

# The Intersection of Applied Linguistics and Technology in Albania: Pedagogical Innovations, Computational Landscapes, and Future Horizons

**Rovena Mera**

Faculty of Sciences Medical and Technique, Department of Nursing and Physiotherapy  
Aldent University, Tirana, Albania

Email address: [rovena.mera@ual.edu.al](mailto:rovena.mera@ual.edu.al)

Orcid ID: <https://orcid.org/0009-0006-9740-0917>

doi: <https://doi.org/10.37745/ejells.2013/vol14n22329>

Published June 15, 2026

---

**Citation:** Mera R. (2026) The Intersection of Applied Linguistics and Technology in Albania: Pedagogical Innovations, Computational Landscapes, and Future Horizons, *European Journal of English Language and Literature Studies*, 14 (2),23-29

---

**Abstract:** *The study investigates the extent of the present incorporation of technology in applied linguistics and language education in Albania, a distinct low-resource language context. We explore Albanian higher education institutions' use of flipped learning, corpus-based methodologies, and communicative grammar education to facilitate greater linguistic competency through a line of study taking into account both CALL and computational linguistics. Simultaneously, the paper looks into some recent developments in Albanian Natural Language Processing (NLP), such as authorship attribution based on machine learning and construction of multi-dialectal corpus from social media sources. Finally, we address the systemic infrastructural, pedagogical, and digital divide barriers that hinder widespread adoption with evidence-based recommendations for policy and research. In the end, sustainable technological integration needs a balanced frame that marries computational innovation and humanistic, student-centred pedagogy.*

**Keywords:** applied linguistics, technology integration, low-resource language, Albanian NLP, flipped learning, corpus linguistics, digital divide.

---

## INTRODUCTION

Albania's emerging applied linguistics has been a vital focus field in language teaching, learning, and multilingual communication in an increasingly digital environment. Indeed, despite the impressive advancement made, the use of technology within the context of the research in linguistics (and languages) itself has remained relatively unexplored in the Albanian academic framework, relative to Western European institutions (Yfeku, 2023). Educational systems are

Publication of the European Centre for Research Training and Development-UK

starting to appreciate the life-changing capacity of digital tools, but adoption has been slow given existing infrastructure constraints and insufficient comprehensive pedagogical frameworks adapted to local circumstances.

In Albania, being a low-resource language context, technological adoption faces challenges, namely on Natural Language Processing (NLP) development and digital tool accessibility (Kryeziu et al., 2022). Globally, low-resource digital systems face serious disparities in the development of language technology owing to structural and commercial bias in favor of high-resource ecosystems (Joshi et al., 2020). In addition to limited digital infrastructure, some of the major challenges that the country's educational institutions face includes a training need for the teachers in these technology-based areas as well as low access to high-performance computers.

Although it holds great significance for cultural and linguistic identity research finds among Albanian NLP researchers, the Albanian language has not yet been well-represented in digital space. This lack of consistent, clean, and standardized datasets in machine learning training is mainly responsible for the digital divide, thus preventing advanced language technology that could support automated language learning and corpus-based research.

Nevertheless, these structural obstacles, there is a sense emerging in education now that technology has emerged as a potential tool to drive positive language outcomes. Some tertiary institutions have incorporated a flipped learning, computer aided instruction and data driven learning (DDL) in a context of the Albanian linguistic climate (Enesi et al., 2023). When tailored based on institutional resources and pedagogical objectives, these technologies mediated methods not only underpin interactive, learner-centered environments but also support fundamental digital literacies among educators and students alike.

## **METHODOLOGY**

This study seeks to evaluate the synthesis of technology and applied linguistics in Albania by using a secondary desk research methodology that provides a meta synthesis of recent empirical studies, ranging from 2022–2025. This methodological framework is divided into two central analytic vectors:

### *Pedagogical Evaluation Framework*

The pedagogical data assesses the efficiency of Technology-Enhanced Language Learning (TELL) in Albanian universities. In particular, we assess evidence from empirical case studies based on FLIP model (Flexible environments, learning through authentic text, Intentional content, and Professional educators) as used in English Language Teaching (ELT) Master students using the Compleat Lex tutor platform (Yfeku, 2023). Furthermore, comparative instructional data from university-based grammar modules (Alternative/Communicative vs. Traditional/Transmission-

based) are further compared and analyzed to understand impact on student reading abilities and communication accuracy (Enesi et al., 2023).

### *Computational Linguistics Assessment*

In the computational dimension, we discuss benchmarking, calibration and execution of machine learning (ML) and deep learning (DL) architectures in Albanian text corpora. The datasets we examine belong to public web-scraped corpora such as OSCAR and CCAligned, as well as localized datasets (Alb dataset) and geotagged social media user text (Twitter/X) with over 2,500 unique users (Çera & Sula, 2023).

The algorithmic architectures that are evaluated are as follows:

- XGBoost (Extreme Gradient Boosting) and Lexical Feature Extraction for stylometric and authorship classification.
- fastText word embeddings and Multilingual BERT transformers for deep-learning-based language comprehension and dialectal text pattern recognition.

