

Lingea: Integration of Language Tools into Workflows

Josef Jon, Isidora Katanić, Efraim Berkovič, and Marek Švanda

Lingea, Brno, Czechia

Abstract. Lingea s.r.o. is a longstanding language service provider and language technology developer from Brno, Czech Republic. This paper details practical applications of Lingea’s technologies across critical sectors—healthcare, global commerce, and law enforcement. It highlights key features such as state-of-the-art machine translation, enhanced search tools, predictive writing assistance, and specialized translation databases that support over 30 languages. A distinctive competitive advantage of our company are the hand-made, in-house dictionaries and other language resources spanning over 50 languages, created in the course of 25 years of the company’s existence by a team of linguists and language professionals. Use cases presented include real-time translation in emergency medical services, a phrase translation management platform for global enterprises, and the ARIEN project for monitoring illegal drug trafficking. Each demonstrates how Lingea’s tools ensure accurate communication, reduce operational delays and allow an efficient localization of services and products, significantly contributing to the effectiveness of multilingual interactions in various professional environments.

Keywords: NLP · MT · NMT · Machine translation · Language technology · Language service provider

1 Introduction

Lingea s.r.o. is a language service and technology provider based in Brno, Czech Republic. For over 25 years, we have developed a robust set of in-house, hand-made dictionaries and offer a suite of language tools designed to address the needs of our clients. We also provide language services, like human translation. We follow the newest trends in the world of Natural Language Processing (NLP) and Neural Machine Translation (NMT) and we focus on implementing practical solutions that address real-world language barriers. This short paper presents examples of how our language tools can be effectively integrated into diverse operational workflows.

- **Machine Translation:** State-of-the-art machine translation [1, 2], currently supporting combinations of more than 30 languages, possible addition of new languages made-to-order. Adaptation to the types of texts needed by the client. Support for formatted documents and enforcing user-defined terminology translation [3].

- **Keyword/Full-text Search:** Enhanced search tools that leverage custom lemmatization and morphological analysis.
- **Semantic Search:** Advanced search functionalities that use large language models to understand and retrieve information based on the meaning embedded in texts.
- **Predictive Writing Assistance:** Systems that suggest text continuations based on the context of user input. Adaptable both before and during the usage of the system.
- **Translation Database of Predefined Phrases:** A collection of frequently used phrases translated and stored for quick access.
- **Diacritics/Accents Completion:** Tools that accurately apply necessary diacritical marks to texts, using our dictionaries, morphological tools, and neural networks.
- **Grammar Correction/Autocorrect:** Systems that improve text by correcting grammatical errors through advanced machine learning algorithms and dictionaries.
- **Adaptable Autocomplete:** Real-time text suggestions that are adjusted based on user data and input.
- **Hand-made Dictionaries:** Extensive, human-made dictionaries developed over decades to support various language services.

Use Case 1: Emergency Medical Services

Accurate real-time translation in emergency medical services can be life-saving. We are working with the Emergency Service of a Pilsen region in Czechia, which borders the German state of Bavaria. Due to the proximity, there often arises a situation where the German ambulance picks up a person, writes an initial report in German, but drives to a Czech hospital, or vice versa. Our technology enables rapid translation of medical reports, diagnoses, and prescriptions between Czech and German. By utilizing predefined medical phrases and real-time translation feedback mechanisms, healthcare professionals can communicate more effectively, ensuring that critical medical information is conveyed accurately and quickly. We also integrate other language technologies, like adaptive predictive writing (i.e. smart autocomplete), grammatical error correction, and specialized terminology integration.

Use Case 2: Phrase Translation Management Platform

Phrase is an advanced translation management platform that streamlines the localization process for companies operating globally. This platform integrates our language tools, mostly machine translation, to handle vast amounts of content, maintaining consistency across translations and reducing turnaround times. With features like integration with existing customer systems and real-time performance measurement of different machine translation providers, Phrase helps

organizations achieve efficient and reliable localization of their services and products. Our cooperation with Pharse allows many new users to access our high-quality MT.

Use Case 3: ARIEN Project

The ARIEN project is part of the European Union Horizon initiative, which utilizes advanced AI tools to establish a real-time framework for monitoring illegal drug activities across the EU. In this collaborative effort, Lingea s.r.o. contributes specialized language technologies to a platform developed with 18 partners from 9 countries. Lingea's role involves integrating its language tools into a complex system that includes various technologies from other partners. This integration is essential for effective communication and data processing within the platform.

Examples of Operational Workflows:

Postal Monitoring:

1. Detection of drugs in a postal item triggers an OCR scan of accompanying documents in a foreign language.
2. Documents are translated and analyzed via keyword searches, with findings stored in a knowledge base.

Dark Web Surveillance:

1. An investigator has a suspicion about drug trade on a Dark Web site in a foreign language, they come up with search queries.
2. Queries are translated into the language of suspected Dark Web sites.
3. Retrieved data is translated back, aiding in investigative analysis.

Phone Call Surveillance:

1. Suspicious phone calls are transcribed with ASR
2. The transcription is translated into the language of the investigator
3. The translation is scrutinized through keyword searches, enhancing investigatory insights.

Social Media Analysis:

1. An investigator is suspicious about drug trafficking on a social network, with posts written in a foreign language
2. Social network posts are translated into the investigator's language
3. The translations undergo semantic search and analysis to uncover relevant data and connections.

These workflows demonstrate Lingea's crucial role in supporting ARIEN's goal of creating an effective, multilingual intelligence framework for combating drug trafficking throughout Europe.

Conclusion

Lingea’s language technologies not only facilitate communication but also significantly enhance the efficiency and accuracy of necessary translations in critical sectors such as healthcare, global commerce, and law enforcement. By adopting recent methods and developing practical solutions, we continue to support and improve multilingual interactions across diverse fields.

Acknowledgements

Funded by the European Union (project ARIEN; Grant agreement ID: 101121329). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

References

- [1] Josef Jon, Martin Popel, and Ondřej Bojar. “CUNI at WMT23 General Translation Task: MT and a Genetic Algorithm”. In: *Proceedings of the Eighth Conference on Machine Translation*. Ed. by Philipp Koehn et al. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 119–127. DOI: 10.18653/v1/2023.wmt-1.8. URL: <https://aclanthology.org/2023.wmt-1.8>.
- [2] Josef Jon, Martin Popel, and Ondřej Bojar. “CUNI-Bergamot Submission at WMT22 General Translation Task”. In: *Proceedings of the Seventh Conference on Machine Translation (WMT)*. Ed. by Philipp Koehn et al. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, Dec. 2022, pp. 280–289. URL: <https://aclanthology.org/2022.wmt-1.21>.
- [3] Josef Jon et al. “End-to-End Lexically Constrained Machine Translation for Morphologically Rich Languages”. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Ed. by Chengqing Zong et al. Online: Association for Computational Linguistics, Aug. 2021, pp. 4019–4033. DOI: 10.18653/v1/2021.acl-long.311. URL: <https://aclanthology.org/2021.acl-long.311>.