

Digital Preservation of Rhodope Dialects Through Intelligent Language Technologies

Radomira Lazarova¹, Veneta Tabakova-Komsalova¹[0000-0002-0617-9844],
Elena Kanevska-Nikolova¹[0000-0001-5899-654X], Stanimir Stoyanov¹[0000-0002-3854-4260]

¹ University of Plovdiv Paisii Hilendarski, 24 Tsar Assen Street, 4000 Plovdiv, Bulgaria
radomira.lazarova@uni-plovdiv.bg, v.komsalova@uni-plovdiv.bg,
elenakanik@uni-plovdiv.bg, stani@uni-plovdiv.bg

Abstract. The article presents a prototype intelligent language modelling system to preserve Rhodope dialects. Implemented in Prolog, it converts dialectal into standard Bulgarian lexical units and vice versa, generating the corresponding grammatical form. It includes a module for phonetic features and demonstrates the application of symbolic AI in digital cultural heritage.

Keywords: Rhodope Dialects, Digitalization, Prolog Logic Programming, Cultural Heritage.

1 Introduction

Dialects are an essential part of a country's cultural and linguistic heritage (Chambers & Trudgill, 1998; Labov, 2006). They not only reflect linguistic diversity, but also provide valuable information about historical, geographical, and cultural processes. They can be used to trace the characteristics of everyday life, migration movements, and the development of society in different regions. Studying and preserving dialects is key to understanding cultural identity. This perspective is central in modern dialectology and sociolinguistics (Chambers & Trudgill, 1998). Bulgaria is characterised by a diverse and valuable dialectal heritage, which has been the subject of systematic scholarly research and documentation (Institute for Bulgarian Language – BAS, n.d.; Prokić et al., 2009; Osenova et al., 2010; Alexander et al., 2016). The present study focuses exclusively on the Rhodope dialects, which have been preserved in a relatively authentic form and exhibit a range of interesting linguistic phenomena. The Rhodope dialects are divided into six subdialects (microdialects): Smolyanski, Shirokolushki, Hvoynenski, Chepinski, Pavlikyanski, and Zlatogradski (Stoykov, 2002). Each of them has certain specific features and characteristics, some of which are quite isolated and homogeneous, such as the Shirokolushki dialect, while others, such as the Chepinski dialect, show dialectal variations even between neighbouring villages. All these factors reflect regional specificity and long-standing historical heritage, playing a significant role in shaping the cultural identity of the communities. Despite the existence of substantial linguistic research and accumulated lexical material, the level of digitization and public

accessibility of the respective dialects remains insufficient, which in turn creates conditions for the loss of their linguistic authenticity. Similar challenges in documenting and digitizing linguistic diversity have been identified in other regions as well (Ryzhikova, 2025). The subject of our research is the digitization of the Rhodope dialects in the form of an intelligent interactive system that models their grammatical and phonetic features and differentiates the specifics between the microdialects. The prototype of the system, presented in the article is based on a combination of dictionary data and procedural rules and is fully implemented in the logic programming language Prolog, which is extremely suitable for language processing. Another advantage is that, unlike large language models, it requires few computational resources. Prolog has traditionally been applied in natural language processing due to its suitability for rule-based linguistic modelling (Pereira, et al., 1987; Nugues, 2014; Bitter et al., 2010).

Cultural and creative industries play a key role in preserving linguistic and cultural diversity by supporting artistic expression, local traditions, and intangible heritage (Kazashka et al., 2017). Through cultural activities, as well as through the integration of digital technologies and educational approaches, they help to maintain regional identity and pass on linguistic heritage to future generations.