## RESULTS

Data from the computational and pedagogical studies based on recent Albanian linguistic projects show very specialized performance metrics in different application areas. Table 1 provides a systematic organization of the empirical findings.

Table 1: *Performance Metrics and Domain Applications of Computational and Pedagogical Technologies for Albanian.*

| <b>NLP technology / Method</b>                              | <b>Application domain</b>  | <b>Data source / Corpus</b>   | <b>Performance metrics</b>            | <b>Observed effectiveness for Albanian</b>                                   |
|---|----------------------------|-------------------------------|---------------------------------------|--|
| <b>Corpus evaluation</b><br>( <i>Statistical analysis</i> ) | Dataset quality assessment | OSCAR, CCAligned, Alb dataset | Consistency & data quality benchmarks | Establishes crucial quality standards for downstream low-resource NLP tasks. |
| <b>XGBoost classification</b>                               | Authorship attribution     | Literary works &              | \$F_1 = 0.982\$ (Literary)            | Yields the highest accuracy among traditional                                |

Publication of the European Centre for Research Training and Development-UK

|  |                                  |   |   |   |
|--|----------------------------------|---|---|---|
|  |                                  | newsroom columns  | \$F_1 = 0.905\$ (Journalism)                        | machine learning classifiers.   |
| <b>Lexical feature extraction</b>          | Stylometric analysis             | Multi-genre Albanian corpus                                 | Lexical variance sensitivity                        | Proven as the most effective linguistic indicator over structural features.                 |
| <b>fastText embeddings</b>                 | Authorship & dialect recognition | Literary, journalistic, & social media text                 | Latent semantic accuracy                            | Highly resource-efficient and effective for sub-word features in morphologically rich text. |
| <b>BERT-multilingual</b>                   | Deep learning language modeling  | Diverse Albanian text genres                                | Contextualized semantic representation              | Competitive with fastText; highly scalable for multi-task linguistic processing.            |
| <b>Machine learning (multi-classifier)</b> | Dialect classification           | Geotagged Twitter/X data (Albania, Kosovo, North Macedonia) | Statistically superior to human annotator baselines | Successfully differentiates and maps subtle regional dialectal patterns.                    |
| <b>Geotagging techniques</b>               | Automated Corpus Construction    | User-generated content (\$N = 2,500+\$ users)               | High-throughput data ingestion                      | Enables large-scale dialectal corpus creation from authentic social web streams.            |
| <b>Geotagging techniques</b>               | Automated Corpus Construction    | User-generated content (\$N = 2,500+\$ users)               | High-throughput data ingestion                      | Enables large-scale dialectal corpus creation from authentic social web streams.            |

## DISCUSSION

### *Technology-Enhanced Language Learning (TELL) in Higher Education*

According to empirical evidence, for Albanian universities, flipped learning presents a promising pedagogical route. Educators can successfully scaffold asynchronous corpus-based activities through the use of free web-applicative platforms such as Compleat Lex tutor, which allows them to circumvent traditional spatial and temporal constraints (Yfeku, 2023). This directly addresses a critical pedagogical challenge: to combine corpus linguistics (the empirical analysis of real-life authentic language data) with active teaching approaches.

Related to a wider computer-assisted language learning literature, Data-Driven Learning (DDL) empowers students to be linguistic researchers, identify situational patterns, and develop inductive grammatical abilities (Boulton, 2021). Moreover, with the combination of communicative language teaching (CLT)/digital resources grammar pedagogy is transformed from dry memorization based on rules to conversational contextualized communication (Enesi et al., 2023). This is due to the fact that, as students are exposed to corpora as authentic linguistic data, grammar is learnt as a functional tool for expression of a real-world and not as an abstract number of rules. But this paradigm shift depends strongly on teacher competence.

Progressing to DDL and to contemporary technologies, we need robust professional development that marries technical ICT capability to evidence-based approaches to digital pedagogy and that is conceptualized at global level through frameworks such as Technological Pedagogical Content Knowledge (Mishra & Koehler, 2006; Yeleussiz & Tankibayeva, 2024).

### *Advancements in Albanian Computational Linguistics*

From the computational viewpoint, it is essential to assess existing corpora, such as OSCAR and CCAligned, to establish baseline quality metrics in low-resource situations (Kryeziu et al., 2022). Because of the structural and morphological intricacy of Albanian, one cannot easily obtain general web-based scraping with noise that can lead to systematic data bias in smaller languages, so dedicated benchmarking to eliminate bias is needed (Joshi et al., 2020). For reliable downstream tools such as machine translation or speech recognition, the need for data cleansing and dataset checks is critical. The high performance of XGBoost classifier ( $F1 = 0.982$  for literature) shows that hand-crafted lexical features can collect unique stylistic fingerprints of Albanian texts (Asllani et al., 2024). This is critical in forensic linguistics and content verification. Moreover, it has revolutionized our understanding of dialectology to use geotagged social media data to train ML models (Çera & Sula, 2023). By detecting regional differences across Albania, Kosovo, and North Macedonia, these models can beat human annotators and offer subtle linguistic changes that standard textbooks have failed to capture, making their significance the influence of computational resources in deepening descriptive sociolinguistics apparent.

