Executive Summary

of

The Thesis Entitled

CONTEXT AWARE SPELL CHECKER FOR GUJARATI LANGUAGE USING DEEP LEARNING BASED HYBRID MODELS

Submitted By

PANCHAL BRIJESHKUMAR YOGESHBHAI (FOTE/1082)

Under Guidance of

PROF. DR. APURVA M. SHAH

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF TECHNOLOGY AND ENGINEERING
THE MAHARAJA SAYAJIRAO UNIVERSITY OF BARODA
VADODARA 390 001

February 2026

TABLE OF CONTENT OF THE EXECUTIVE SUMMARY

Table of Contents of Thesis

INTRODUCTION

At present, researchers stand at a juncture where the relationship between technology and language is not just an option but a necessity. The researcher presented here discusses “How technologies like Artificial Intelligence (AI) Natural Language Processing (NLP) can play a promising role in the revival of Indian languages, especially Gujarati. Along with this, the question arises: will technology advance or destroy the language? Researcher seeks solutions to such questions, which are related to the vitality of the language, research, employment, and the future of the youth, as well as this research clarifies the role of the Gujarati language in the technological age in the context of Artificial Intelligence [1].

In current scenario NLP is an ascension research work around the world, but when one can talk about India only, one question should ask by researcher or product-based software company, why all are working more on NLP orientated work, compared to other technologies. So, answer is here, according to research done by Google in collaboration with KPMG (Klynveld Peat Marwick Goerdeler) in 2017, the period of English being control on Internet use in India may be coming to a fall. According to research "Indian Languages: Defining India's Internet, "there are around 234 million plus Indian language handlers online in India and 175 million English language handlers. The number of internet users using the Indian language increased by 234 million at a CAGR (Compound Annual Growth Rate) of 41% between 2011 and 2016. Due to this impressive development, Indian language Internet users have surpassed their English-speaking counterparts. From 2017, end user of local languages has been increasing at a rate of 18% per year to reach 536 million (Mn) plus, compared to the 3% growth of English-speaking users to 199 million. Approximately, 75% plus of India's online consumers would be speaking their native language. Hindi-language content consumption is expected to surpass that of English, according to Google. In 2017, the survey discovered that services like messaging (169 Mn), digital entertainment (167 Mn), social media (115 Mn), and digital news have been the most popular in driving Internet usage via local languages (106 Mn). In 2021, messaging increased to 396 million users, digital entertainment to 392 million users, social media to 301 million users, and digital news to 284 million users [2]. From this discussion, it is evident that intelligent spell and grammar checkers for Indic languages are now more necessary than ever.

Problem Description and Research Gap

Gujarati, like numerous Indian languages, requires automated grammar and orthography correction systems that prioritize contextual accuracy, despite the growing prevalence of digital

platforms for regional languages. Jodani and Saras are two examples of solutions that mostly use rules-based methods. These methods function well, but they are not flexible or deep enough semantically to handle mistakes in the real world, inflectional morphology, and changes in context. These algorithms are not very good at figuring out linguistic patterns that are hard to understand, which makes them less useful in real-world NLP applications. Also, there are not many annotated corpora, linguistic resources, or complicated computational models available, which make it tougher for these tools to work in Gujarati.

The main research gap addressed in this proposed study is the absence of an accurate spell checker for the Gujarati language. Researchers definitely need a clever Gujarati spell checker that can fix both real and made-up words by using both lexical correction methods and deep learning models that can pick up on little changes in meaning and syntax. A hybrid model that uses both rule-based spelling logic and Deep Learning transformer-based and sequence-based contextual learning (like IndicBERT and GRU) can make spell checkers work better and be more effective for Gujarati and other Indian languages with few resources.

