World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Review Article)

Check for updates

Computational linguistics at the crossroads: A comprehensive review of NLP advancements

Shahid Iqbal Rai [1], Maida Maqsood [2, *], Bushra Hanif [3], Muhammad Ali Adam [4], Muhammad Arslan [5], Hira Shafiq [6] and Muhammad Sijawal [7]

[1] Department of Enterprise Engineering, University of Rome Tor Vergata, Italy.
[2] Department of English, Government College Women University, Sialkot, Pakistan.
[3] Faculty of Asian and African Studies, Guangdong University of Foreign Studies, 510420, Guangzhou, Guangdong, China.
[4] MS Computer Science, College of Computing and Digital Media (CDM), Depaul University, Chicago, United States of America.
[5] School of Computer Science Chenab College of Advance Studies Faisalabad, Pakistan.
[6] Department of Computer science, Institute of Southern Punjab, Multan, Pakistan.
[7] Deputy Hostel Provost, Liaquat University of Medical & Health Sciences Jamshoro, Pakistan.

World Journal of Advanced Engineering Technology and Sciences, 2024, 11(02), 578–591

Publication history: Received on 11 March 2024; revised on 20 April 2024; accepted on 23 April 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.11.2.0146

Abstract

New NLP breakthroughs have put Computational Linguistics at a crossroads. NLP's past, present, and future are covered. This review explains computational linguistics' creation with a brief history of linguistics and computer science. Early solutions processed and understood natural language using rule-based systems using manually constructed linguistic rules. Over time, these tactics became increasingly problematic as language became more complex and obscure. Statistical approaches transformed operations. Neural network-based machine learning methods are leading the area because they can learn complicated patterns and representations from large text collections. A data-driven model revolution in natural language processing enhanced language modelling, machine translation, and sentiment analysis. Next, NLP improvements for several tasks and applications are evaluated. Language understanding models that capture semantic nuances and contextual relationships use deep learning frameworks. Word embeddings and transformer-based architectures like GPT and BERT perform well on benchmark datasets for text classification, question answering, and named item identification. The paper also shows how NLP interacts with computer vision, voice processing, and other domains to show the merits and cons of cross-disciplinary research. Multimodal techniques that combine text, graphics, and audio may increase natural language processing and interpretation. The review discusses NLP's effects on prejudice, justice, and privacy. Responsible development and implementation are needed when NLP technology becomes widespread due to algorithmic bias and data privacy concerns. NLP research directions and concerns are reviewed. Existing models may meet standards but fail in practice.

Keywords: Computational Linguistics; Machine Learning; Language; Natural Language Processing; Cognitive Psychology

1. Introduction

At these crossroads of language study and technology growth, computational linguistics will advance in discovery and creativity. This review, "Computational Linguistics at the Crossroads: A Comprehensive Review of NLP Advancements," examines Natural Language Processing (NLP) to determine its importance, breadth, and goals in shaping language technology (Abdallah et al., 2024).

* Corresponding author: Maida Maqsood

1.1. Overview of Computational Linguistics

The rapidly growing subject of computational linguistics (natural language processing) studies, understands and generates human language using linguistics, computer science, artificial intelligence, and cognitive psychology. Linguists want computers to read and analyze natural language like humans. Computing methods for studying language patterns and structures in the mid-20th century created computational linguistics. Machine learning, neural networks, and massive data processing have accelerated growth. Computational linguistics includes machine translation, speech recognition, sentiment analysis, discourse modelling, and syntactic analysis. Statistical models, rule-based systems, and deep learning architectures analyze language data well in computer linguistics. Dialogue systems, virtual assistants, language teaching platforms, information retrieval, and text mining use computational linguistics (Smidt et al., 2024).

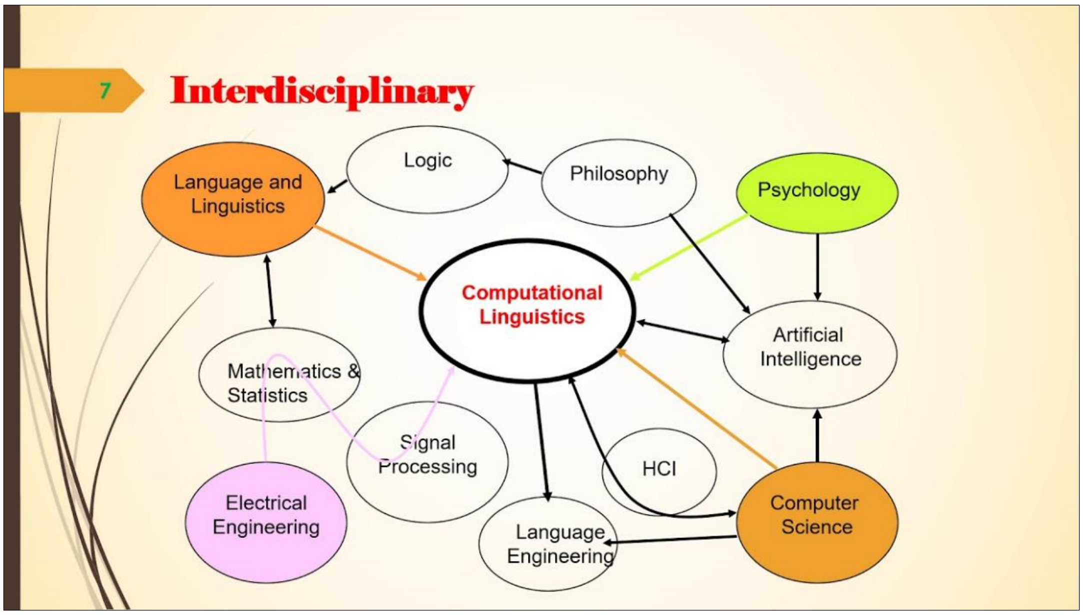

Figure 1 Introduction to Computational Linguistics (Shokhrukh & Abror, 2022)