*Digital Divides, Risks, and Structural Challenges*

Notwithstanding these technology advances, structural restrictions also pose a barrier for the fair advancement of the Albanian educational system:

- **Infrastructural & Geographical Gaps:** There remains a severe digital divide in Albania – between urban areas and remote or low-income regions. Inconsistent Internet connection, regular power outages, and outdated hardware lead to disparities in access to educational technologies (Sadigzade, 2025). This creates an imbalance that threatens to marginalize minority linguistic communities and rural students who most benefit from digital tools.
- **Pedagogical & Cognitive Risks:** Overreliance on automated feedback or poorly optimized AI tools present different linguistic risks. Automated platforms without scaffolding have been known to present learners with non-standard input (Sadigzade, 2025), limit social-pragmatic interaction or cause cognitive overload through distracting user interfaces. As Reinders and Benson (2017) stress, “educational technologies should help improve student autonomy and critical thinking, rather than reduce language learning to rote drilling with a machine model.”
- **Lack of teacher preparation courses:** Although language technologies have made incredible strides, the way we train teachers has been a lot behind the curve. A lot of language teachers do not have comprehensive training in tools of data analysis, corpus analysis software, or digital instructional design (Fahad, 2025). Without selective intervention, the real potential of these digital tools will stay enshrouded by a digital literacy deficit.

## **CONCLUSION AND RECOMMENDATIONS**

The integration of sustainable technology into Albanian applied linguistics will require an integrated programme of technical innovation and pedagogical strictness. Albania would have to develop strategies based on multiple points to move away from an under-provided digital environment to a technology-fueled digital ecosystem: Recommendation for Strategic Policy and Research.

- **Infrastructure and Equity Investments:** Government and institutional policies should prioritize building broadband capacity in rural regions and subsidizing digital devices through free school grants for students from low-income backgrounds to meet technology-enhanced learning goals worldwide.
- **Integrated in Curriculum Teacher Training:** Education is not only to focus on generalist learning on computers. They should invest in digital pedagogy and in educational frameworks structural mastery (TPACK), that is in a way helping teachers make the best use of corpus tools within student-centered curricula (Mishra & Koehler, 2006; Yeleussiz & Tankibayeva, 2024).
- **Development of Low-Resource Language Models:** Research funding should be directed towards the curation of clean, representative, and ethical Albanian dataset. AI and ML for Albanian

Morphology to protect the digital linguistic footprint (Ding et al., 2024; Joshi et al., 2020). A major focus must be on native Large Language Models (LLMs) and NLP architectures for Albanian morphology.

- Ethical AI Governance: Educational organizations should provide strong guidelines around ethical considerations regarding learners' privacy, data integrity, and algorithmic fairness in artificial intelligence-based assessment and language learning tools (Reinders & Benson, 2017). By tackling the digital divide, offering comprehensive teacher training, and facilitating collaboration in applied linguistics, data science and policymaking, Albania can leverage digital tools while maintaining the critical communicative and social dimensions of language teaching.

## REFERENCES

- Asllani, M., Çekrezi, E., Keta, A., & Fusha, E. (2024). Automatic authorship attribution in Albanian texts. *Journal of Quantitative Linguistics*, 31(2), 145–162.
- Boulton, A. (2021). Data-driven learning: From research to practice. *Language Teaching*, 54(4), 493–511.
- Çera, E., & Sula, R. (2023). Using twitter to collect a multi-dialectal corpus of Albanian using advanced geotagging and dialect modeling. *Languages*, 8(3), 201–219.
- Ding, Z., Liu, Z., Jiang, H., Gao, Y., Zhai, X., Liu, T., & Liu, N. (2024). Foundation models for low-resource language education (vision paper). *arXiv preprint arXiv:2403.11234*.
- Enesi, M., Strati, E., & Trifoni, A. (2023). Alternative grammar teaching vs. Traditional grammar teaching in Albanian tertiary education. *International Journal of Education and Practice*, 11(2), 294–305.
- Fahad, G. M. S. I. (2025). Digital literacy and self-efficacy of English language teachers in south Asia. *International Journal of Social Science and Human Research*, 8(1), 412–423.
- Joshi, P., Santy, S., Budhiraja, G., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293.
- Kryeziu, L., Shehu, V., & Çausi, A. (2022). Evaluation and verification of NLP datasets for the Albanian language. *Slovenian Journal of Artificial Intelligence*, 26(1), 35–48.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Reinders, H., & Benson, P. (2017). Research agenda: Technology and autonomy. *Language Teaching*, 50(4), 561–578.
- Sadigzade, Z. (2025). *Linguistic risks and resilience in digital language learning*. Academic Press.
- Yeleussiz, A., & Tankibayeva, M. Kh. (2024). Assessing teacher's media literacy competencies in secondary education. *Pedagogy and Education Review*, 14(2), 88–97.
- Yfeku, J. (2023). Alternative flipped learning approach scaffolding student engagement and collaboration: A case study with ELT master students in corpora in applied linguistics subject. *International Journal of Management, Knowledge and Learning*, 12, 115–131.