Motivation

The primary motivation for this research is the lack of a robust Gujarati spell checker developed using a real-world, authentic dataset. Existing tools are either limited in accuracy, based on synthetic or insufficient data, or not designed to meet the linguistic and research requirements of Gujarati language users. There are a lot of multilingual AI systems out there, such as ChatGPT, Claude, Google Gemini, Zapier Agents, Microsoft Copilot, DeepSee, Perplexity, Meta AI, Zapier Chatbots, Grok, and ChatSUTRA. As a researcher, you should question yourself: Where do these huge AI systems store their multilingual datasets? Their strong resource base contributes to improved task accuracy.As India's regional material goes increasingly digital, Gujarati, Hindi, and Assamese are being utilized more and more in social media, e-governance, education, and digital communication. This means that it is very necessary to have accurate NLP techniques, especially for fixing spelling mistakes, to maintain the text quality good. Mistakes in spelling when you talk to someone online might make you look less trustworthy, make the law less clear, and even make your computer less safe. When you spell things wrong, it makes NLP tasks like machine translation, text summarization, sentiment analysis, and ChatBots less effective. English and other global languages have more tools, but Indian languages don not have as many datasets, resources, or tools.

There are not many good spell-checking tools for Gujarati. Most Gujarati spell checkers, like Jodani and Saras, use Rules-Based and Heuristic methods. These systems don't get context, have difficulties with homophones, and don not work well in the actual world. The script for Gujarati

comprises a variety of distinct forms, diacritics, compound characters, and inflections, which makes it tougher to discover and rectify faults. These kinds of language distinctions can't be found with a basic dictionary search. A lot of errors in Gujarati are real-word mistakes, which implies that words that are correct are used in the wrong way. You need to understand the context through deep learning and transformer models in order to remedy these kinds of problems. Using both old and new methods, such Norvig's probabilistic method and GRU and IndicBERT, is a good way to deal with both lexical and contextual data. This motivates the development of hybrid systems for rectifying Gujarati orthography.

Problem Statement

The growing use of the Gujarati language online shows a big problem: spell checkers do not work for mistakes that happen in context. They have a hard time fixing words that are spelled correctly but used in the improper grammatical or semantic context, or finding fake words. A far smarter Gujarati spell checker is clearly and urgently needed. This complex tool must do more than just mend basic spelling problems; it must also have a comprehensive understanding of Gujarati grammatical rules, sentence structure, and how words can signify different things in different situations.

A hybrid strategy is necessary to reach this level of accuracy and understanding of the situation. This method would work well because it would use the proven effectiveness of the Peter Norvig spelling correction algorithm, which is great for finding non-words and common typos, along with the strong contextual understanding abilities of deep learning models like IndicBERT (which was specifically trained on Indian languages) and GRU neural networks. Combining these methods is the greatest way to make suggestions that are truly dependable and take context into account. This will greatly improve the accuracy and overall reliability of Gujarati language processing for all users, making it easier to access and communicate clearly.

Objectives

The objectives of the research work are as follows:

- To develop a dataset of very high quality that has both correct and incorrect Gujarati sentences. The suggested models will be trained on this dataset, and their performance will be measured using natural language processing metrics such as F1-score, recall, accuracy, and precision.
- To create hybrid spell-checking models for Gujarati sentences that take into account the spelling's context and surroundings. These models will employ a mix of strategies to help people understand context and sequence better. Some of these algorithms are GRU, IndicBERT, and Peter Norvig's probabilistic spelling correction.
- To optimize the proposed models through hyperparameter tuning, ensuring improved

performance on contextually complex linguistic inputs.

- To conduct a comprehensive comparative analysis between the proposed models and existing spell-checking systems to evaluate their effectiveness and accuracy.

Scope

The primary scope of this research is the development of two novel hybrid models GUJAPUBRIJ (Norvig + GRU) and GUJBRIJAPU (Norvig + IndicBert) for Gujarati spell correction by considering contextual understanding and surrounding circumstances. Presented a comparative analysis with existing tools (Jodani and Saras), highlighting the superiority of deep learning models in terms of precision, recall, and adaptability. The Secondary Contribution are Developed a Gujarati dataset of 20,000 sentences (correct and incorrect) sourced from publicly available resources, custom-made for spelling correction research.Explain Gujarati Grammatical rules of RHASVA, DĪRGHA AND ANUSVĀRA in plainly English for Guajarati NLP research. Introduced an explainable and extendable architecture that can be scaled with additional data or adapted for other low-resource Indic languages in future research.