1.2. Significance of NLP Advancements

Natural language processing (NLP) has transformed communication, information processing, and human-machine interaction. Intelligent systems must recognize and respond to natural language inputs for human-technology collaboration. Chatbots, virtual assistants, and intelligent agents enhance accessibility, efficiency, and comfort. Deciphering vast amounts of unstructured content online and in other digital archives requires natural language processing improvements. NLP analyzes text, measures user attitude, and extracts relevant data to provide companies with an edge. Due to advances in natural language processing, machine translation allows fluent linguistic communication. Transformer architectures and neural machine translation models improved automated translation quality and fluency, increasing worldwide collaboration and understanding.

1.3. Scope and Objectives of the Review

A comprehensive review of Computational Linguistics and Natural Language Processing including current trends, problems, and advances. This review highlights key studies, unique methods, and practical applications to explain natural language processing. Computational linguistics review critically analyzes and synthesizes significant discoveries, methodologies, and prospects. This study helps organize field research and technical breakthroughs by identifying new research trends, unsolved problems, and multidisciplinary cooperation prospects. Natural Language Processing is constantly changing, but "Computational Linguistics at the Crossroads: A Comprehensive Review of NLP Advancements" helps scholars, practitioners, and enthusiasts.

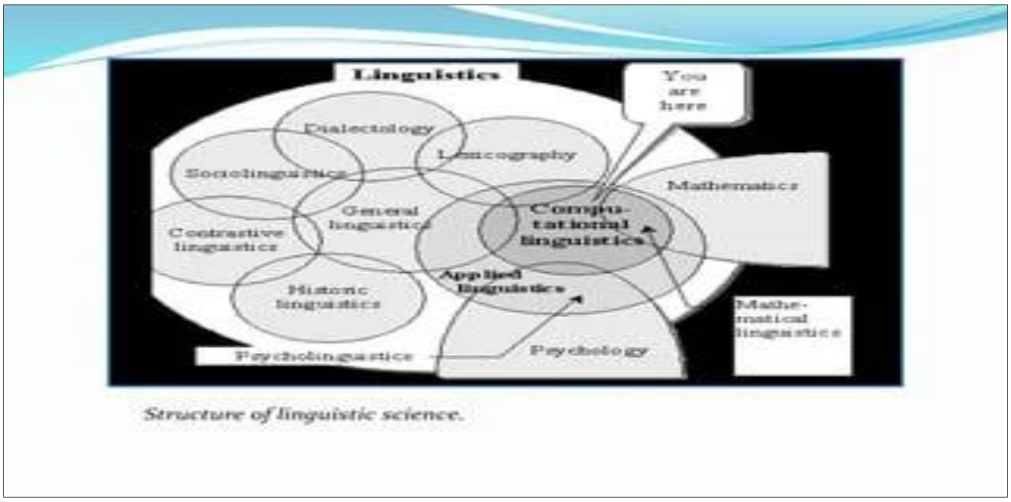

Figure 2 Computational Linguistics: Models, Resources, Applications (Abdallah et al., 2024)

2. Foundations of Computational Linguistics

2.1. Historical Development

In its interesting history, computer linguistics has progressed with linguistics and computing. Computer science, mathematics, and languages have intertwined since the mid-20th century when computers were introduced. WWII machine translation spawned computational linguistics. George Zipf and Warren Weaver tested automatic translation with basic computer approaches. NLP improvements were enabled by them. Noam Chomsky's formal language theory advanced the area throughout the 1950s and 1960s. Chomsky's generative grammar theory explained language structure and inspired computational models. Early NLP systems like the Georgetown-IBM Experiment translated Russian to English and were groundbreaking (Shokhrukh & Abror, 2022). These systems interpreted and produced human speech using rule-based and custom language rules (Abdallah et al., 2024).