2 Exposition of the Investigation

The presented prototype is under active development and currently incorporates core mechanisms for morphological and phonetic processing. These components will be discussed in the present article. The limitations of the model are due to the fact that, at present, it does not fully cover all phonetic and morphological features, and their extension and refinement are part of ongoing work. To create the prototype and processing rules, the morphological and phonetic features of the Rhodope dialects were first analysed. This is a highly complex process, because although they share a common basis, the individual microdialects have specific features, some of which are not homogeneous, and there may be local differences even within the same microdialect (Chepinski, Smolyanski). The architecture of the system (Fig. 1) consists of three levels: a distributed knowledge base, management, and a user interface. The user interface is simple and suitable for users without in-depth knowledge of dialectology and linguistics and is currently under development and in interactive dialogue mode. Management is carried out by a server and a translator program. The server processes user requests and manages the interaction between the different levels of the system and the morphological processing modules. Currently, the system operates locally, using the Prolog language for both management and language modelling due to its effectiveness in formalizing grammatical dependencies and extracting linguistic information. Rule-based and symbolic approaches remain particularly effective for modelling morphologically rich languages (Sag & Wasow, 2000; Reyle & Rohrer, 1988).

The knowledge base is differentiated by category according to the type of lexical units and contains separate modules for nouns, adjectives, verbs, uninflecting particles, and idiomatic expressions. A `dictionary_suffixes` module has also been added,

which contains suffixes and other morphemes for word formation, and additional separate modules can be added if necessary.

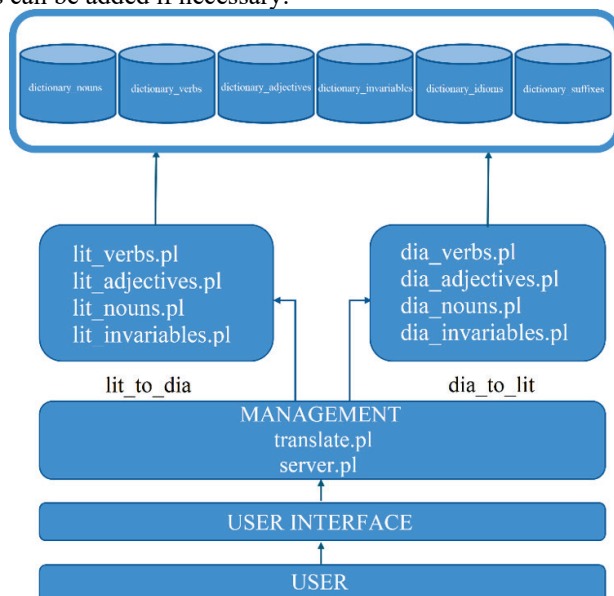


Fig. 1. Prototype architecture.

Each basic lexical unit in the knowledge base matches two sets of lexical units – one is a set of a given standard Bulgarian word plus its corresponding synonyms, the other is the set of synonyms of the dialectal equivalent of the word. This avoids repetition of lexical units and improves the technical maintenance of the database, as well as making it very easy to add more categories if necessary. The structure of the dictionary entries in the categories "nouns", "adjectives" and "verbs" also allows for the possibility of more than one meaning, with priority given to dialectal vocabulary.

The morphological processing mechanisms work in two steps: recognizing a valid word form when user input is entered, and finding a match and constructing a word form that grammatically corresponds to the entered word. This mechanism works both in the direction from standard words to dialectal words and from dialectal to standard words, with an analogy in both directions, but not a complete correspondence due to some differences in the morphological processing of dialectal and standard units. This required the processing programs to be separated into two modules – `lit_to_dia`¹ – containing programs for processing in the direction of the entered standard word form – derived dialectal correspondence, and `dia_to_lit` – for the reverse process. Computational modelling of inflectional morphology is a central topic in contemporary NLP research (Lieber, 2009; Neef, 1998).

The modules for processing the different lexical categories are: `lit_verbs.pl`; `lit_adjectives.pl`; `lit_nouns.pl`; `lit_invariables.pl` and, similarly,

¹ In variable names, "lit" denotes the standard form of Bulgarian language, not literary works.

dia_verbs.pl; dia_adjectives.pl; dia_nouns.pl; dia_invariables.pl, with the selection of the relevant processing modules being made according to user preferences.