Research Methodology for Work Done

As an initial start in looking at spell checking for Gujarati, two attempts were undertaken. The first test of Peter Norvig's probabilistic model used 16,937 words and got 80–90% of them right. However, its English-origin design did not work well for Gujarati because of the language's complexity, lack of context, and inadequate handling of rich morphology. This shows that adaptive solutions are needed. Second, a system that merged Norvig's algorithm with rule-based methods used rhasva, dīrgha, and anusvāra to include phonetic features and the barakhadi chart for vowels and nasalization. Despite adding contextual logic (e.g., adjacent words), it retrieved correct words in only 0–25% of test cases. Both attempts showed Norvig’s method lacks the linguistic depth and contextual sensitivity needed for morphologically rich, low-resource Gujarati. The restrictions of rule-based and statistical approaches underscored the requirement for advanced techniques. Consequently, both studies propose a deep learning-based hybrid model using RNN/LSTM/Bi-LSTM/GRU/IndicBERT architectures to address these challenges and improve Gujarati spell correction accuracy [29] [73].

As per above discussion researchers finally proposed models for the Gujarati language use Peter Norvig's algorithm for correcting spelling to ensure spelling accuracy by breaking the supplied text into smaller parts. Peter Norvig's proposed approach uses probability and edit distance to handle mistakes. Using a Gujarati lexicon, the first version of the Gujarati spell checker followed

Peter Norvig's methods. The groundwork for finding and fixing spelling mistakes has been laid by this method. Finding and fixing mistakes that do not contain words is a breeze using this procedure. While it works well with a specified lexicon, understanding context is a bit of a difficulty. The first GUJAPUBRIJ model Gated Recurrent Unit (GRU)-based uses a neural network to check context of the text, while the second model GUJBRIJAPU an IndicBERT based uses a neural network. Both methods enhance the Spelling checking process by keeping an eye on contextual interdependencies [82].

- Proposed two novel hybrid models GUJAPUBRIJ and GUJBRIJAPU for Gujarati spell correction by considering contextual understanding and surrounding circumtances:
    - Peter Norvig + GRU – GUJAPUBRIJ (sequence modeling for contextual meaning handling)
    - Peter Norvig + IndicBERT – GUJBRIJAPU (transformer-based contextual analysis)
- Developed a curated Gujarati Dataset of 20,000 sentences (correct and incorrect) sourced from publicly available resources, tailored for spelling correction research.
- Achieved high accuracy and contextual reliability, with the IndicBERT-based model outperforming all others:
    - Accuracy: 93.49%
    - Precision: 94.46%
    - Recall: 90.13%
    - F1 Score: 91.59%
- Presented a comparative analysis with existing tools (Jodani and Saras), highlighting the superiority of deep learning models in terms of precision, recall, and adaptability.
- Introduced an explainable and extendable architecture that can be scaled with additional data or adapted for other low-resource Indic languages in future research.

CONCLUSION

Outcome of Research Work

According to the literature survey, Natural Language Processing (NLP) needs spelling checkers because researchers help make sure that text data is clear, correct, and of high quality. A lot of work has gone into building these kinds of tools for languages with a lot of resources, like English. But Indian regional languages like Gujarati are still very challenging to deal with because their morphology is intricate, they have a lot of phonetic variants, and they use distinct scripts. It is challenging to design error detection systems that are both accurate and aware of the context because of how complicated these languages are. Jodani and other classic rule-based tools are fast and do not use a lot of computer power, but they do not always work well with grammar and do not have the semantic context needed to fix mistakes in the real world.To deal with these problems, this study introduces and tests two new hybrid models GUJAPUBRIJ / GUJBRIJAPU that combine Peter Norvig's probabilistic spelling correction algorithm with advanced deep learning frameworks. These are Gated Recurrent Unit (GRU) networks and IndicBERT, a transformer-based language model that has been trained on many Indian languages. Researcher test the suggested methods on a carefully chosen collection of Gujarati sentences to see how well they can fix spelling mistakes by checking surrounded circumstances and context of the sentence. The model that combined Peter Norvig's method with IndicBERT - GUJBRIJAPU consistently did better than the GRU GUJAPUBRIJ model on all important measures, including as accuracy (93.49%), precision (94.46%), recall (90.13%), and F1-score (91.59%).