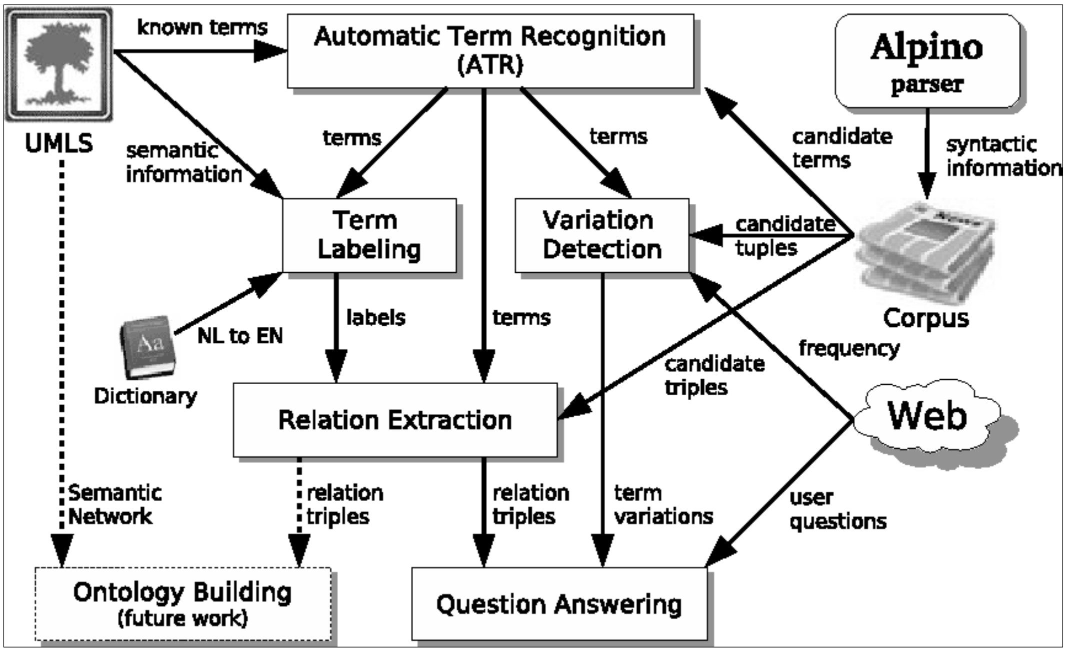

Figure 3 Computational Linguistics and History of Science (Lenci & Padó, 2022)

Massive corpora and machine learning algorithms moved NLP toward statistics in the 1970s. Researchers improved machine translation, speech recognition, and parsing with n-gram language models and hidden Markov models. Statistics and symbols were used to develop advanced NLP systems in the 1980s and 1990s. The Penn Treebank and

DARPA-supported SUR advanced speech understanding. Probabilistic parsing and machine translation are based on these pioneering works. Modern deep learning has altered natural language processing (NLP), allowing neural network models to surpass older methods in numerous applications (Becker et al., 2023).

2.2. Basic Concepts and Terminology

Computational linguistics relies on natural language processing principles and terminology. All aspects of language processing, knowledge representation, and human language structure are covered. The hierarchical structuring of words, phrases, and sentences is the focus of computational linguistics. Computers construct phrase structures with dependency networks and syntax trees to understand natural language. Language elements like semantic roles, syntactic dependencies, and part-of-speech tags matter. Features are used in natural language processing to analyze text, identify entities, and determine sentiment (Shokhrukh & Abror, 2022).

Many algorithms and approaches process natural language in computational linguistics. Rule-based methods evaluate and generate text using bespoke language rules, whereas statistical methods detect patterns from data using probabilistic models and machine learning algorithms. Due to its ability to replicate complicated linguistic processes, deep learning dominates computational linguistics. Transformers and RNNs produce excellent sentiment analysis, text synthesis, and language translation. Machine translation, speech recognition, dialogue systems, and information retrieval are computational linguistics. Each area has unique challenges and study opportunities, sparking discoveries (Smidt et al., 2024).

2.3. Key Milestones in NLP Research

NLP milestones have shaped the field and advanced language technology. These advances advanced computational linguistics theory, technique, and application. The first machine translation systems were developed in the 1950s, advancing natural language processing. The Georgetown-IBM Experiment demonstrated rule-based automatic translation for machine translation and cross-lingual communication. Natural language processing statistics changed in the 1970s. Hidden Markov models and n-gram language models by Frederick Jelinek and Martin Kay changed machine translation and speech recognition (Pirozelli & Câmara, 2022).

Natural language processing research needed corpus linguistics in the 1980s. Researchers trained and tested NLP systems using large annotated datasets from the Penn Treebank and Brown Corpus to build more accurate and reliable models. Text mining and information retrieval boomed in the 1990s as the internet and digital media emerged. Researchers employed text categorization, document clustering, and information extraction to analyze huge textual data. Google changed how people find and use information online in the early 2000s. Search engines are essential for knowledge discovery and information retrieval because they use advanced NLP algorithms to analyze user queries and find relevant web resources (Becker et al., 2023).

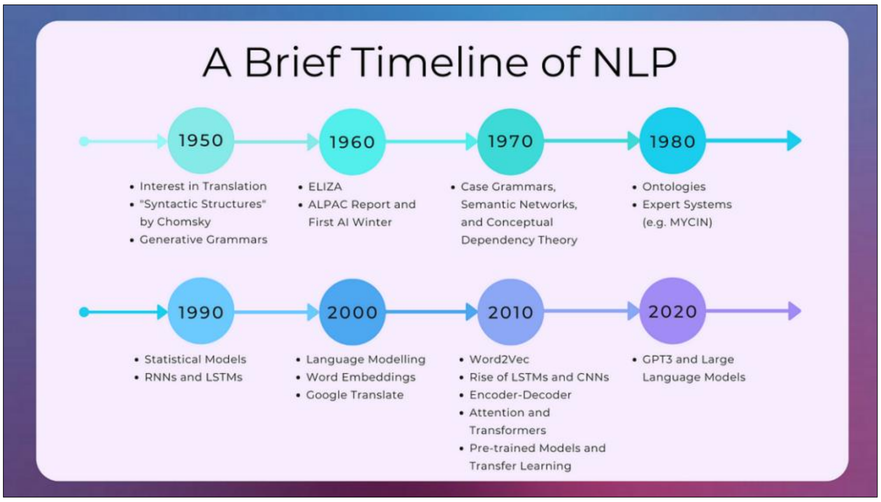

Figure 4 A Brief Timeline of NLP (Ive, 2022)

3. Core Techniques in Natural Language Processing

Computational linguistics is reshaping NLP. Statistical Methods, Machine Learning Approaches, Deep Learning Models, and Rule-based Systems are the main NLP advancement methods examined in this big overview study. These methods have helped natural language processing (NLP) flourish with their benefits and insights (Lenci & Padó, 2022).

