Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.16, No.2, 2026, pp.731-737

DOI: https://doi.org/10.51983/ijiss-2026.16.2.73

Computational Approaches to Morphosyntactic Analysis in Under-Resourced Languages

Solieva Malika Abduzukhurovna[1*], Xodjiyeva Gulchehra Norovna[2], Bahramov Otabek Tursunovich[3], Akhmedova Muyassar Khadimatovna[4], Kushmatova Dildora Ergashevna[5] and Aliyeva Shakhnoza Aybekovna[6]

[1*]Associate Professor, Department of Theory of Translation and Comparative Linguistics, National University of Uzbekistan named after Mirza Ulugbek, Tashkent, Uzbekistan

[2]Teacher Trainee, Samarkand State Institute of Foreign Languages, Samarqand, Uzbekistan

[3]Associate Professor, Faculty of Physical Education and Sports, Department of Sports and Methods of Teaching, Jizzakh State Pedagogical University named after Abdulla Qodiriy, Jizzakh, Uzbekistan

[4]Acting Professor, Department of Psychology Executor, Tashkent University of Applied Sciences, Tashkent, Uzbekistan

[5]Assistant, Samarkand State Medical University, Samarkand, Uzbekistan

[6]Faculty of Philology, Department of Uzbek Linguistics, Andijan State University named after Zahiriddin Muhammad Babur, Andijan, Uzbekistan

E-mail: [1]m.soliyeva@nuu.uz; malikakhon76@yandex.com, [2]gulchehraxodjiyeva74@gmail.com, [3]baxromovotabek7999@gmail.com, [4]muyassaraxmedova1968gmail.com, [5]kushmatova.dildora67@gmail.com, [6]aliyeva.shahnoza.83@mail.ru

ORCID: [1]https://orcid.org/0009-0001-9366-6471, [2] https://orcid.org/0009-0005-4849-6767, [3]https://orcid.org/0000-0003-4584-1703, [4]https://orcid.org/0009-0002-0914-3457, [5]https://orcid.org/0009-0007-8644-7514, [6]https://orcid.org/0009-0007-3803-3640

(Received 08 April 2026; Revised 12 May 2026, Accepted 21 May 2026; Available online 05 June 2026)

Abstract - **In recent years, Natural Language Processing has improved significantly, but most developments favor high-resource languages. Due to the absence of annotated corpora, lexical databases, and computational tools many languages are under-resourced. Morphosyntactic analysis is a fundamental component of several NLP tasks, including part-of-speech tagging, syntactic parsing, and machine translation. It focuses on understanding words forms and sentence structure in a language. The under-resources languages have limited linguistic resources and contain complex morphological structures which makes the development of computational approaches for morphosyntactic analysis challenging. This study reviews different computational methods used for morphosyntactic analysis in under-resourced languages. 19 studies were reviewed to examine the technique used, research areas and performance trends in the literature. The results show that the percentage of neural network-based is about 42% of the analyzed literature, then statistical ML methods 32% and the rule-based approaches 26%. In Performance comparison the neural network models (82%) achieve higher accuracy compared to statistical models (75%) and rule-based models (70%). Also, it shows that the morphological processing and language resource development is the most investigated research areas. These developments and performance of the models are affected by limited annotated datasets and linguistic diversity. The results show the need for better linguistic resources and hybrid computational approaches for morphosyntactic analysis in under-resourced languages. These can help guide future research to develop better NLP tools for these languages.**

Keywords: **Computational Linguistics, Morphosyntactic Analysis, Under-Resourced Languages, Natural Language Processing, Neural Network Models, Language Resources**

I. INTRODUCTION

Advancements in ML and the accessibility of large linguistic datasets have significantly improved the progress of natural language processing (NLP). However, current NLP research mainly focuses on high-resource languages, and many other languages still lack computational resources. The absence of essential linguistic resources such as annotated corpora, lexical databases, and computational tools limits the development of NLP systems. In many NLP applications, morphosyntactic analysis plays a vital role particularly in tasks such as part-of-speech tagging, syntactic tagging, and machine translation.