These results reveal that GUJBRIJAPU - IndicBERT is considerably better at recognizing context and fixing language inputs that are complicated and have a lot of different forms. This makes it particularly reliable for real-world NLP tasks that use Gujarati text. The comparative research also demonstrates that Jodani and Sarash are good for simple spelling correction tasks that do not require a lot of computing power, but they do not have the grammatical knowledge and contextual depth needed for high-accuracy applications. The IndicBERT-enhanced GUJBRIJAPU model, on the other hand, is particularly good at recognizing little problems in context, minimizing false positives, and recommending changes that make sense. This makes it a perfect candidate for things like smart writing assistants, educational software, government documentation systems, and Gujarati-language chatbots, where language correctness is highly crucial.The research also gives us not just technically sound models, but also a helpful annotated dataset and a scalable architectural framework that may be utilized for a number of Indic languages with little effort. Future research may concentrate on enhancing the efficiency of the IndicBERT-based model by exploring lighter transformer alternatives,

quantization techniques, or model distillation. Also, more work may be done to add other dialects and make changes for certain fields. This will make the model even more usable in more circumstances. This research work establishes a robust foundation for the growth of intelligent spelling checkers for the Gujarati language that possess environmental awareness. It is a huge step forward in building NLP tools for low-resource languages since it connects simple rule-based systems with more complex deep learning systems. This will have long-term impacts on keeping regional languages alive, making them easier to find, and include them in the digital world. As a result, all four research objectives are solidly based on the gaps identified from the literature studies.

Objective 1 :

To develop a dataset of very high quality that has both correct and incorrect Gujarati sentences. The suggested models will be trained on this dataset, and their performance will be measured using natural language processing metrics such as F1-score, recall, accuracy, and precision.

Achievement:

To develop a Gujarati language dataset for the study, data was sourced from publicly available resources provided by Gujarati Wikipedia, Gujarati Vishwakosh Trust, Bhagavadgomandal, Gujarati Lexicon, Ekatra Foundation, E-shabad by Cygnet Infotech and Navjeevan Trust. Over 100,000 sentences were initially collected from these sources. After performing a thorough data cleaning and preprocessing process to ensure quality and relevance, a final dataset comprising 20,000 sentences was created. This refined dataset includes both correct and erroneous sentence pairs, creation it appropriate for responsibilities such as spell error correction and language model training.Thus, the first objective is justified because of development of required Gujarati language Data Set.

Objective 2 :

To create hybrid spell-checking models for Gujarati sentences that take into account the spelling's context and surroundings. These models will employ a mix of strategies to help people understand context and sequence better. Some of these algorithms are GRU, IndicBERT, and Peter Norvig's probabilistic spelling correction.

Achievement:

This objective is achieved by creating hybrid models of spell correction by combining the

probabilistic spell correction method proposed by Peter Norvig with state-of-the-art deep learning models such as Gated Recurrent Units (GRU) and IndicBERT. First, the Norvig algorithm identifies misspelled Gujarati words by comparing them with a specially compiled Gujarati dictionary and suggesting possible correction candidates based on edit distance operations. Probabilities are then assigned based on word frequency, making it possible to correct non-word spelling errors correctly. To make the model contextually aware, the GRU and IndicBERT models are used. The GRU model identifies sequential patterns by analyzing word patterns and understanding the relationships between adjacent words, thus identifying the most contextually appropriate correction. At the same time, IndicBERT, a pre-trained language model specifically designed for Indian languages, evaluates the suggested corrections based on contextual embeddings and semantic analysis at the sentence level. By combining probabilistic lexical correction and neural context analysis, the proposed hybrid model ensures accurate and contextually aware spell correction for Gujarati text, thus meeting the requirements of Objective 2.

Objective 3 :

To optimize the proposed models through hyperparameter tuning, ensuring improved performance on contextually complex linguistic inputs.