Nouns – lexical units are contained in the `dictionary_nouns` module and are processed by the `lit_nouns/dia_nouns` programs, with each unit having the following structure: `- dictionary_noun([mean(N, [lit(LiteratureNoun, LiteratureGender, LiteraturePluralForm)]), [region(Region, [dia(DialectNoun, DialectGender, DialectPluralForm, Stress)])])`. As in standard Bulgarian, nouns in its dialectal variations change according to number and definiteness. The lexeme in the singular, indefinite form, is accepted as the basic dictionary form, but due to the diversity of plural formation in Bulgarian and the existence of some irregular forms, another basic form has been added to the dictionary unit-predicate: plural, indefinite form. The other possible cases of word formation are singular, definite form, and plural, definite form. For standard nouns, the process of recognition and, accordingly, of word form construction involves checking whether the entered word is in its basic dictionary form, and if so, simply displaying its meaning. If this step fails, lemmatization is performed – the ending is separated and it is checked whether the resulting lemma matches a valid dictionary form, as well as whether the ending is a valid ending specified in the `lit_noun_ending/3` inflection predicates. With regard to dialectal forms, the mechanisms become more complicated, since in the Rhodope dialects, with very few exceptions, the phenomenon of "three-way definite article system" is observed – when inflecting a word, we may have a definite form with the morpheme indicator “-H”, “-C” or “-T”. Three definite article forms are used in specific contexts, reflecting a three-way definite article system with distinct proximate and distal forms. Even when deriving the meaning of individual words, they must be taken into account as an important feature of the dialect.

Such complex inflectional systems pose challenges for computational representation, particularly in dialectal contexts (Dunn, 2019; Zampieri & Nakov, 2021). Initially, the formation of dialectal definite forms in the system was modelled analogously to those in Standard Bulgarian—by adding a suffix to the lemma; however, at a later stage, as the modelling was refined in greater detail, numerous changes were introduced. What is special here is that the endings for declension depend not only on gender and number, but also on the sound at the end of the base form of the word, as well as whether the vowel before the morpheme indicator is stressed or not. The morpheme vowel itself also varies depending on the microdialect and region to which the word belongs (Kanevska-Nikolova, 2006). This variability leads to numerous combinations and, accordingly, requires differences in the mechanisms for constructing dialectal word forms, therefore, unlike in standard Bulgarian, where the ending is simply added to the lemma, in dialectal words the ending is constructed dynamically by selecting the vowel corresponding to the region and adding it to the inflectional morpheme.

Adjectives – dictionary entries are contained in the `dictionary_adjectives` module and are processed by the `lit_adjectives/dia_adjectives` programs, with each entry having the following structure: `dictionary_adjective([mean(N, [lit('LiteratureAdjective', Type)]), [region(Region, [dia('DialectAdjective', Type)])])`. Here, the variability of word

forms increases, as adjectives change according to gender, number, and definiteness, with some of them losing the penultimate vowel when forming word forms. A similar analogy can be observed in dialectal adjectives, where there is another category of adjectives that do not change, so that in this category too, the processing mechanisms do not fully overlap with those in the standard language.

Verbs – dictionary entries are contained in the `dictionary_verbs` module and are processed by the `lit_verbs/dia_verbs` programs, with each entry having the following structure: `dictionary_verb([mean(N , [lit(LiteratureVerb, VerbEnding, LiteratureConjugation, Reflexive, Impersonal)])], [region(Region, [dia(DialectRoot, DialectEnding, DialectConjugation, Reflexive, Impersonal)])])` – the arguments `Reflexive` and `Impersonal` indicate whether the verb is reflexive or not, and whether it is impersonal or not, respectively, with possible values being `yes/no`.