Morphosyntactic analysis becomes important in languages with rich morphological structures, where words often appear in multiple forms and with complex grammar patterns. But for under-resourced languages the lack of linguistic resources and annotated datasets makes it difficult to develop computational models.

These challenges have been addressed by several studies to develop language technologies and linguistic resources for low-resource languages.

Many computational tools and language technologies for African and indigenous languages have been explored to support digital language processing (Puttkammer & Du Toit, 2021; Don & Knowles, 2022). Likewise, computational analysis in low-resource environments has been supported by the development of linguistic datasets and annotation methods (El-Haj et al., 2015; Felt et al., 2014). Advances in machine learning and neural network techniques have further improved the ability to process morphologically rich languages. Neural models such as BiLSTM-based tagging systems and neural morphological tagging models have shown good results in morphosyntactic analysis tasks (Yu et al., 2025; Chakrabarty et al., 2019).

Research on computational approaches for morphosyntactic analysis in under-resourced languages is still distributed across different methods and linguistic settings. Therefore, a systematic synthesis of existing studies is needed to identify the main computational techniques, evaluate their performance and highlight research gaps in this area.

The key contributions of this study are

1. Systematic review of computational approaches for morphosyntactic analysis in under-resourced languages is provided.

2. The distribution of computational techniques (neural models, statistical methods, and rule-based approaches) is analyzed.

3. Research trends, challenges and emerging directions in the development of NLP systems for under-resourced languages are examined.

The remainder of this paper is organized as follows. Section 2 reviews related works on the proposed approaches. Section 3 is the research methodology used for the systematic review. Section 4 presents the findings obtained from the reviewed studies. Section 5 discusses the implications and limitations of the findings. Finally, Section 6 concludes the study and outlines the future research.

II. LITERATURE REVIEW

Several areas such as the development of linguistic resources, computational models and language technology tools have been the focus of research on the processing of under-resourced languages. The key research area involves developing lexical databases, morphological dictionaries, and annotated corpora for languages with limited resources (El-Haj et al., 2015; Bosch & Griesel, 2017; Espla-Gomis et al., 2017). Few other research focuses on the development of NLP tools and computational approaches for specific languages. Several studies have proposed new language technologies and NLP toolkits for languages such as Welsh, Runyakitara, and other indigenous languages (Cunliffe et al., 2022; Katushemererwe et al., 2021). These works support the linguistic analysis and improve the availability of computational tools for under-resourced languages. Machine learning models have also been widely used to deal with morphosyntactic analysis challenges.

In morphological tagging and syntactic analysis, neural network models have shown strong performance. For example, the models with neural morphological tagging have been applied to low-resource languages to improve morphosyntactic processing (Chakrabarty et al., 2019). Also, BiLSTM-based tagging approaches have been used to improve morphological and lexical analysis in NLP tasks (Yu et al., 2025). Statistical models and speech processing systems have also been explored for under-resourced languages in addition to neural approaches. Languages with limited linguistic resources are being processed using speech recognition and language modeling techniques (Kurimo et al., 2017; Tachbelie et al., 2014). Researchers have also developed language identification methods to assist multilingual environments (Selamat & Akosu, 2016). Recent studies highlight the importance of multilingual learning and transfer learning techniques. These methods are used to apply models trained on high-resource languages to low-resource languages (Nzeyimana, 2024; Wiemerslage et al., 2022). Additional studies highlight the need to address resource asymmetry in multilingual NLP systems and to expand linguistic datasets for under- resourced languages (Pakray et al., 2025; Ranathunga & De Silva, 2022).

The existing studies shows that the research in this field has improved from resource development towards ML based computational approaches. Despite these advancements, challenges such as limited annotated datasets and linguistic diversity affect the development of morphosyntactic analysis for under-resourced languages.

III. MATERIALS AND METHODS

Study Design

A systematic literature review is adopted with synthesizing the qualitative evidence to investigate computational approaches used for morphosyntactic analysis in under-resourced languages. This approach helps in identification and evaluation of findings from existing studies. It also synthesis the studies related to computational morphology, part-of -speech tagging and syntactic analysis in low-resource language contexts. The results from multiple studies were combined and analyzed. This identified research trends, commonly used computational techniques and existing research gaps in NLP. The review process has structures that are applied during systematic review methods. It also complied with PRISMA in the identification, screening and selection of the relevant studies.