Achievement:

The third objective is achieved by the optimization of both hybrid models, GUJAPUBRIJ (which combines Peter Norvig's approach with GRU) and GUJBRIJAPU (which combines Peter Norvig's approach with IndicBERT), through the use of systematic hyperparameter optimization techniques. In the GRU-based GUJAPUBRIJ model, Random Search as implemented in Keras Tuner is used to explore the hyperparameter space efficiently, considering parameters such as embedding sizes, numbers of GRU units, dropout values, and learning rates. Model selection is made based on validation accuracy and F1-score, which helps to achieve a balance between model complexity and generalization performance while avoiding overfitting. For the IndicBERT-based GUJBRIJAPU model, fine-tuning is performed in a regression-based sentence scoring setting. Key parameters such as learning rate, batch size, number of epochs, and weight decay are optimized carefully using the AdamW optimizer. The use of probability-based sentence labels helps the model to better capture nuanced differences in context and grammar. The above optimization strategies help to improve the accuracy, reliability, and robustness of both models in handling contextually complex Gujarati linguistic inputs, thus successfully achieving the third objective.

Objective 4 :

To conduct a comprehensive comparative analysis between the proposed models and existing spell-checking systems to evaluate their effectiveness and accuracy.

Achievement:

The fourth objective is achieved through a comprehensive comparative analysis of the proposed hybrid models, GUJAPUBRIJ and GUJBRIJAPU, with the existing Gujarati spell-checking models, Jodani and Saras. The comparative study uses the conventional performance metrics of accuracy, precision, recall, and F1-score. Based on the experimental results, it is found that the GUJBRIJAPU model performs better with 93.49% accuracy, 94.46% precision, 90.13% recall, and an F1-score of 91.59%, outperforming Jodani with 91.56% accuracy and Saras with 64.57% accuracy. The comparative analysis also investigates the architectural differences between the models. While Jodani and Saras are largely dependent on rule-based approaches with less emphasis on contextual understanding, the proposed models incorporate deep learning approaches, namely GRU and IndicBERT, to better understand the semantic and contextual relationships between words. This architectural superiority enables the proposed models to provide more precise corrections for non-word errors and contextually incorrect real-word errors. Based on the performance comparison and architectural analysis, it can be concluded that the proposed hybrid model GUJBRIJAPU (Norvig + IndicBert) provide improved accuracy, better contextual understanding, and better overall functionality, thus successfully meeting the fourth objective.

Research Contribution

This work contributes to the Gujarati Natural Language Processing community by constructing and testing contextually informed hybrid spell-checking models for a morphologically complex and resource-poor language. A large dataset of 20,000 carefully crafted Gujarati sentences, including linguistically correct as well as deliberately erroneous examples, was developed, thus providing a useful resource for future research. Two new hybrid models, GUJAPUBRIJ and GUJBRIJAPU, are proposed. GUJAPUBRIJ is a combination of Peter Norvig's probabilistic approach and GRU learning, while GUJBRIJAPU is a combination of Peter Norvig's approach and IndicBERT for improved contextual understanding. These models combine the strengths of lexical probability and deep contextual understanding for better correction accuracy. Using hyperparameter optimization and careful evaluation techniques, the proposed models have shown excellent performance, with GUJBRIJAPU performing exceptionally well in terms of accuracy and contextual understanding. A comprehensive comparison with traditional systems such as Jodani and Saras has highlighted a paradigm shift from rule-based word-level correction to semantically aware sentence-level correction. In addition to its contributions, this work provides a template for application to other Indic languages, thus facilitating the overall goals of digital inclusion, preservation of linguistic heritage, and intelligent language processing for real-world NLP applications.

FUTUREWORK

The IndicBERT-based model (GUJBRIJAPU) works better when it comes to accuracy and comprehending context, but it needs a lot of computer power to do it. This makes it rigid to use in actual period on low-end devices. The existing models are trained on standard Gujarati and do not specifically include regional dialects or spoken variations. This could make them less useful in real-world situations. Using pretrained models like IndicBERT limits the ability to customize for Gujarati-specific subtleties unless further pretraining is conducted on domain-specific corpora. Even while models are better at comprehending context, they still have trouble with homophones and statements that are grammatically valid but do not make sense in context. This technique also does not find mistakes in punctuation. These constraints create opportunities for the researcher in the future. Future research will focus on reducing computational overhead through model distillation, quantization, or lighter transformer architectures to enable deployment on mobile and embedded devices. Embedding the proposed models into real-world systems such as Gujarati word processors, educational apps, and smart keyboards for real-time spelling and grammar correction. Enhancing the system to handle code-mixed inputs (e.g., Gujarati-English) and developing multilingual support for related Indo-Aryan languages. Developing fine-grained error typologies to better distinguish between phonetic, morphological, and syntactic errors, and tailoring correction strategies accordingly.

