1. Developing Lexical Resources of Saraiki Verbs: A Corpus Based Study  
   In January 2024, Azher M., et al., published a research paper on developing Lexical Resources of Saraiki Verbs. There was no model used as the purpose of the project was to create a single corpus of data for the Saraiki Indic language. It was a valuable study that documented the efforts for compiling of the corpus. Various online and physical sources, including poems, dramas, letters, diaries etc were used in compiling the corpus. The research paper went on to analyse the Saraiki language verbs based on word roots and other linguistic properties of language. However there was no Machine Learning, Deep Learning or Natural Language Processing methods or concepts used in the paper as it seemed to be a purely linguistic study of the language and literature of the Sariaki lexicon.  
   Cite: Awais, M., Azher, M., & Arslan, M. F. (2023). Developing Lexical Resources of Saraiki Verbs: A Corpus Based Study . Linguistic Forum - A Journal of Linguistics, 5(3), 136–158. https://doi.org/10.53057/linfo/2023.5.3.10
2. Explicitly unsupervised statistical machine translation analysis on five Indian languages using automatic evaluation metrics  
   In February 2022, Saxena S. et al., published a research paper for an Unsupervised Machine Translation Model (UMNT) which used the NLTK toolkit of python and the BLEU assessment metric to create a Machine Translation model on the IndicNLPSuite dataset for English, Hindi, Tamil, Telugu and Kangri languages. This was a revolutionary paper as it was the first one to introduce a Unsupervised MT model. Even though the model used was an ordinary NLTK model, a new, unsupervised method of Machine Translation was developed by the authors. However the model was only rudimentary and only processed four out of 200 Indic languages. Despite these limitations, this is an important paper as it shows the possibilities of a completely unsupervised statistical machine translation system, capable of performing analysis on five Indic languages using automated evaluation metrics.  
   Cite: Saxena, S., Chauhan, S., Arora, P. et al. Explicitly unsupervised statistical machine translation analysis on five Indian languages using automatic evaluation metrics. Sādhanā 47, 106 (2022). https://doi.org/10.1007/s12046-022-01855-w
3. Improving Multilingual Neural Machine Translation System for Indic Languages  
   In June 2023, Das S.B. et al., published a paper that improved Multilingual Neural Machine Translation systems for Indic languages. The team used the Samanantar Corpus of data as a dataset, which contained over 49.6 million pairs of English and Indic language sentence pairs. The team created Multilingual Machine Translation Models for translating to and from the Indic Languages and English. These included many Indic languages, and was able to translate successfully to and from the two groups. The paper lists various examples of the Neural Machine Translation system translating sentences between the languages, with specific examples for specific languages. Each example is numbered according to the order of execution. The model was, however, unsuccessful in translating accurately for the Indic Languages Bengali, Gujarati, Hindi, Kannada, Malayalam, Odia and Punjabi. Despite that, it is an excellent development as it shows the possibility of developing a larger scale lexical analysis tool for Indic Languages.  
   Cite: Sudhansu Bala Das, Atharv Biradar, Tapas Kumar Mishra, and Bidyut Kr. Patra. 2023. Improving Multilingual Neural Machine Translation System for Indic Languages. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22, 6, Article 169 (June 2023), 24 pages. https://doi.org/10.1145/3587932
4. Towards Leaving No Indic Language Behind: Building Monolingual Corpora, Benchmark and Models for Indic Languages  
   In July 2023, Doddapaneni S. et al., published a paper for building a Monolingual Corpora, benchmark and models for Indic Languages. The IndicCOPA, IndicQA, IndicXParaphrase, IndicSentiment, IndicXNLI, Naamapadam, MASSIVE and FLORES datasets were used for training and testing. The models used were IndicBERT-MLM and IndicBERT-SS models, with the evaluation being IndicXTREME benchmark. The models use large text corpi, however it would take time to train, test and run the models.  
   Cite: Sumanth Doddapaneni, Rahul Aralikatte, Gowtham Ramesh, Shreya Goyal, Mitesh M. Khapra, Anoop Kunchukuttan, and Pratyush Kumar. 2023. Towards Leaving No Indic Language Behind: Building Monolingual Corpora, Benchmark and Models for Indic Languages. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 12402–12426, Toronto, Canada. Association for Computational Linguistics.