Search Strategy

Detailed literature search was conducted to find out the studies which are relevant and do concentrate on computational methods of morphosyntactic analysis of under-resourced languages. The search was conducted in

several academic databases that offer a wide range of coverage of research publication in the field of computational linguistics and artificial intelligence. These included ACL Anthology, IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar.

The publications produced between 2015-2025 are considered to reflect recent developments in computational techniques for morphosyntactic processing A keyword-based search strategy was employed using combinations of terms related to morphosyntactic analysis and low-resource natural language processing. Morphosyntactic analysis, computational morphology and part of speech tagging are included as search queries. They also included terms like low-resourced languages, under-resourced languages in NLP, neural models for morphology and multilingual language models. Boolean operators were used to combine these terms and to find relevant publications. In addition to database searches, the reference lists of selected articles were also checked manually to identify additional relevant studies.

Eligibility Criteria

Inclusion and exclusion criteria were applied during the screening process to ensure the relevance and quality of the selected studies as shown in table I.

TABLE I INCLUSION AND EXCLUSION CRITERIA

Criteria Type	Description
Inclusion	Studies on morphosyntactic analysis or computational morphology
Inclusion	Research on under-resourced or low-resource languages
Inclusion	Studies using rule-based, statistical, or neural computational approaches
Inclusion	Peer-reviewed journal articles or conference papers
Inclusion	Publications between 2015 and 2025
Exclusion	Studies on high-resource languages
Exclusion	Studies lacking computational methodology
Exclusion	Non-scholarly sources such as blogs or editorials
Exclusion	Duplicate or incomplete studies

Study Selection

All the records retrieved were imported into a reference management system, where duplicate entries were identified and removed. The remaining studies were screened in two stages.

At first, the title and abstract of the retrieved records were examined and their relevancy to morphosyntactic analysis in under resourced languages are determined. Studies that did not meet the inclusion criteria were excluded during this stage. Next, the full text of the remaining articles was reviewed to determine the eligibility for inclusion in the systematic review. The study that satisfies all inclusion criteria were only retained for further analysis. Any differences about study inclusion were resolved through discussion to ensure consistency in the selection process. For study identification and selection process PRISMA framework was followed. This framework includes stages such as identification, screening, eligibility assessment and final inclusion of studies. The complete study selection process is illustrated in fig. 1.

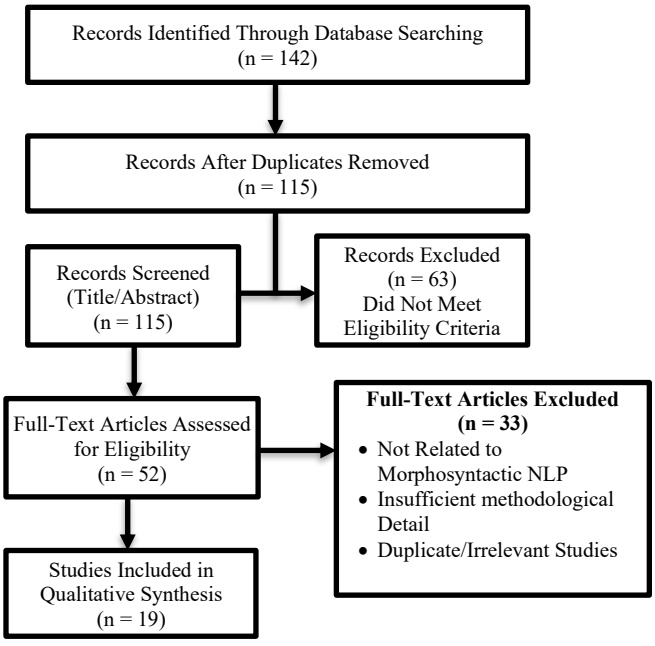

Fig. 1 PRISMA Flow Diagram of Study Identification and Selection

Initially, a total of 142 studies were initially identified through database searches. 27 duplicate records were removed and the remaining 115 were kept for title and abstract screening. At this stage, 63 studies were excluded because they did not meet the pre-defined inclusion criteria. The remaining 52 full text articles were assessed for eligibility.

