

¹ LexiRus: Word-Aligned Translation of Russian Learning Materials for International Students

³ Nataliya Kondrashova Vladimirovna¹ and Elijah Edward Masanga  ²

⁴ 1 ITMO University, Saint Petersburg, Russian Federation 2 Damotiva, Mbeya, Tanzania

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Open Journals](#) 

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright¹⁶ and release the work under a¹⁷ Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).¹⁸

⁵ Summary

⁶ Russia attracts numerous international students seeking advanced education in modern technologies and specialized academic programs. However, a significant barrier exists: most⁷ undergraduate courses and technical subjects are taught exclusively in Russian. To address⁸ this challenge, preparatory courses have been established to teach the Russian language to⁹ international students before they begin their main studies.

¹⁰ A critical problem in these programs is that Russian is often taught using Russian itself, despite¹¹ students arriving with diverse native backgrounds—including **Swahili, French, Spanish, English,**
Portuguese, Arabic, Chinese, and Umbundu. This necessitates constant manual translation by¹² students, significantly impeding their learning progress. Furthermore, the occasional adoption¹³ of English as a bridge language by instructors creates an “instructional asymmetry,” providing¹⁴ an unfair advantage to students with English proficiency while further marginalizing those from¹⁵ non-Anglophone backgrounds.

¹⁶ We present LexiRus, an open-source software system that provides word-level translation of¹⁷ Russian learning materials into students’ native languages. By enabling immediate vocabulary¹⁸ comprehension, preserving the visual structure of learning materials, and providing transparent¹⁹ sentence structure analysis, this tool accelerates Russian language acquisition and helps²⁰ international students transition more rapidly into their intended academic programs.²¹

²³ Statement of Need

²⁴ Language acquisition fundamentally requires mastering vocabulary (nouns and verbs) and²⁵ understanding syntactic patterns ([Krashen, 1982](#); [Nation, 2013](#)). International students in²⁶ Russian preparatory programs face a unique challenge: they must simultaneously decode²⁷ Russian instruction while learning the language itself. Traditional teaching methods that use²⁸ the target language to teach the target language create a “circular dependency” that extends²⁹ learning time and increases cognitive load.

³⁰ Existing translation tools typically operate at the sentence or paragraph level, often obscuring the³¹ word-to-word correspondence essential for language learning ([O’Shea et al., 2020](#); [Warschauer](#)
[& Healey, 1998](#)). Students need to understand not just the general meaning of a sentence,³² but how individual words combine to create meaning—a critical step in developing linguistic³³ competence.

³⁴ LexiRus addresses this gap by providing word-level translations that:

- ³⁵ **1. Preserve structural relationships:** Students see direct correspondences between Russian
and their native language syntax.
- ³⁶ **2. Enable pattern recognition:** Grammatical structures become visible through aligned
translations.

- 40 3. **Reduce cognitive friction:** Immediate vocabulary access eliminates the translation
- 41 bottleneck and reliance on external dictionaries.
- 42 4. **Accelerate learning:** Students progress faster when they can focus on language patterns
- 43 rather than tedious dictionary lookups.
- 44 5. **Ensure Linguistic Equity:** By supporting a wide range of native languages, the tool
- 45 removes the dependency on English as an intermediary bridge.

46 Implementation

47 LexiRus is built as a modular system designed to handle the complexities of Russian morphology
48 and PDF structure preservation.

49 Architecture

50 The system implements a modular architecture with the following core components:

- 51 1. **Document Parser:** Extracts text and layout metadata from various learning material
- 52 formats (primarily PDF).
- 53 2. **Tokenization Engine:** Segments Russian text into individual words while preserving
- 54 grammatical context using spaCy and NLTK (Bird et al., 2009; Honnibal et al., 2020).
- 55 3. **Translation Module:** Interfaces with high-performance translation APIs to provide
- 56 context-aware, word-level translations.
- 57 4. **Alignment System:** Maintains correspondence between source and target language words
- 58 across different syntactic structures.
- 59 5. **User Interface:** Presents aligned translations in an accessible, learner-friendly PDF
- 60 format.

61 Core Features

- 62 ■ **Word-level granularity:** Each Russian word is individually translated and aligned with
- 63 the original text.
- 64 ■ **Context preservation:** Grammatical relationships are maintained in the translation to
- 65 ensure accuracy.
- 66 ■ **Multiple language support:** Full support for **Swahili, Umbundu, French, Spanish, English,**
- 67 **Portuguese, Arabic, and Chinese.**
- 68 ■ **Interactive learning:** Users can comprehend new vocabulary while studying the original
- 69 layout of their course materials.
- 70 ■ **Export functionality:** Processed materials are saved in PDF format for high-fidelity offline
- 71 study.
- 72 ■ **Caching mechanism:** A **PostgreSQL** database stores previously translated words to
- 73 improve system performance and reduce latency.