BIBLIOGRAPHY

1. બ. પંચાલ અને અ. શાહ, "ગુજરાતી ભાષાની ટેક્નોલૉજી યુગમાં ભૂમિકા: આર્ટિફિશિયલ ઇન્ટેલિજન્સના સંદર્ભમાં," *VIDYA – A Journal of Gujarat University*, vol. 4, no. 2, pp. 397–402, 2025.
2. B. Y. Panchal and A. Shah, "NLP Research: A Historical Survey and Current Trends in Global, Indic, and Gujarati Languages," in *4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, 2024.
3. N. S. Bhirud, "Grammar Checkers for Natural Languages: A Review," *International Journal on Natural Language Computing (IJNLC)*, vol. 6, 2017.
4. N. G. Patel and D. D. B. Patel, "Research Review of Rule-Based Gujarati Grammar Implementation with the Concepts of NLP," *Journal of Emerging Technologies and Innovative Research (JETIR)*, vol. 5, no. 9, 2018.
5. N. P. Desai and V. K. Dabhi, "Resources and Components for Gujarati NLP Systems: A Survey," *Artificial Intelligence Review*, vol. 55, pp. 1–19, 2022.
6. S. Singh and S. Singh, "HINDIA: A Deep Learning-Based Model for Spell Checking of Hindi Language," *Neural Computing and Applications*, vol. 33, no. 8, pp. 3825–3840, 2021.
7. B. Y. Panchal and A. Shah, "A Survey on Gujarati NLP Research Work," *Aibi Research, Management and Engineering Journal*, vol. 13, no. 1, pp. 234–249, 2025.
8. W. Frawley, *International Encyclopedia of Linguistics*, Oxford University Press, 2003.
9. T. Vyas and A. Ganatra, "Gujarati Language: Research Issues, Resources and Proposed Method on Word Sense Disambiguation," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 2, pp. 2277–3878, 2019.
10. B. Waghmar, *Concise Encyclopedia of the Languages of the World*, Elsevier, 2009.
11. M. Gokani and R. Mamidi, "GSAC: A Gujarati Sentiment Analysis Corpus from Twitter," in *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, Association for Computational Linguistics, 2023.
12. J. Baxi and B. Bhatt, "GujMORPH: A Dataset for Creating Gujarati Morphological Analyzer," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC)*, 2022.
13. B. Thakur et al., *Gujarati: A Textbook for Learning Gujarati Through Hindi*, Central Institute of Indian Languages.
14. W. Tisdall, *A Simplified Grammar of the Gujarati Language*, London: Kegan Paul, Trench, Trübner, 1892.
15. P. J. Mistry, "Gujarati Writing," in *The World's Writing Systems*, Oxford University Press, pp. 391–394, 1996.
16. G. Cardona, *A Gujarati Reference Grammar*, University of Pennsylvania Press, 1965.
17. "Gujarati," Languages Gulper. [Online]. Available: https://www.languagesgulper.com/eng/Gujarati.html. Accessed 2025.
18. N. Patel and D. Patel, "Implementation Approach of Gujarati Grammar 'Sandhi' Using Rule-Based NLP," in *8th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2021.
19. B. Suthar, "A Brief Outline of Gujarati Parts-of-Speech," University of Pennsylvania, 2003.
20. R. Soni, *Gujarati Lekhan-Paddhati*, Gurjar, 2019.
21. Y. Vyas, *Gujarati Bhashaunun Vyakran*, Balvinod Prakashan, 2018.
22. R. B. K. P. Trivedi, *Higher Grammar of the Gujarati Language*, Macmillan, 1919.

23. A. S. Bhuva and D. Mishra, "Gujarati Optical Character Recognition Using Efficient Text Feature Extraction Approaches," *Informatica*, vol. 49, no. 28, 2025.

24. A. A. Desai, "Gujarati Handwritten Numeral Recognition Through Neural Network," *Pattern Recognition*, vol. 43, no. 7, pp. 2582–2589, 2010.