3.1. Statistical Methods

NLP has long used statistical methods to express probability and patterns in language. Statistic NLP trains and analyzes big text corpora. In the past, HMMs and n-gram models used statistics to deduce text patterns. Language modelling, named entity identification, and part-of-speech tagging were superb. Statistical methods handle language variation and ambiguity well. This probability model represents the natural language's context dependence and ambiguity. However, complex linguistic demands make long-range and semantic relationships challenging to record (Pirozelli & Câmara, 2022).

Statistics has improved with algorithms and computers. Probabilistic graphical models like PCFGs and CRFs enable more complicated language structure investigations. Statistical NLP models are more adaptable and resilient thanks to Bayesian inference and maximum likelihood estimations. Data-driven methods, like deep learning, challenge statistical methods despite their benefits. Due to their reliance on handcrafted qualities and limited semantic representation capacity, new data-driven paradigms have emerged, though they are still useful (Ive, 2022).

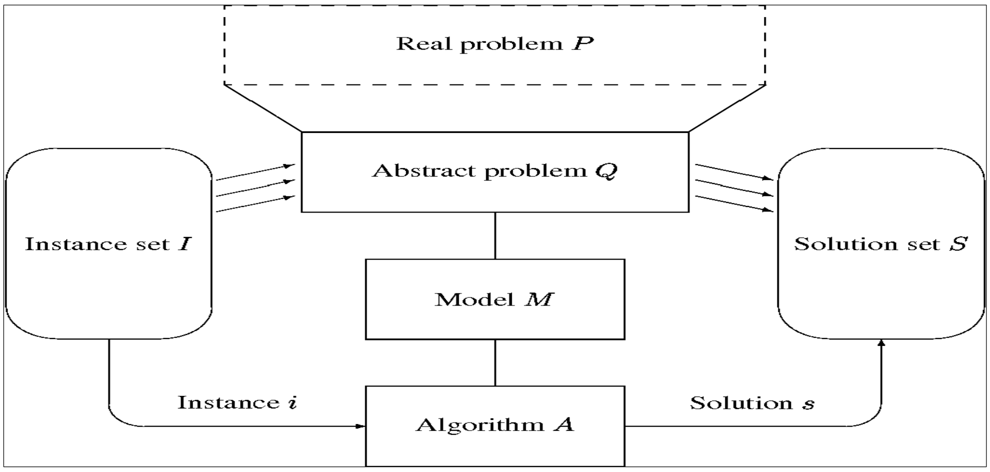

Figure 5 Statistical Methods in Natural Language Processing (Satibaldieva, 2024)

3.2. Machine Learning Approaches

Machine learning lets systems learn from data and improve performance, revolutionizing natural language processing. Random Forests and SVMs are important in sentiment analysis, text categorization, and machine translation. These algorithms anticipate outcomes from unlabeled data using patterns from labelled datasets. Natural language processing has improved with dimension reduction and unsupervised grouping. LDA and Word Embeddings (Word2Vec, GloVe) can find latent semantic structures in text data. Word association representations simplify document categorization, subject modelling, and semantic similarity (Oyewole et al., 2024).

To use labelled and unlabeled data better, active and semi-supervised learning arose. Iteratively picking informative annotation cases enhances model performance without human involvement. Machine learning can scale and adapt, so natural language processing (NLP) systems can handle different linguistic phenomena and vast data sets. For maximum benefit, feature engineering and parameter tweaking must be done properly. Also, black-box algorithms can limit interpretability and application (Satibaldieva, 2024).

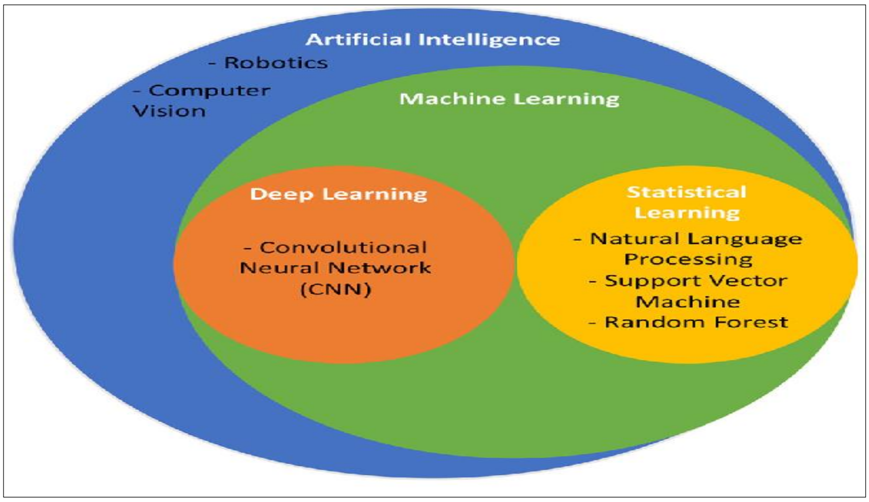

Figure 6 Artificial intelligence and machine learning (Satibaldieva, 2024)

3.3. Deep Learning Models

Deep learning has entered a new NLP era thanks to neural networks' hierarchical text representation. RNNs, CNNs, and transformer topologies boost NLP performance in numerous tasks. CNNs find text syntactic structures and local patterns using hierarchical convolutional layers. They excel at sentiment analysis, named entity identification, and text categorization. CNNs extract characteristics at multiple levels of abstraction to encode rich semantic information from raw text messages. RNNs better express a language's sequential dependencies and large contexts. LSTM and GRU networks improve language modelling, machine translation, and speech recognition (Lenci & Padó, 2022).