74 Technical Stack

75 The implementation leverages a modern, robust technology stack:

- 76 ■ **Python 3.8+:** The core engine for text processing and logic.
- 77 ■ **NLP Libraries:** spaCy and NLTK for tokenization and linguistic analysis.
- 78 ■ **AI Backend:** Contextual word alignment and translation powered by the **Llama 3 Open**
- 79 **Source Model**, which utilizes the Transformer architecture (Vaswani et al., 2017).
- 80 ■ **Web Framework:** Django for backend management and user data handling.
- 81 ■ **Frontend:** NuxtJS for a responsive and intuitive user interface.

82 Installation and Usage

```
# Installation
pip install lexirus

# Basic usage
from lexirus import PDFTranslator

# Initialize translator for Swahili
translator = PDFTranslator(target_language='sw')

# Translate a document
translated_doc = translator.translate_document('learning_material.pdf')

# Save output
translated_doc.save('learning_material_sw.pdf')
```

83 Research Context

84 This work builds upon established principles in second language acquisition (Krashen, 1982)
85 and computer-assisted language learning (CALL) (Warschauer & Healey, 1998). The word-level
86 translation approach aligns with research showing that explicit vocabulary instruction combined
87 with contextual exposure produces optimal learning outcomes (Laufer, 2005; Nation, 2013).
88 Previous studies have demonstrated that learners benefit from seeing structural correspondences
89 between their native language and the target language (Jarvis & Pavlenko, 2008; Lado, 1957).
90 LexiRus operationalizes this insight by making these correspondences explicit and immediately
91 accessible. The approach is particularly relevant for morphologically rich languages like Russian,
92 where understanding word formation patterns and case systems is crucial (Timberlake, 2004).

93 Comparison with Existing Tools

94 Unlike general-purpose translation tools (such as Google Translate, DeepL, and Yandex.Translate),
95 our system provides unique pedagogical value:

- 96 ▪ **Word-Level Alignment:** Specifically designed for language learning and structural analysis
97 rather than just general communication.
- 98 ▪ **Grammatical Visibility:** Helps the learner “see” the underlying grammar instead of hiding
99 it behind a fluent, sentence-level translation.
- 100 ▪ **Direct Material Integration:** Specifically optimized for the PDF formats and technical
101 layouts used in academic settings.

102 Compared to existing language learning platforms like Duolingo, Babbel, or Rosetta Stone
103 (Vesselinov & Grego, 2012), LexiRus offers distinct advantages for university-level students:

- 104 ▪ **Authentic Materials:** Works with actual university textbooks and specialized technical
105 curricula rather than fixed, generic exercises.
- 106 ▪ **Curriculum-Agnostic:** Allows instructors to continue using their established curricula
107 while providing an automated layer of accessibility.
- 108 ▪ **Technical Focus:** Specifically optimized for the complex vocabulary challenges found in
109 Russian preparatory programs.

110 Future Work

111 Planned enhancements for the LexiRus ecosystem include:

- 112 1. **Audio Integration:** Implementation of pronunciation guides and text-to-speech (TTS)
113 for both the translated words and the original Russian terms to aid phonetic acquisition.
- 114 2. **Mobile Applications:** Development of native iOS and Android applications to facilitate
115 mobile learning and real-time translation during university lectures.
- 116 3. **Morphological Tagging:** Visualizing Russian case endings and verbal aspects through
117 color-coded highlights to further assist in grammar recognition.

118 Acknowledgments

119 We acknowledge ITMO University and Associate Professor Nataliya Kondrashova Vladimirovna.
120 Special thanks to Nataliya Kondrashova Vladimirovna for her pedagogical oversight and the
121 students who participated in the pilot program, providing the feedback necessary to refine the
122 system.

123 References

- 124 Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: Analyzing*
125 *text with the natural language toolkit*. O'Reilly Media, Inc. <https://www.nltk.org/book/>
- 126 Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength*
127 *natural language processing in python*. <https://doi.org/10.5281/zenodo.1212303>
- 128 Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.
129 <https://doi.org/10.4324/9780203935927>
- 130 Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon
131 Press. ISBN: 9780080286280
- 132 Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University
133 of Michigan Press.
- 134 Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook*,
135 5(1), 223–250. <https://doi.org/10.1075/eurosla.5.11lau>
- 136 Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge
137 University Press. <https://doi.org/10.1017/CBO9781139858656>
- 138 O'Shea, J., Crockett, K., Bandar, Z., & Buckley, K. (2020). Neural machine translation: A
139 review. *Journal of Artificial Intelligence Research*, 69, 347–418. <https://doi.org/10.1613/jair.1.12007>
- 140 Timberlake, A. (2004). *A reference grammar of russian*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511613951>
- 141 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., &
142 Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing*
143 *Systems*, 30. <https://arxiv.org/abs/1706.03762>
- 144 Vesselinov, R., & Grego, J. (2012). The duolingo effectiveness study. *City University of New*
145 *York, USA*. http://static.duolingo.com/s3/DuolingoReport_Final.pdf
- 146 Warschauer, M., & Healey, D. (1998). Computer-assisted language learning: An introduction.
147 *Language Teaching*, 31(2), 57–71. <https://doi.org/10.1017/S0261444800012970>