33 studies were excluded out of these, due to insufficient methodological information or lack of relevance to the analysis in under-resourced languages. In the final dataset 19 studies were included for synthesis and analysis.

Data Extraction

Structured extraction framework was used for data extraction to ensure consistency in recording information from the selected studies. Relevant information from each study was recorded systematically. This included the author and publication year, the language or dataset used, and the computational methods applied. It also included the morphological tasks studied, evaluation metrics used and the reported performance results. The extracted information was reviewed to verify its accuracy and completeness before proceeding with further analysis.

Data Synthesis and Analysis

The selected studies were categorized according to the computational technique used for morphosyntactic analysis. During this process the identified methods were grouped into 3 major categories, namely rule-based approaches, statistical ML approaches, and neural - network based approaches. Rule-based approaches mainly depend on linguistic rules and handcrafted morphological grammars. Probabilistic models are included by statistical approaches. Neural network-based approaches include deep learning architectures such as recurrent neural networks, LSTM networks, and transformer-based language models. A comparative analysis was conducted to examine the effectiveness and applicability of these computational methods in under-resourced languages. The analysis is based on the performance of model, data requirements, scalability, and the ability of various computational approaches to handle complex morphological structures. Then the synthesized findings were used to identify major research trends and highlight promising directions for future research in low-resource NLP.

IV. RESULTS

4.1 Characteristics of Included Studies

Various computational approaches used for processing under -resourced languages are reviewed in the study. Those include morphological tagging, syntactic analysis, language resource development, speech recognition, language identification. The selected studies represent various linguistic contexts, with research which focused on African languages, Indigenous languages and other low-resource linguistic groups. Challenges such as limited annotated datasets and complex morphological structures are addressed in these studies through the analysis of various computational techniques.

4.2 Distribution of Computational Approaches

The computational techniques were grouped into three main categories- neural network-based approaches, statistical machine learning models, and rule-based systems to understand the working principle of the reviewed studies.

TABLE II SUMMARY OF COMPUTATIONAL APPROACHES IDENTIFIED IN THE REVIEWED STUDIES

Method	No. of Studies	Percentage (%)	Common Techniques
Neural Network-Based	8	42%	LSTM, BiLSTM, CNN, Transformer models
Statistical Machine Learning	6	32%	HMM, CRF, Maximum Entropy
Rule-Based Approaches	5	26%	Finite-state rules, handcrafted grammars

Table II present the summary of the various computational approaches used in this study. The largest proportion of computational techniques used in the reviewed studies consists of neural network approaches (42%). This shows that the field depends on deep learning models for morphosyntactic processing tasks. Statistical machine learning approaches shows 32% but are still widely used as they can work effectively with smaller datasets. Rule-based systems are used in conditions where linguistic resources are very limited.

4.3 Performance of Morphosyntactic Models

Performance metrics for computational models including accuracy, precision and F1-score are reported in the reviewed studies. General performance trends can be identified from the literature, although the reported values vary across different datasets and languages.

TABLE III ESTIMATED PERFORMANCE COMPARISON OF COMPUTATIONAL APPROACHES

Method	Average Accuracy (%)	Average F1-Score (%)	Average Precision (%)
Neural Network-Based	82	80	78
Statistical Machine Learning	75	73	71
Rule-Based Approaches	70	68	65

Table III shows that neural network models achieve higher performance compared to a statistical and rule-based methods. These models are capable of identifying the relationships between contextual linguistic patterns and

complex morphological patterns. In case of computational resources or training datasets. In cases where computational resources or training datasets are limited, statistical and rule-based approaches are still widely applied.

4.4 Publication Trend of Studies

The publication years of the reviewed studies were analyzed to examine the growth of research activity in this field.