Random Neural Networks (RNNs) can survive sentence structure and word order changes since they store temporal dynamics. Transformers change natural language processing by enhancing global dependencies and contextual information acquisition through self-attention. GPT and BERT excel in NLP benchmarks for question answering, text production, and NLU. Deep learning models outperform and scale better than NLP. Their data requirements, computational complexity, and intricacy make them difficult to grasp. The research aims to resolve these challenges (Satibaldieva, 2024).

3.4. Rule-based Systems

Instead of natural language processing, rules-based systems assess texts using predefined linguistic patterns. These systems encode domain information and linguistic constraints with custom rules, lexicons, and grammars. Early rule-based uses of finite-state machines and regular expressions included tokenization, morphological analysis, and syntactic parsing. Rule-based systems are visible and interpretable, so developers can quickly fix logic. Language expertise and domain-specific limitations improve precision and resilience in confined domains (Oyewole et al., 2024).

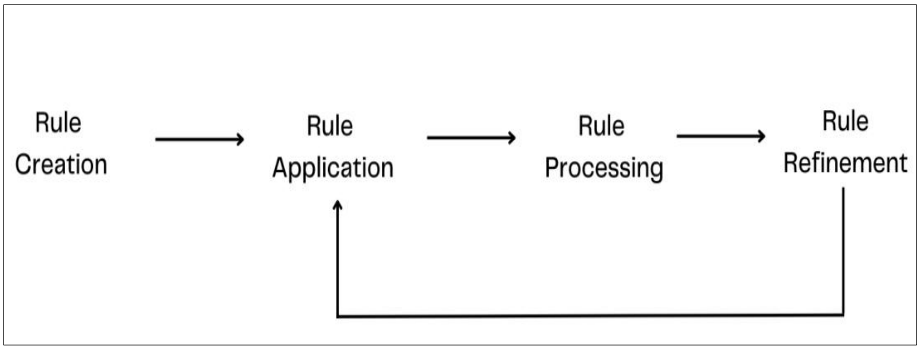

Figure 7 Rule-Based Approach in NLP (Satibaldieva, 2024)

Language variety and changing patterns challenge rule-based systems. Rules can be laborious and error-prone to set and manage manually, especially in complicated and ever-changing linguistic circumstances. Finally, these systems need many rules and struggle with ambiguity. In recent years, rule-based machine learning hybrid approaches have grown.

This strategy uses both paradigms' benefits. Rule-based and statistical or neural network hybrid systems improve NLP precision and adaptability (Nilufar, 2024).

4. Language Representation and Understanding

4.1. Exploring the Pillars of NLP Advancements

NLP helps computers interpret and produce human language using precise and nuanced language representations and understanding. Word Embeddings, Contextualized Word Representations, Semantic Parsing, Syntax, and Grammar Analysis are examined in detail. The importance, methodologies, and contemporary breakthroughs are explained (Nahli et al., 2022).

4.2. Word Embeddings

Syntactic patterns and semantic linkages were captured as dense, continuous vectors in a high-dimensional space by the Word2Vec paradigm, revolutionizing NLP. Algorithms use semantic similarity to understand semantics and word relationships. GloVe and FastText revised word embeddings utilizing massive datasets for better accuracy. Recently developed contextualized word embeddings can learn from their environment and improve natural language understanding (Satibaldieva, 2024).

4.3. Contextualized Word Representations

Instead of static word embeddings, contextualized word representations show how context affects meaning. Elmo, GPT, and BERT pioneered contextualized embeddings with deep neural networks. By evaluating the complete input sequence, these models create context-aware representations that capture complex syntactic and semantic information. Transformers like BERT help contextualized embeddings succeed in sentiment analysis, machine translation, and NLP (Xudaybergenova et al., 2024).

4.4. Semantic Parsing

Unstructured speech is parsed into semantic graphs for NLP, which helps computers understand language. Classic semantic parsing systems used handwritten rules or statistical methods and were rarely generalized across linguistic occurrences. Neural architectures can directly map plain English utterances to meaning representations via deep learning, revolutionizing semantic parsing. Due to attention processes and graph-based methodologies, Seq2Seq excels at semantic parsing, information retrieval, dialogue systems, and question-answering (Yu et al., 2024).

