Gulsana Kenzhebek E-mail: g.kenzhebek17@gmail.com

GitHub

Skills

• **Programming Languages**: Python, R, SQL
• **Frameworks & Libraries**: spaCy, NLTK, pymorphy2, Pandas, NumPy, Matplotlib, Seaborn, tidyverse, ggplot2, scikit-learn (basic), Hugging Face (intro), Stanza
• **NLP Tools**: Rasa, Universal Dependencies, CoreNLP (basic), Gensim, Stanford NER
• **Statistical Methods**: t-test, ANOVA, regression, PCA, clustering, hypothesis testing
• **Technologies**: Git, Github, Docker, Linux, Jupyter Notebook, JupyterLab
• **Web Development:** HTML, CSS
• **Languages**: Kazakh (Native), Russian (Native), English (C1), French (Pre-Intermediate), German (A1)

Experience

• **National Research University Higher School of Economics** — Moscow, Russia
Student, Professional Retraining in Computational Linguistics | 02/2025 – Present

• **Rasa-English-Russian-chatbot · April 2025 / Rasa-chatbot-demo**
Developed a Rasa-based conversational agent to support language learning with a focus on practicing correct word order. Designed dialogue flows and intent classification to provide feedback and improve syntactic awareness in learners.
• **Parsing Difficulty Simulation & ANOVA (Python) · Mar 2025**
Simulated and analyzed sentence processing difficulty using Python and one-way ANOVA to compare reading times across syntactic structures (SVO, Passive, OSV). Applied statistical methods and data visualization to demonstrate the impact of syntax on comprehension in a cognitive science context.
• **Chekhov Text Analysis & Visualization (Python) · Mar 2025**
Processed and analyzed literary text using Python to extract linguistic features such as word frequency, bigrams, and TF-IDF scores. Generated word clouds and statistical summaries to visualize lexical patterns and syntactic tendencies, demonstrating methods of text interpretation in a computational linguistics context.
• **Chinese Text Processing & Word Frequency Analysis (Python) ·Feb-Mar 2025**
Analyzed Chinese text using Python, Jieba for word segmentation, and pandas to calculate word frequencies and bigrams. Generated insights into the text's structure and provided visualizations of word distributions, showcasing the use of natural language processing tools for text analysis in a cross-linguistic context.
Web Scraping and Analysis of Nutritional Terms in News Articles (Python) · Feb 2025
Created a Python web scraper to collect articles from news websites, focusing on nutritional terms like vitamins, minerals, and macronutrients. Used BeautifulSoup for scraping, spaCy for natural language processing, and Matplotlib to visualize how often these terms appeared in the articles. The project highlighted trends in nutritional topics in media and showcased the use of NLP and data visualization for content analysis.

Publications

◦ Published paper on frequency of word orders in Russian and English (using Python, spaCy, pymorphy2), — G.Kenzhebek. Young Scientist (moluch.ru), No. 14 (565), April 2025
◦ Built NLP pipeline for syntactic analysis, POS tagging, lemmatization using SpaCy and NLTK
◦ Applied statistical methods for word order distribution analysis (Pandas, Matplotlib, Seaborn)
◦ Researched linguistic aspects of word order focusing on topic-comment structure and variation

• **Cellular Automata Project, SFPC — March 2023**
Participant, School for Poetic Computation
◦ Developed Cellular Automata simulations using Python, experimenting with rule-based systems to model complex behaviors.
◦ Created interactive visualizations to explore dynamic systems and computational patterns.
◦ Collaborated with peers to design custom algorithms, blending creativity with computation.
◦ Applied critical thinking to assess biases in computational systems and understand their social implications.

• *Independent Contract Translator/Simultaneous interpreter* | 2017 – 2024
◦ Participated in multiple United Nations-related projects, providing critical support in areas such as research, data collection, and reporting.
◦ Enhanced proficiency in using translation management tools and CAT software.
◦ Contributor to GIS Projects. Continued providing freelance translation services, with a focus on high-level technical translation, including GIS-related materials for the United Nations projects.
◦ Collaborated with international teams, offering language support for multilingual projects, ensuring smooth communication across diverse stakeholders.

• *Simultaneous Interpreter and Translator*
◦ Austrian Pavilion, EXPO 2017 – Astana, Kazakhstan
◦ Delivered simultaneous interpretation for international visitors and delegations.
◦ Translated official written materials from English into Kazakh.

Education
• **National Research University Higher School of Economics** — Moscow, Russia
Professional Retraining in Computational Linguistics | 02/2025 – Present
• **Relevant Courses:** Python, Data Analysis, Lemmatization, Parsing, Corpus Linguistics, Text Models
• **The Eurasian Humanities Institute** — Astana, Kazakhstan
B.Ed. in Foreign Language: Two Foreign Languages: (English/French) | 2014 – 2017
GPA: 3.66 / 4.00 **(equivalent to 1.5 in the German grading system – *Sehr gut*)**
• **Thesis:** *Using Fairy Tales during English Lessons in Inclusive Schools*
• **Relevant Courses:** Introduction to Linguistics, Morphology, Syntax, Psycholinguistics, Cognitive Linguistics, Pedagogy, Child Development. Focus on linguistic theories related to language acquisition and communication disorders.

Certificates

Cognitive Science – The Human Brain
Massachusetts Institute of Technology (MIT OpenCourseWare)
Non-credit, self-paced course | Completed: March 2025

◦ Explored how the brain processes perception, language, memory, and attention, with a focus on specialized brain regions

◦ Developed a working understanding of cognitive neuroscience tools like EEG, fMRI, and lesion studies

◦ Independently completed 17 video lectures, empirical readings, and assessments