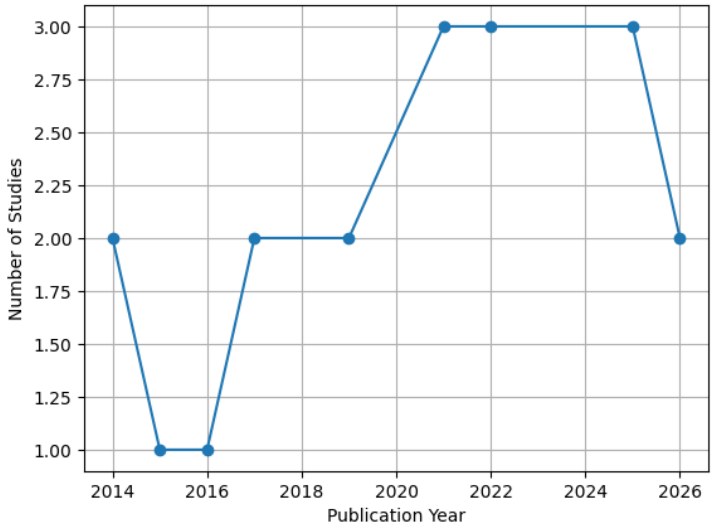

Fig. 2 Publication Trend of Studies on Under-Resourced Language Processing

Fig. 2 shows the distribution of the reviewed studies over time. It also demonstrates how the activity of research in this direction has been gradually growing over the years. The prior research was dedicated to language development and linguistic recording, and the recent research is based on machine learning and neural networks to perform morphosyntactic analysis.

4.5 Distribution of Research Focus Areas

The literature was analyzed further to establish the main research themes in the literature.

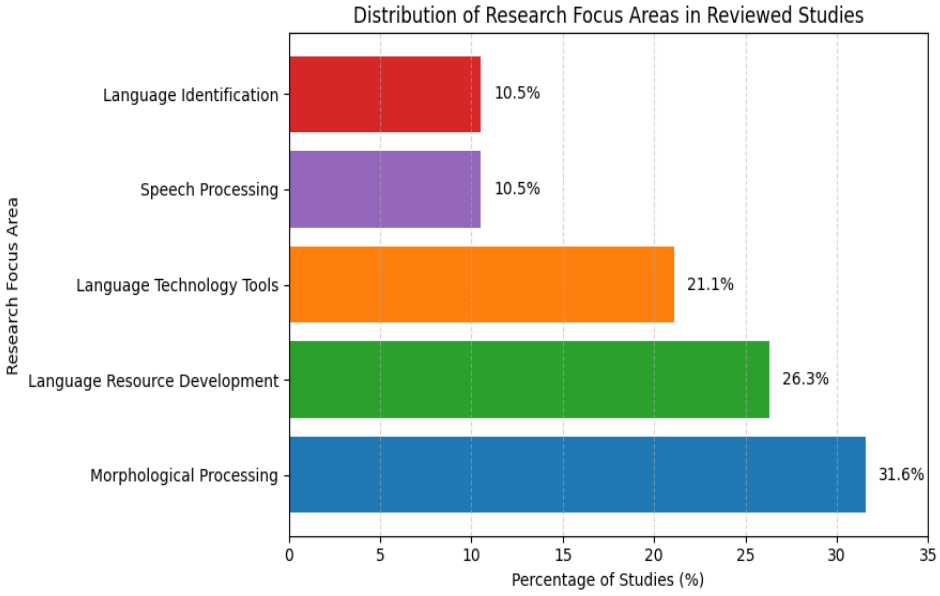

Fig. 3 Distribution of Research Focus Areas in Reviewed Studies

Morphological processing represents the most commonly researched area which was about 31.6% of the reviewed papers, as illustrated in fig. 3. Another popular field is the language resource development that emphasizes the role of annotated datasets and linguistic resources in computational analysis. Language technology tools, speech processing and language identification are some of the few areas that lead to the development of NLP systems of under-resourced languages.