Verbs, both in the standard Bulgarian language and in the Rhodope dialects, are categorized by conjugation and change according to person, number, and tense. Modelling verbal inflection computationally has been widely explored in morphological NLP research (Wiemerslage et al., 2018). Initially, only the present tense was developed in the system, with the verb root as the basis. Since there is variability in the endings of first and second conjugation verbs when forming the first person singular and third person plural, the ending for the first person was added as an argument in the predicate, and the other endings were added to the `dictionary_suffixes` module. In order to extend the initial version with more verb tenses, a more thorough restructuring and refinement of the morphology was necessary. The `dictionary_suffixes` module was supplemented by adding predicates for verb endings for the past perfect tense, past imperfect tense, and other verb forms.

Uninflecting parts of speech – contained in the `dictionary_invar` module, each unit has the following structure: `dictionary_invar(LiteratureWord, [region(Region, DialectWord)], Type)` – this module contains adverbs, prepositions, conjunctions, particles, and interjections, with pronouns added for simplicity, as most of them do not change according to morphological rules and the addition of particles and suffixes. The structure of the predicate is simplified as much as possible, but a third argument, `Type`, is added to determine the type of uninflecting part. `Type` can have the following values: `adverbs`, `preposition`, `conjunction`, `particle`, `interjection`, `pronoun`, `possessive`, `universal`, with the last two referring to possessive pronouns and general pronouns, which differ from the others in that they can take definite forms.

Phonetic phenomena and transformations – unlike the standard language, where there are generally accepted rules for spelling and pronunciation of words, in dialects it is very important to take into account phonetic features in writing. This is important, on the one hand, in order to preserve information about their correct pronunciation, but in addition, phonetic phenomena and transformations also convey information about belonging to a particular microdialect, as well as being associated with certain local regional features. The interaction between phonology and morphology is central in computational phonology (Bird, 2002). Not all such features are universally valid for all Rhodope dialects, nor do they apply entirely to a specific category of lexical units. In certain cases, there is a transformation of vowels when certain words are inflected, a

transition of consonants, or variations in pronunciation. All these specifics necessitate in-depth research and precise modelling of the phonetic phenomena in the Rhodope dialects, which is presented in a separate module for processing `phonetics.pl`, which is still under development and is periodically supplemented. A subset of these predicates is illustrated in Figure 2. In the Rhodope dialect, stress in certain nouns shifts to the definite morpheme upon definiteness marking, which is reflected in the lexical predicate through the argument `Stress`. In the example, `stress_transition/2` predicates encode stressed vowels along with their corresponding alternants in forms with mobile stress, while the `stress_removal/2` and `char_replace/4` predicates identify the stressed vowel in a given word and substitute it with its corresponding unstressed counterpart when the `Stress` argument of the word has the value `mobile`.

The following examples provide a detailed illustration of the functional capabilities of the developed system for processing and bidirectional conversion between standard Bulgarian and the Rhodope dialect. Specific input data are used to demonstrate the lexical, morphological, and word-formation mechanisms implemented within the system.

The input words are entered in the lower text field, while the extracted meaning is displayed in the upper text field, with the output words shown in bold. The system can return meanings for both a single input word and multiple words separated by commas.

Figure 3 shows an example of automatic extraction of the meaning of idiomatic expressions. In this case, the system recognizes the fixed phrase as a whole semantic unit, rather than a mechanical collection of individual lexemes, and provides its corresponding interpretation. This ensures the correct interpretation of phraseological constructions whose semantics cannot be derived through direct lexical analysis of their constituent words.

Figure 4 demonstrates the processing when a word from the Rhodope dialect is entered. The system performs morphological identification of the input form and outputs the dialectal meaning of the corresponding standard noun – both in its indefinite and definite forms (including forms with a triple article). In addition, a phonetic transformation of the vowel is performed, which shows that the algorithm takes into account the characteristic phonetic features of the dialect when generating counterparts.

Figure 5 shows the output of the standard meaning of a dialect noun. Both indefinite and definite forms are presented, which indicates a complete morphological correspondence between the dialect and standard Bulgarian paradigms. In this way, the system ensures the correct correlation of the grammatical categories of gender, number, and definiteness.