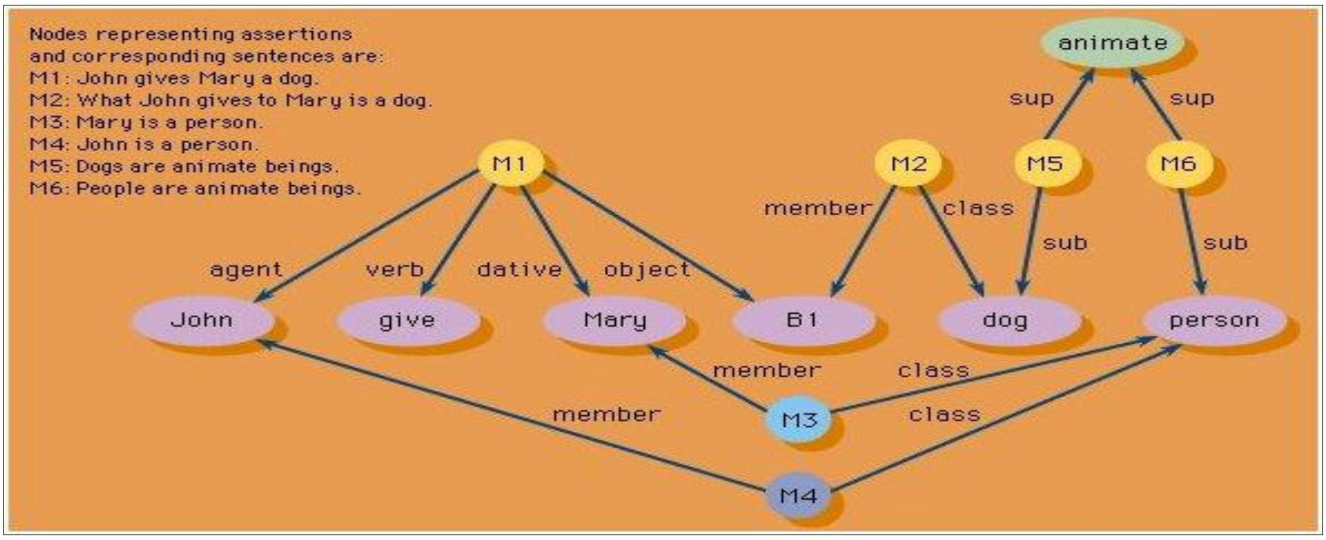

Figure 8 Working of semantic parsing work (Yenduri et al., 2023)

4.5. Syntax and Grammar Analysis

Language-processing computers grasp human language conventions and structures through syntax and grammar. In traditional syntax and grammar analysis, statistical models or rule-based parsers have ambiguity and linguistic variability difficulties. Deep learning improves this area by data-drivenly discovering syntactic patterns. Neural network-based methods increase dependence, constituency, and syntactic role labelling. Graph-based parsers and Transformer-based SyntaxNet have captured grammatical structures and syntactic linkages, enabling more sophisticated and precise language understanding systems (Xudaybergenova et al., 2024).

Advanced natural language processing (NLP) allows computers to represent and understand human language with remarkable precision and complexity. NLP's contextualized representations, semantic parsing, syntax and grammar analysis, and static word embeddings enhance language intelligence and adaptability. Where computational linguistics is studying verbal understanding, machines may conduct meaningful conversations with people, showing empathy and providing insightful analysis (Nahli et al., 2022).

5. Natural Language Generation

5.1. Text Generation Models

Text generation models are used in NLP applications like machine translation, chatbots, and content development. These models have progressed from rule-based to deep learning. Early text creation systems used manual templates and guidelines. Though competent for lesser tasks, these systems could not generate complicated language. As computer power rose and deep learning became prominent, researchers studied neural network-based text production techniques. RNNs are leading models in this field because they can sequentially recognize data dependencies. GRUs and LSM RNNs are used for language modelling and sequence generation. These models generate unified, contextually relevant data in amazing ways (Yu et al., 2024).

Transformer-based models became popular because RNNs exhibited vanishing gradients and long-range dependencies. Vaswani et al.'s Transformer altered text production. Transformers use self-attention to grasp input sequence global dependencies and produce more coherent and contextually rich text than RNNs. Transformer-based models aid text summarizing and translation. BERT and GPT perform well on NLP benchmarks. Due to their fine-tuned and pre-trained on massive text datasets, these models can readily create text for many jobs (Nilufar, 2024).

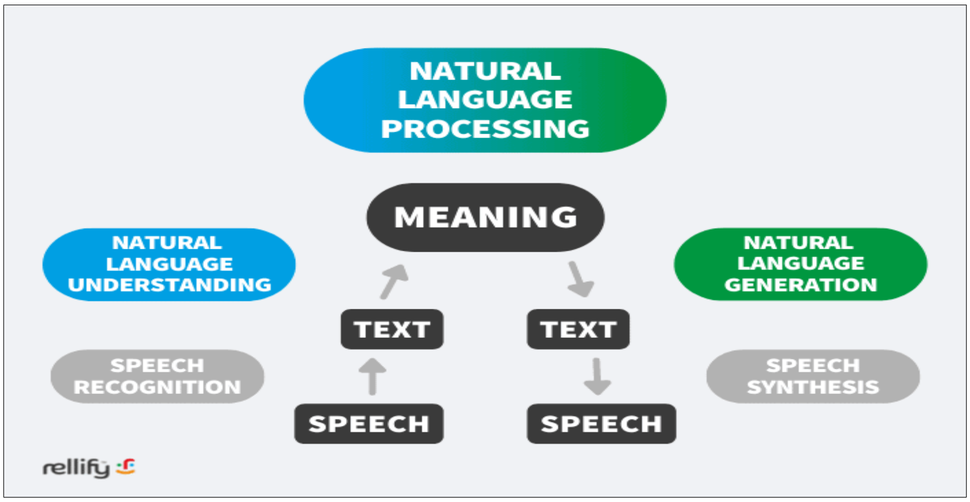

Figure 9 How Natural Language Generation Can Boost Your Business (Becker et al., 2023)

5.2. Dialogue Systems

Another prominent computational linguistics topic is dialogue systems or chatbots. These technologies allow virtual assistants, customer service bots, and educational tutors to have authentic, engaging conversations with humans. Initial dialogue systems used rule-based, hand-written scripts to reply to user input. While useful for rudimentary interactions, these systems struggled to interpret and produce actual language. Machine learning and deep learning were utilized to build complex dialogue systems. Natural language understanding (NLU), which derives meaning from user inputs, challenges dialogue systems. Natural language's intricacy and ambiguity challenged handwritten NLU algorithms (Yenduri et al., 2023).

Neural network models like sequence-to-sequence architectures and transformers are strong NLU tools since deep learning. Conversational systems now understand human input better. Dialogue systems provide coherent, context-relevant responses in multiple ways. Data-driven systems use reinforcement learning and sequence-to-sequence learning to answer input context, while rule-based systems use scripts or templates. Recently developed transformer topologies like GPT have enhanced text-generating model response quality and fluency (Zendaoui et al., 2023).