V. DISCUSSION

The paper highlights many important patterns in computational approaches used for morphosyntactic analysis in under-resourced languages. The main findings indicate that neural network-based models are being increasingly used, especially deep learning models like LSTM and transformer-based models (Grifoni et al., 2016). Such strategies are more successful in morphosyntactic tasks because they can more effectively interpret the contextual and sequential linguistic patterns compared to traditional models (Chakrabarty et al., 2019). It is also revealed that, when annotated datasets are small, statistical ML techniques and rule-based models are effective (Pakray et al., 2025). Many under-resourced languages do not have large-scale datasets needed to effectively train deep learning models (Ranathunga & De Silva, 2022). And statistical models and rule-based methods still provide a practical solution in cases where there is limited data. The significance of language resource development is also mentioned in the review. A significant portion of the studies reviewed are related to the development of annotated corpora, lexical databases, and morphological dictionaries (Grifoni et al., 2016).

These resources are essential for training and evaluating computational models and it remain a important requirement for improving research in low-resource language processing. Although these advances have been made, the review has a few current challenges like the small availability of data sets, and complex morphological structures (Erdmann & Habash, 2018). These problems are present in most under-resourced languages to create challenges to computational systems. Complex languages have complex word form and they need models that are capable of accommodating various morphological variations. According to the current trends in research, multilingual and transfer learning methods can be a solution to these issues (Ronny Mabokela et al., 2025). These methods employ the information obtained in the rich languages to enhance the performance of the models in low-resource languages. In general, the results indicate that there is a great improvement in this field (Erdmann & Habash, 2018). However, to continue to broaden the range of linguistic datasets in most of the under-resourced languages, as well as to learn to create computational models capable of competently addressing their intricate morphological patterns, more work is necessary in the future.

VI. CONCLUSION

A systematic synthesis was used to review 19 studies to assess the suitability of computational methods of morphosyntactic analysis of under-resourced languages. Neural network-based methods are among the most commonly known computational methods, and were reported in approximately 42% of the studies reviewed. Statistical machine learning techniques indicated 32% of the studies, and rule-based techniques reported 26%. The findings suggest that more complex morphosyntactic tasks are increasingly being modeled with deep learning models like LSTM and transformer-based systems. The reviewed studies demonstrate the effectiveness of neural models with an average accuracy of about 82% as opposed to 75 in statistical models and 70 in rule-based systems. The benefits of deep learning models are evident in its capacity to comprehend contextual linguistic patterns in morphologically rich languages. The most common research focus areas identified were morphological processing (31.6%), and language resource development (26.3%). These results underscore the significance of linguistic resources and computational tools to support under resource languages. Although the developments have been made, the absence of annotated datasets, and complicated morphological structures are the challenges that slow down the development in the field. Future directions are in the direction of increasing multilingual datasets in under-resourced languages. It is also able to analyze hybrid models, which integrate neural methods with linguistic knowledge.

REFERENCES

[1] Bosch, S. E., & Griesel, M. (2017). Strategies for building wordnets for under-resourced languages: The case of African languages. *Literator: Journal of Literary Criticism, Comparative Linguistics and Literary Studies*, *38*(1), 1-12. https://doi.org/10.4102/lit.v38i1.1351

[2] Chakrabarty, A., Chaturvedi, A., & Garain, U. (2019). NeuMorph: Neural Morphological Tagging for Low-Resource Languages—An Experimental Study for Indic Languages. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, *19*(1), 1-19. https://doi.org/10.1145/3342354

[3] Cunliffe, D., Vlachidis, A., Williams, D., & Tudhope, D. (2022). Natural language processing for under-resourced languages: Developing a Welsh natural language toolkit. *Computer speech & language*, *72*, 101311. https://doi.org/10.1016/j.csl.2021.101311

Don, Z. M., & Knowles, G. (2022). The digital humanities and re-imagined language description: A linguistic model of Malay with potential for other languages. *Digital Scholarship in the Humanities*, *37*(4), 1084-1096. https://doi.org/10.1093/llc/fqab101

[4] El-Haj, M., Kruschwitz, U., & Fox, C. (2015). Creating language resources for under-resourced languages: methodologies, and experiments with Arabic. *Language Resources and Evaluation*, *49*(3), 549-580. https://doi.org/10.1007/s10579-014-9274-3

