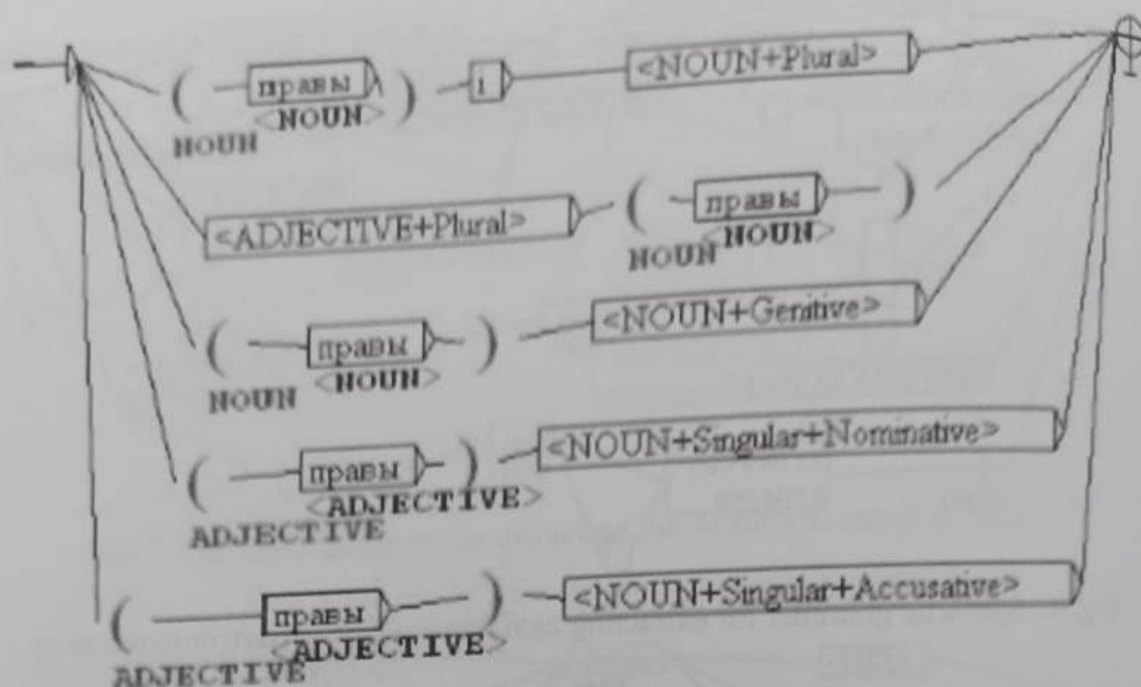


Fig. 10. Syntactic grammar for extracting part-of-speech homonym правы 'not left'; 'liberty', 'permission'.

In the Belarusian language, there is a case when a word form has a different lexical meaning but entirely coincides with a particular grammatical form. Thus, the words акт 'act' (masculine gender) and акт 'octave' (feminine gender) coincide in the plural form of the genitive case, and also have different accents. Therefore, the grammar developed for this case also helps the correct voicing of the word form by the text-to-speech synthesizer (Fig. 11). For example,

- <акт, NOUN+FLX=AAMIЦЭT+Meaning=Common+Case=Genitive+Animation=Inanimate+Gender=Masculine+Number=Plural>

Пацвярджэнне адпаведнасці транспартных работ патрабаванням тэхнічных нарматыўных прававых **актаў** у галіне тэхнічнага нарміравання. 'Confirmation of compliance of transport works with the requirements of technical regulatory **legal acts** in the field of technical regulation.'

- Noun+Feminine:

<актава, NOUN+FLX=АБЛАВА+Meaning=Common+Gender=Feminine+Case=Genitive+Animation=Inanimate+Number=Plural>

Белыя клавiшы (...) заўсёды імкнуліся правільна перадаць гучанне асноўных тонаў музычных **актаў**. 'White keys (...) always tried to correctly convey the sound of the main tones of musical **octaves**.'

One can only use semantic features of words that form the context to recognize homonyms of this type. The developed grammar illustrates specific words that can indicate the semantic and generic affiliation of the noun *актаў* in the genitive plural.