25. S. Antani and L. Agnihotri, "Gujarati Character Recognition," in *Proc. ICDAR*, Bangalore, India, 1999.

26. C. P. B. Tailor, "Chunker for Gujarati Language Using Hybrid Approach," in *Advances in Intelligent Systems and Computing*, 2021.

27. K. Suba et al., "Hybrid Inflectional Stemmer and Rule-Based Derivational Stemmer for Gujarati," in *Proceedings of WSSANLP*, 2011.

28. H. Patel, B. Patel, and K. Lad, "Jodani: A Spell Checking and Suggesting Tool for Gujarati Language," in *11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021.

29. B. K. Y. Panchal and A. Shah, "Spell Checker Using Norvig Algorithm for Gujarati Language," in *International Conference on Smart Data Intelligence*, Singapore, 2024.

30. T. A. Gal, "Natural Language Processing (NLP) Pipeline," *Medium*, Oct. 23, 2023.

31. P. Patel, K. Popat, and P. Bhattacharyya, "Hybrid Stemmer for Gujarati," in *Proceedings of the 1st Workshop on South and Southeast Asian NLP*, 2010.

32. M. Parikh and A. Desai, "Recognition of Handwritten Gujarati Conjuncts Using CNN Architectures," in *ICACDS 2022*, 2022.

33. B. Y. Panchal and A. Shah, "NLP-Based Spellchecker and Grammar Checker for Indic Languages," in *Natural Language Processing for Software Engineering*, Scrivener Publishing, pp. 43–70, 2025.

34. C. Tailor and B. Patel, "Sentence Tokenization Using Statistical and Rule-Based Approach for Gujarati," in *Advances in Intelligent Systems and Computing*, 2018.

35. S. Sooraj et al., "Deep Learning Based Spell Checker for Malayalam," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1427–1434, 2018.

36. S. Murugan et al., "Symspell and LSTM Based Spell-Checkers for Tamil," in *Tamil Internet Conference*, 2020.

37. N. Hossain et al., "Panini: A Transformer-Based Grammatical Error Correction Method for Bangla," *Neural Computing and Applications*, vol. 36, pp. 3463–3477, 2024.

38. R. Phukan et al., "Deep Learning Based Approach for Spelling Error Detection in Assamese," in *ICCCNT*, 2023.

39. J. L. Peterson, "Computer Programs for Detecting and Correcting Spelling Errors," *Communications of the ACM*, vol. 23, no. 12, pp. 676–687, 1980.

40. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

41. E. Mays, F. J. Damerau, and R. L. Mercer, "Context Based Spelling Correction," *Information Processing & Management*, vol. 27, no. 5, pp. 517–522, 1991.

42. T. Brants et al., "Large Language Models in Machine Translation," in *EMNLP-CoNLL*, 2007.

43. A. R. Golding and D. Roth, "A Winnow-Based Approach to Context-Sensitive Spelling Correction," *Machine Learning*, vol. 34, no. 1, pp. 107–130, 1999.

44. G. Wilcox-O'Hearn and A. Budanitsky, "Real-Word Spelling Correction with Trigrams," in *International Conference on Intelligent Text Processing*, 2008.

45. S. Deode et al., "L3Cube-IndicSBERT," arXiv preprint arXiv:2304.11434, 2023.

46. J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *NAACL-HLT*, 2019.

47. D. Kakwani et al., "IndicNLPSuite," in *Findings of EMNLP*, pp. 4948–4961, 2020.

48. D. Pruthi et al., "Combating Adversarial Misspellings with Robust Word Recognition," arXiv:1905.11268, 2019.

49. G. Hirst and A. Budanitsky, "Correcting Real-Word Spelling Errors by Restoring Lexical Cohesion," *Natural Language Engineering*, vol. 11, no. 1, pp. 87–111, 2005.

50. K. Shaalan et al., "Analyzing and Correcting Spelling Errors for Non-Native Arabic Learners," in *INFOS*, 2010.

51. K. Cho et al., "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches," arXiv:1409.1259, 2014.

52. J. Chung et al., "Empirical Evaluation of Gated Recurrent Neural Networks," arXiv:1412.3555, 2014.

53. Z. Lan et al., "ALBERT: A Lite BERT," arXiv:1909.11942, 2019.

54. K. S. Jones, "Natural Language Processing: A Historical Review," 1994.

55. P. Johri et al., "Natural Language Processing: History, Evolution, Application, and Future Work," in *ICCIN 2020*, 2021.