[5] Erdmann, A., & Habash, N. (2018, October). Complementary strategies for low resourced morphological modeling. In *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology* (pp. 54-65). https://doi.org/10.18653/v1/W18-5806

[6] Espla-Gomis, M., Carrasco, R. C., Sánchez-Cartagena, V. M., Forcada, M. L., Sánchez-Martínez, F., & Pérez-Ortiz, J. A. (2017). Assisting non-expert speakers of under-resourced languages in assigning stems and inflectional paradigms to new word entries of morphological dictionaries. *Language Resources and Evaluation*, *51*(4), 989-1017. https://doi.org/10.1007/s10579-016-9360-9

[7] Felt, P., Ringger, E. K., Seppi, K., Heal, K. S., Haertel, R. A., & Lonsdale, D. (2014). Evaluating machine-assisted annotation in under-resourced settings. *Language resources and evaluation*, *48*(4), 561-599. https://doi.org/10.1007/s10579-013-9258-8

[8] Grifoni, P., D'Ulizia, A., & Ferri, F. (2016). Computational methods and grammars in language evolution: a survey. *Artificial Intelligence Review*, *45*(3), 369-403. https://doi.org/10.1007/s10462-015-9449-3

[9] Katushemererwe, F., Caines, A., & Buttery, P. (2021). Building natural language processing tools for Runyakitara. *Applied Linguistics Review*, *12*(4), 585-609. https://doi.org/10.1515/applirev-2020-2004

[l0] Kurimo, M., Enarvi, S., Tilk, O., Varjokallio, M., Mansikkaniemi, A., & Alumäe, T. (2017). Modeling under-resourced languages for speech recognition. *Language Resources and Evaluation*, *51*(4), 961-987. https://doi.org/10.1007/s10579-016-9336-9

[11] Nzeyimana, A. (2024, June). Low-resource neural machine translation with morphological modeling. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 182-195). https://doi.org/10.18653/v1/2024.findings-naacl.13

[12] Pakray, P., Gelbukh, A., & Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. *Natural Language Processing*, *31*(2), 183-197. https://doi.org/10.1017/nlp.2024.33

[13] Puttkammer, M., & Du Toit, J. S. (2021). Canonical segmentation and syntactic morpheme tagging of four resource-scarce nguni languages. *Journal of the Digital Humanities Association of Southern Africa (DHASA)*, *3*(03), 1-7. https://doi.org/10.55492/dhasa.v3i03.3818

[14] Ranathunga, S., & De Silva, N. (2022, November). Some languages are more equal than others: Probing deeper into the linguistic disparity in the NLP world. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 823-848). https://doi.org/10.18653/v1/2022.aacl-main.62

[15] Ronny Mabokela, K., Primus, M., & Celik, T. (2025). Advancing sentiment analysis for low-resourced african languages using pre-trained language models. *PloS one*, *20*(6), e0325102. https://doi.org/10.1371/journal.pone.0325102

[16] Selamat, A., & Akosu, N. (2016). Word-length algorithm for language identification of under-resourced languages. *Journal of King Saud University-Computer and Information Sciences*, *28*(4), 457-469. https://doi.org/10.1016/j.jksuci.2014.12.004

[17] Tachbelie, M. Y., Abate, S. T., & Besacier, L. (2014). Using different acoustic, lexical and language modeling units for ASR of an under-resourced language–Amharic. *Speech Communication*, *56*, 181-194. https://doi.org/10.1016/j.specom.2013.01.008

[18] Wiemerslage, A., Silfverberg, M., Yang, C., McCarthy, A. D., Nicolai, G., Colunga, E., & von der Wense, K. (2022, May). Morphological processing of low-resource languages: Where we are and what's next. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 988-1007). https://doi.org/10.18653/v1/2022.findings-acl.80

[19] Yu, H., Cho, Y., Park, G., & Kim, M. (2025). KRongBERT: Enhanced factorization-based morphological approach for the Korean pretrained language model. *Information Processing & Management*, *62*(3), 104072. https://doi.org/10.1016/j.ipm.2025.104072