Figure 6 illustrates the derivation of the standard meaning of a dialectal adjective. Here, the correspondence between the dialectal form and its standard equivalents is traced with regard to the categories of gender, number, and definiteness, when relevant.

Figure 7 shows the derivation of the dialectal meaning of a standard adjective. In the example considered, the dialectal form is uninflecting, so it does not change regardless of the grammatical forms of the standard equivalent entered. This demonstrates the system's ability to take into account structural differences between the two language systems and apply the appropriate transformation rules.

```

vowel(a). vowel(ä). vowel(o). vowel(y). vowel(e). vowel(и).

stress_transition(ä,ä).
stress_transition(ä,о).
stress_transition(ä,а).
stress_transition(е,е).
stress_transition(и,и).

stress_removal(Atom, NewAtom) :-
    stress_transition(OldChar, NewChar),
    atom_chars(Atom, CharList),
    nth1(Position, CharList, OldChar),
    char_replace(Position, CharList, NewChar, NewCharList),
    atom_chars(NewAtom, NewCharList).

char_replace(Position, CharList, Char, Result) :-
    nth1(Position, CharList, _, Rest),
    nth1(Position, Result, Char, Rest).

```

Fig. 2. Subset of phonetic transformation predicates.

Речник

Преживял е най-мъчното

Пририпнал йе арапкцото

Изпрати

☐ Книжовен - Дialeкт ☒ Дialeкт - книжовен

Fig. 3. Deriving the meaning of idiomatic expressions.

Речник

мѐж, мѐжѐт, мѐжѐс, мѐжѐн

мѐж, мѐжѐт

Изпрати

☒ Книжовен - Дialeкт ☐ Дialeкт - книжовен

Fig. 4. Deriving the dialectal meaning of a standard noun – indefinite and definite forms (three-way definite article system), with phonetic transformation of a vowel.

Речник

боровинка, боровинки, боровинките

брусница, брусници, брусниците

Изпрати

☐ Книжовен - Дialeкт ☒ Дialeкт - книжовен

Fig. 5. Deriving the standard meaning of a dialectal noun – indefinite and definite forms.

Речник

тъжен, тъжна, тъжно

балан, бална, бално

Изпрати

☐ Книжовен - Дialeкт ☒ Дialeкт - книжовен

Fig. 6. Deriving the standard meaning of a dialectal adjective.

Речник

кайл

съгласен, съгласна

Изпрати

☒ Книжовен - Дialeкт ☐ Дialeкт - книжовен

Fig. 7. Deriving the dialectal meaning of a standard adjective; in this case, the dialectal form is uninflecting and does not change, regardless of the forms of the standard Bulgarian language that are introduced.

Речник

мърморя, мърмориш, мърмори, мърморим,
мърморите, мърморят

папунйсѣм, папунйсѣш, папунйсѣ,
папунйсѣме, папунйсѣте, папунйсѣт

Изпрати

☐ Книжовен - Диалект ☒ Диалект - книжовен

Fig. 8. Deriving the standard meaning of a dialectal verb in all persons of the singular and plural.

Речник

пригрибам, пригрибаш, пригрибѣ,
пригрибаме, пригрибате, пригрибѣт

прегрѣшам, погрѣшаш, погрѣша,
погрѣшаме, погрѣшате, погрѣшат

Изпрати

☒ Книжовен - Диалект ☐ Диалект - книжовен

Fig. 9. Deriving the dialectal meaning of a standard verb in all persons of the singular and plural.

Figure 8 shows the derivation of the standard meaning of a dialectal verb in all persons of the singular and plural. The system generates the complete verb paradigm by matching the dialectal forms with the corresponding standard equivalents, taking into account the grammatical categories of person and number.

Figure 9 shows the reverse process – deriving the dialectal meaning of a standard verb in all persons of the singular and plural. In this case, too, a complete paradigmatic transformation is performed, based on predefined morphological and phonetic rules.