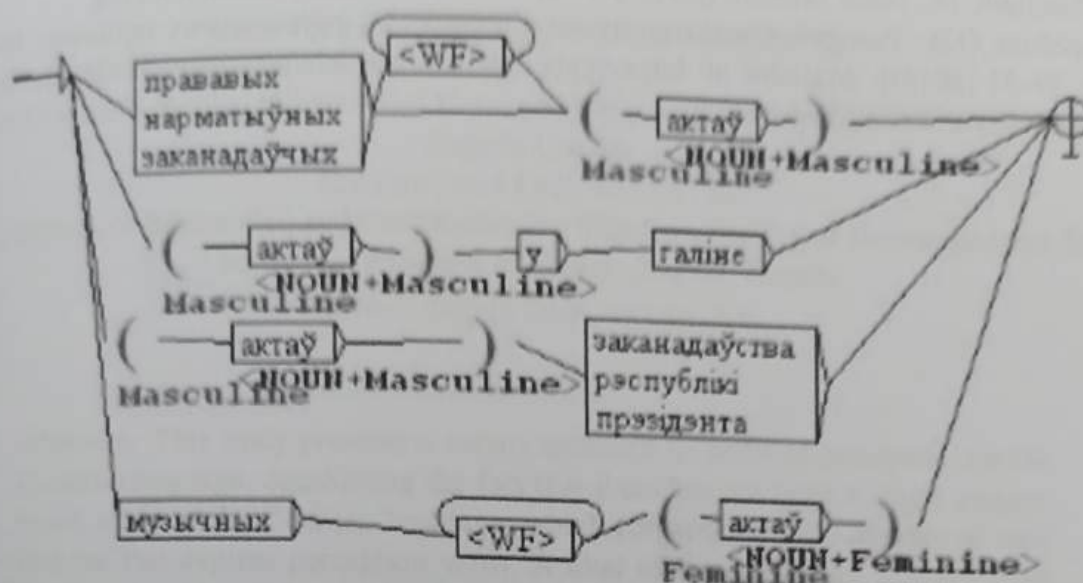


Fig. 11. Syntactic grammar for extracting plural noun *актаў* in the genitive case.

5 Conclusion

Specialists of the Speech Synthesis and Recognition Laboratory of UIIP of NASB translate legal codes from Russian into Belarusian to ensure equal access for speakers of both official languages. First, the texts are translated automatically and checked and edited by translators, linguists and lawyers. More than 20 of the 26 codes have already been prepared (more details are available at <https://ssrlab.by/en/7804>). We used the NooJ system for optimization of work and research purposes. It incorporates tools that help in solving various types of ambiguity (homonymy and polysemy) in Belarusian legal texts: lexical, grammatical, part-of-speech homonymy and mixed homonymy. We prepared an experimental legislative corpus of some fragments for the study of homonymy in the language of legal texts. Ambiguities of various types are prevalent in Belarusian legal texts: 9,000 tokens of 52,500 tokens. The most common types are grammatical rather than lexical: homonymy of prepositions (government of different cases) and homoforms (coincidence of cases of the same word). Several NooJ grammars for reducing ambiguity are presented above. They make it possible to remove homonymy, to reduce the number of incorrect definitions of grammatical or lexical meaning, to distinguish a noun from a verb, an adjective from an adverb, etc. Our findings will be employed in the automatic linguistic analysis of Belarusian texts, machine translation and speech synthesis.

References

1. Bouarroudj, W., Boufaïda, Z., Bellatreche, L.: Named Entity Disambiguation in Short Texts over Knowledge Graphs. *Knowl. Inf. Syst.* **64**(2), 325–351 (2022)

2. Воронович, В. В.: Словарь лексических валентно-стей в системе русско-белорусского машинного перевода. In: А. И. Головня (ed.) Третьи чтения, посвященные памяти В. А. Карпова: сб. науч. ст., pp. 88–92. РИВШ, Минск (2009). Available at: <http://elib.bsu.by/handle/123456789/10645>
3. Silberztein, M.: NooJ Manual. (2003–). Available at: www.nooj-association.org
4. Барабаш, О.В.: Разграничение омонимии и полисемии юридических терминов. Рема 2, pp. 39–51 (2015)]. Available at: <https://cyberleninka.ru/article/n/razgranichenie-omonimii-i-polisemii-yuridicheskikh-terminov>