56. J. Léon, "Early Machine Translation," 2014.

57. J. Hutchins, "Two Precursors of Machine Translation," *International Journal of Translation*, vol. 16, no. 1, pp. 11–31, 2004.

58. A. M. Turing, "Computing Machinery and Intelligence," 2007.

59. N. Chomsky, *Syntactic Structures*, Walter de Gruyter, 2002.

60. G. L. Steele, "Report on the 1980 LISP Conference," *ACM SIGPLAN Notices*, vol. 17, no. 3, pp. 22–36, 1982.

61. S. S. Jamwal and P. Gupta, "Hybrid Approach for Dogri Spell Checker," in *IDEA 2021*, 2022.

62. M. Das et al., "Design and Implementation of a Spell Checker for Assamese," in *Language Engineering Conference*, 2002.

63. S. Iqbal et al., "Urdu Spell Checking: Reverse Edit Distance Approach," in *WSSANLP*, 2013.

64. A. A. Lawaye and B. S. Purkayastha, "Kashmiri Spell Checker and Suggestion System," *The Communications*, vol. 21, no. 2, p. 123, 2012.

65. B. Kaur and H. Singh, "HINSPELL—Hindi Spell Checker Using Hybrid Approach," *International Journal of Scientific Research and Management*, vol. 3, no. 2, pp. 2058–2062, 2015.

66. R. Sankaravelayuthan, "Spell and Grammar Checker for Tamil," 2015.

67. A. A. Lawaye and B. S. Purkayastha, "Design and Implementation of Spell Checker for Kashmiri," *International Journal of Scientific Research*, vol. 5, no. 7, 2016.

68. R. Sakuntharaj and S. Mahesan, "Hybrid Approach to Detect and Correct Spelling in Tamil," in *ICIAfS*, 2016.

69. U. M. G. Rao et al., "Telugu Spell-Checker," *Vaagartha*, 2012.

70. S. Saha et al., "Bangla Spell Checker and Suggestion Generator," 2019.

71. S. Singh and S. Singh, "Systematic Review of Spell-Checkers for Highly Inflectional Languages," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4051–4092, 2020.

72. B. Bhagat and M. Dua, "Enhancing Performance of Gujarati ASR Using Improved Spell Corrector," in *ITM Web of Conferences*, 2023.

73. B. Y. Panchal and A. Shah, "Gujarati Spell Checker Using Norvig Algorithm with Grammar Rules," *AiBi Revista de Investigación*, vol. 13, no. 2, pp. 234–249, 2025.

74. S. Khanuja et al., “MuRIL: Multilingual Representations for Indian Languages,” arXiv:2103.10730, 2021.

75. A. Conneau et al., “Unsupervised Cross-Lingual Representation Learning,” arXiv:1911.02116, 2019.

76. J. Pfeiffer et al., “AdapterHub,” arXiv:2007.07779, 2020.

77. M. Nejja and A. Yousfi, “The Context in Automatic Spell Correction,” *Procedia Computer Science*, vol. 73, pp. 109–114, 2015.

78. A. K. Ingason et al., “Context-Sensitive Spelling Correction and Rich Morphology,” in *NODALIDA*, 2009.

79. A. Yunus and M. Masum, “Context-Free Spell Correction Using Supervised ML,” *International Journal of Computer Applications*, vol. 176, no. 27, pp. 36–41, 2020.

80. P. Gupta, “Context-Sensitive Real-Time Spell Checker,” in *IEEE ICSC*, 2020.

81. J. Sheth and B. C. Patel, “Gujarati Phonetics and Levenshtein Based String Similarity,” in *5th National Conference on Indian Language Computing*, 2015.

82. B. Y. Panchal and A. Shah, “Hybrid Context Aware Gujarati Spell Correction Using Norvig Algorithm, GRU, and IndicBERT,” *Informatica*, vol. 49, no. 34, pp. 427–442, 2025. DOI: https://doi.org/10.31449/inf.v49i34.9836