In addition to processing individual lexemes, the system also supports the analysis of entire sentences. When a sentence in the Rhodope dialect is entered, a semantic translation into standard Bulgarian is generated and a morphological analysis of each individual word is performed. The analysis includes determining the part of speech and the corresponding grammatical characteristics. When a sentence in standard Bulgarian is entered, the system performs a morphological and lexical analysis of each word, after which it generates the corresponding dialect forms. In addition, the dialect group to which the correspondences belong is noted. The generation of dialect forms is carried out through reverse construction, analogous to the analysis of the standard Bulgarian form, which ensures structural symmetry and consistency in the two-way transformation.

3 Theoretical Context

As can be seen from the examples, basic morphological processing mechanisms for the main parts of speech, as well as some of the phonetic features, are currently modelled in the system. Given the need for precise and authentic reproduction of both the Bulgarian standard language and the Rhodope dialects, it can be expected that the system is undergoing continuous transformation and will continue to develop. In accordance with the linguistic features, the formulated rules include many special cases and exceptions, which had to be described in the most flexible and formalized way possible. Pho-

netic differences and variability within microdialects pose a particularly serious challenge, and the development and refinement of phonetic rules remains one of the main priorities of the system at present.

The integration of computational methods in dialect research reflects broader developments in statistical and rule-based NLP (Manning & Schütze, H., 1999; Jurafsky & Martin, 2026). Recent work in computational dialectology further demonstrates the potential of digital tools for analysing syntactic and morphological variation (Dunn, 2019; Zampieri & Nakov, 2021).

4 Conclusions

The present study demonstrates that intelligent language technologies can play a significant role in the digital preservation of dialectal heritage, with a specific focus on the Rhodope dialects. By developing a prototype system implemented in Prolog, we have shown that symbolic artificial intelligence and rule-based modelling are effective approaches for representing complex morphological and phonetic phenomena characteristic of these dialects. The system successfully recognizes lexical units, processes grammatical categories, and generates corresponding forms in both directions – from standard Bulgarian to dialectal variants and vice versa – while accounting for region-specific features and microdialectal variation.

The modular architecture of the prototype, including separate components for nouns, adjectives, verbs, uninflecting parts of speech, and phonetic transformations, ensures flexibility, scalability, and maintainability. Particular attention has been given to challenging linguistic features such as three-way definite article system, variable morphemes, and phonetic alternations, which are essential for preserving the authenticity of the Rhodope speech forms. Although the phonetic module and some extended verb forms are still under development, the current results confirm the feasibility of formalizing intricate dialectal structures within a computational framework.

Beyond its technical contribution, the system represents a step toward safeguarding intangible cultural heritage through accessible digital tools. By enabling structured storage, analysis, and interactive exploration of dialectal data, the prototype contributes to the long-term preservation, study, and transmission of regional linguistic identity. Future work will focus on expanding the lexical database, refining phonetic modelling, incorporating additional grammatical categories, and developing a fully functional user interface to enhance accessibility for researchers, educators, and the wider public. The combination of symbolic AI and linguistic theory continues to offer a transparent and resource-efficient alternative to purely data-driven approaches.

Acknowledgements

This publication was made possible through the financial support of the Scientific Research Fund of “Paisii Hilendarski” University of Plovdiv, within the project: "The Rhodopes and their people in an era of global challenges and transformations", contract FP25-FS-003.