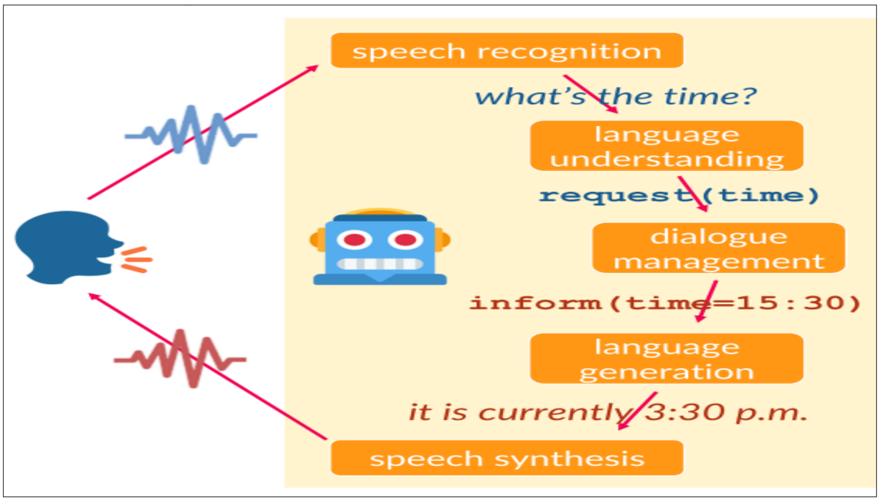

Figure 10 Dialogue systems (Zendaoui et al., 2023)

5.3. Summarization Techniques

Natural language processing uses text summarizing to extract crucial information from a document while keeping its structure and content. Extractive and abstractive summarizing have merits and downsides. Extractive summarizing employs source sentences or phrases to summarize. We score sentences by freshness, relevance, and importance using algorithms. Machine learning-based methods rank phrases by phrase length, word frequency, and semantic similarity; graph-based methods network sentences as nodes and edges to show sentence links. Both methods are used for extractive summarization (Choi & Lee, 2023).

Abstractive summary paraphrases and rephrases to make it more cohesive. This strategy requires the model to understand the input text's meaning and context to generate key phrases. Early rule-based and template-based abstractive summarizing models struggled to provide cohesive summaries. Transformers and sequence-to-sequence neural network models are suited for abstractive summarization now that deep learning is prevalent. Their summaries are more natural and context-appropriate. Despite advancements in abstractive summarizing, short and effective summaries that eliminate repetition and retain the text's spirit remain difficult. Long, complicated texts and sensitive phrases and language patterns challenge abstractive summarization methods (Legallois & Koch, 2020).

6. NLP Applications and Use Cases

6.1. Sentiment Analysis

Popular NLP applications include sentiment analysis and opinion mining. Text is analyzed for subjective information to determine the author's sentiment. Marketing, reputation management, social media monitoring, and consumer feedback analysis use it. An organization can automatically classify textual data as positive, negative, or neutral using sentiment analysis. This helps them gauge public opinion, mood, and client satisfaction. By eliminating human analysis, this automated system saves time and money and provides real-time insights. Human language is complex, making sentiment analysis difficult. Complexity, cultural context, and ambiguity make natural language sentiment analysis difficult. Academics in natural language processing have created sentiment lexicons, deep learning models, and machine learning algorithms to better sentiment analysis (Zendaoui et al., 2023).

6.2. Named Entity Recognition

Natural language processing requires named entity recognition (NER) to identify and label nouns, adjectives, adverbs, verbs, dates, places, and numbers. Unstructured text file retrieval, entity linkage, and knowledge extraction are easier using NER. Information extraction, document summarization, question answering, and knowledge graph development using NER. Named entity recognition (NER) systems improve text comprehension and information retrieval by automatically classifying named entities. Even with significant progress, named entities' diversity and ambiguity across domains and languages make NER problematic. Noise and confusion worsen the situation, needing advanced NLP (Karakanta, 2022).

6.3. Machine Translation

Machine translation (MT) is a new NLP technique. It helps overcome cultural and language barriers and provides information in multiple ways. Early machine translation systems translated documents using dictionaries and language standards. Statistics and neural machine translation (NMT) have improved translation fluency and quality. NMT is popular because it uses neural networks to learn and provide end-to-end translations. NMT models understand complex language patterns and context linkages using vast parallel corpora and sophisticated deep learning architectures, producing more accurate and natural-sounding translations (Zock, 2020).

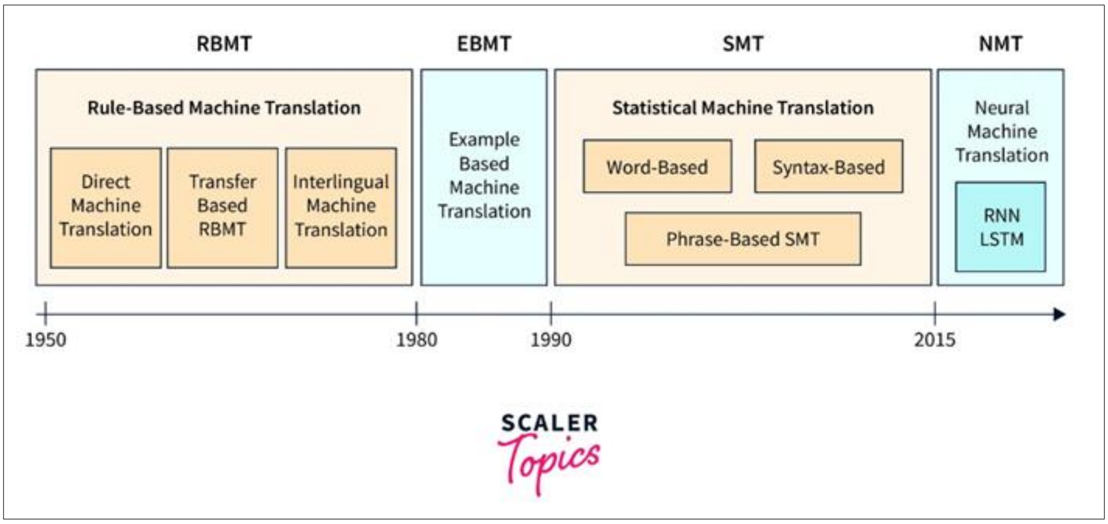

Figure 11 Machine Translation in NLP(Karakanta, 2022)

