**Literature Review**

**[13]**

**A Semi‑automated Approach for Bengali Neologism**

In 2023, Apurbalal Senapati developed Lexiconovation: Navigating the Future of Bengali Language with Cutting-Edge NLP. Researchers historically relied on manual analysis of literature, newspapers, and technical texts for neologism identification, a time-consuming process. The advent of digital technology led to automated techniques, such as web-as-corpus, news corpora and Wikipedia. The CORD19 dataset and Canadian government glossary serve as extensive resources during the COVID-19 pandemic.Various models, including morphological analysis, linguistic perspective exploration, medical domain reflection, and social media analysis, have been employed to study COVID-19 neologisms. Utilized support vector machines for French neologism detection. Studies have categorized and analyzed COVID-19 neologisms, shedding light on grammatical aspects, linguistic relationships, and societal reflections. Automated methods enable efficient neologism extraction from vast datasets, offering insights into language evolution during significant events like the COVID-19 pandemic.Challenges include noisy data, informal language, and misspellings in web sources, necessitating preprocessing techniques. Transferability to Indic languages, like Bengali, remains a challenge.

**Cit--->** Senapati, A. (2023). A Semi-automated Approach for Bengali Neologism. SN Computer Science, 4(5), 428.

**[14]**

**Multilingual Personalised Hashtag Recommendation for Low Resource Indic Languages using Graph-based Deep Neural Network**

In 2023, Shubhi Bansal et al. unveiled TAGALOG, a groundbreaking system addressing the void in hashtag recommendation for low-resource Indic languages. Leveraging user-guided and language-guided attention mechanisms alongside a graph-based neural network, TAGALOG exhibited a remarkable F1-score improvement of 12.3% and 12.8%, surpassing pre-trained models. The introduction of the Indic Hash dataset, featuring tweets across diverse languages, enriched the research. While TAGALOG demonstrated exceptional performance in low-resource settings, challenges persist in adapting to evolving user interests, prompting the need for further exploration in broader linguistic contexts. The study marks a significant stride in personalized hashtag recommendations for multilingual engagement and content discoverability.

**Cit--->** Bansal, S., Gowda, K., & Kumar, N. (2024). Multilingual personalized hashtag recommendation for low resource Indic languages using graph-based deep neural network. Expert Systems with Applications, 236, 121188.

**[15]**

**IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages**

In 2023, Jay Gala et al. introduced "IndicTrans2," a groundbreaking initiative addressing the linguistic diversity of India. The paper emphasizes the absence of robust translation models for all 22 scheduled Indian languages, hindering effective communication. IndicTrans2 boasts key contributions: the extensive Bharat Parallel Corpus Collection (BPCC), the first n-way parallel benchmark (IN22), and the development of the first translation model supporting all 22 languages. With a focus on societal impact, especially in governance, legal proceedings, and education, the paper envisions improved accessibility and quality in translation models. IndicTrans2, supported by BPCC, achieves commendable performance, including IndicTrans2-M2M for 462 translation directions, promoting linguistic inclusivity and fostering open collaboration while facing challenges in user interest evolution and commercial integration.

**Cit--->** Gala, J., Chitale, P. A., AK, R., Doddapaneni, S., Gumma, V., Kumar, A., ... & Kunchukuttan, A. (2023). IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages. arXiv preprint arXiv:2305.16307.

**[16]**

**LIDOMA@DravidianLangTech: Convolutional Neural Networks for Studying Correlation Between Lexical Features and Sentiment Polarity in Tamil and Tulu Languages**

In 2023, Shahiki-Tash et al. presented a pioneering study on Sentiment Polarity Detection in Code-Mixed Texts, focusing on Dravidian languages like Tamil and Tulu. Leveraging datasets from DravidianLangTech, the research provides a gold standard corpus for sentiment analysis in code-mixed social media content. Overcoming class imbalance challenges, the approach integrates lexical features and a meticulously tuned 6-layered Convolutional Neural Network (CNN). Demonstrating robust performance, the model achieves notable macro F1 scores of 0.542 and 0.199 for Tulu-English and Tamil-English datasets, respectively. The combination of lexical insights and CNN architecture proves effective, addressing complexities in sentiment analysis within diverse linguistic contexts, despite acknowledging challenges related to class imbalance and evolving linguistic phenomena in social media communication.

**Cit--->** Tash, M., Armenta-Segura, J., Ahani, Z., Kolesnikova, O., Sidorov, G., & Gelbukh, A. (2023, September). Lidoma@ dravidianlangtech: Convolutional neural networks for studying correlation between lexical features and sentiment polarity in tamil and tulu languages. In Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages (pp. 180-185).