References

- Alexander, R., Zhobov, V., & Kolev, G. (2016). *Bulgarian dialectology as living tradition*. Bulgarian Dialectology. <https://bulgariandialectology.org/>
- Bird, S. (2002). *Computational phonology* (2nd ed.). Oxford International Encyclopedia of Linguistics. <https://doi.org/10.48550/arXiv.cs/0204023>
- Bitter, C., Elizondo, D. A., & Yang, Y. (2010). Natural language processing: A Prolog perspective. *Artificial Intelligence Review*, 33, 151–173. <https://doi.org/10.1007/s10462-009-9151-4>
- Chambers, J. K., & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511805103>
- Dunn, J. (2019). Global syntactic variation in seven languages: Toward a computational dialectology. *Frontiers in Artificial Intelligence*, 2(15). <https://doi.org/10.3389/frai.2019.00015>
- Institute for Bulgarian Language, Bulgarian Academy of Sciences. (n.d.). *Interactive dialect map of the Bulgarian language*. <https://ibl.bas.bg/interactive-dialect-map-of-the-bulgarian-language/>
- Jurafsky, D., & Martin, J. H. (2026). *Speech and language processing* (3rd ed.). web.stanford.edu.
- Kanevska-Nikolova, E. (2006). *Triple articulation in the Rhodope dialects*. Paisii Hilendarski University Press. <https://www.rodopskistarini.com/2013/05/y.html>
- Kazashka, V., Ruseva, M., & Stoyanova, P. (2017). Mapping and ranging of cultural and creative industries in the city of Plovdiv. *CBU International Conference Proceedings*, 5, 639–642. <https://doi.org/10.12955/cbup.v5.999>
- Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511618208>
- Lieber, R. (2009). *Inflectional morphology: A theory of paradigm structure*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486333>
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press. <https://mitpress.mit.edu/9780262133609/foundations-of-statistical-natural-language-processing/>
- Neef, M. (1998). Spencer, Andrew/Zwicky, Arnold M. (Hrsg.): The handbook of morphology. *Zeitschrift für Sprachwissenschaft*, 17(2), 287–296. <https://doi.org/10.1515/zfs.1998.17.2.287>
- Nugues, P. M. (2014). *Language processing with Perl and Prolog: Theories, implementation, and application* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-642-41464-0>
- Osenova, P., Heeringa, W., & Nerbonne, J. (2010). A quantitative analysis of Bulgarian dialect pronunciation. *Zeitschrift für Slavische Philologie*, 66(2), 425–458.
- Pereira, F. C. N., & Shieber, S. M. (1987). *Prolog and natural-language analysis*. Semantic Scholar. <https://api.semanticscholar.org/CorpusID:264203475>
- Prokić, J., Nerbonne, J., Zhobov, V., Osenova, P., Simov, K., Zastrow, T., & Hinrichs, E. (2009). The computational analysis of Bulgarian dialect pronunciation. *Serdica Journal of Computing*, 3, 269–298.

- Reyle, U., & Rohrer, C. (1988). *Natural language parsing and linguistic theories*. Springer. <https://doi.org/10.1007/978-94-009-1337-0>
- Ryzhikova, T. R. (2025). Digitisation of linguistic research as a means of preserving linguistic diversity in Siberia (database of dialogical speech of minority ethnic groups in the Novosibirsk Region and neighbouring regions). *Vestnik NSU, Series: Linguistics and Intercultural Communication*, 22(3), 112–121. <https://doi.org/10.22204/2587-8956-2025-122-03-112-121>
- Sag, I. A., & Wasow, T. (2000). Syntactic theory: A formal introduction. *Computational Linguistics*, 26(2), 295. <https://doi.org/10.1162/coli.2000.26.2.295a>
- Stoykov, S. (2002). Balgarska dialektologiya [Bulgarian dialectology]. In *Dialektizmite v ezika na hudozhestvenata literatura* [Dialectisms in the language of fiction literature] (pp. 368–380). „Prof. Marin Drinov“.
- Wiemerslage, A., Silfverberg, M., & Hulden, M. (2018). Phonological features for morphological inflection. In *Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology* (pp. 161–166). Brussels, Belgium.
- Zampieri, M., & Nakov, P. (2021). Dialectology for computational linguists. In *Similar Languages, Varieties, and Dialects: A Computational Perspective* (pp. 96–118). Cambridge University Press. <https://www.cambridge.org/core/books/abs/similar-languages-varieties-and-dialects/dialectology-for-computational-linguists/C21B76D0E1DF4BBC12C28C9BB50CB1F3>

Received: March 13, 2026

Reviewed: May 04, 2026

Finally Accepted: May 15, 2026