7. Question Answering Systems

Question-answering (QA) systems automatically answer user questions using NLP. Simple factoid-based systems search ordered databases, while complex systems may reason with unstructured material. Knowledge representation, machine learning, and NLU have improved QA systems. To deliver reliable and relevant findings, modern QA systems use information retrieval, knowledge reasoning, semantic parsing, and natural language comprehension. QA systems help with customer service, information retrieval, virtual assistants, and instructional technologies. Quality assurance systems speed up and accurately respond to user requests, enhancing efficiency, information availability, and satisfaction (Oyewole et al., 2024).

8. Challenges and Limitations

8.1. Ambiguity and Polysemy

NLP's major problems are polysemy and ambiguity. Words, phrases, and sentences with many meanings are ambiguous. Polysemy is a noun with multiple meanings. Language occurrences in spoken language can mislead natural language processing systems and produce misinterpretations. Ambiguity makes it difficult to understand a word in context. Consider "bank," which could mean a riverbed or bank. Information retrieval, sentiment analysis, and machine translation require ambiguity resolution. A word's nuanced or remote meaning may not be clear from context (Choi & Lee, 2023).

The polysemy also hinders natural language processing. Due to finer-grained ambiguity, algorithms cannot reliably identify the intended meaning of multi-meaning words. Light means "lightweight," "illuminating," or electromagnetic radiation. Natural language processing (NLP) research struggles to identify these senses due to extensive semantic comprehension. Complex solutions are needed to manage ambiguity and polysemy beyond study (Karakanta, 2022).

Figure 12 Computational Exploration of Linguistic Ambiguity

8.2. Data Bias and Ethical Concerns:

NLP systems are challenging to design and use due to data bias and ethical difficulties in computer linguistics. Biassed training data can make automated systems discriminate and perpetuate inequality. Included or removed opinions in training data influence natural language processing. Web and social media datasets may unintentionally include gender, racial, ethnic, and socioeconomic biases. NLP sentiment analysis, demographic profiling, and language production are affected by these biases. The assumptions and design of NLP algorithms can introduce biases. English-trained algorithms may be biased in multilingual situations because they cannot generalise to languages with differing linguistic patterns or cultural norms (Nilufar, 2024).

8.3. Evaluation Metrics and Benchmarking

Evaluation metrics and benchmarking are crucial for NLP system development and performance but have downsides. To make meaningful field comparisons and progress, assessment metrics must appropriately describe job objectives and performance standards. The difference between computerised and human perspectives hurts evaluation metrics. Automatic NLP system evaluation approaches like BLEU and ROUGE are efficient and scalable, but they do not match human quality and fluency assessments. An automated human NLP model evaluation gap could mislead academics and practitioners regarding model capabilities (Choi & Lee, 2023).

9. Emerging Trends and Future Directions

New technology, methods, and linguistic data have transformed natural language processing (NLP) in recent years. New computational linguistics trends and future breakthroughs are affecting natural language processing research and applications. This extensive study explores four NLP trends: multimodal, zero-shot learning, explainable AI, and cross-lingual, low-resource.

9.1. Multimodal NLP

Multimodal natural language processing (NLP) includes pictures, videos, and audio that alter human language perception and processing. Due to the development of online multimedia content and smart devices with various sensors, NLP models that interpret and generate content across modalities are needed (Ding et al., 2020). Recently developed multimodal natural language processing systems interpret and produce content across modalities. These systems describe multimodal data using RNNs, CNNs, and Transformer models. Multimodal natural language processing models can better interpret multimodal content by capturing complex semantic links between textual and non-textual data utilising fusion and cross-modal attention (Chen et al., 2022)

9.2. Zero-shot Learning

In NLP, zero-shot learning for information shortage and space variation is imaginative and energizing. Conventional managed to learn ideal models request huge volumes of named information, which might be unthinkable in speciality fields or dialects with restricted assets. Zero-shot learning permits models sum up to new classes or areas without human mediation. Zero-shot learning utilizes the move of figuring out how to work on pre-prepared models on one assignment and adjust them to one more with restricted named information. Area variation, not many shot learning, and meta-learning let NLP models apply important errand or space information to new settings. Zero-shot learning decentralizes NLP across dialects by disposing of the requirement for labelled information. This empowers NLP in speciality fields and underrepresented dialects.

9.3. Cross-lingual and Low-resource NLP

The need for low-resource, multilingual NLP systems drives demand for cross-lingual and low-resource NLP. Due to the expanding number of languages used online and the amount of multilingual material, natural language processing systems that can interpret, create, and translate text in numerous languages, even with minimal linguistic capabilities, are needed. Unlabeled input and similar linguistic patterns allow natural language processing models to generalise across languages and adapt to resource-constrained scenarios. (Chen et al., 2022).

9.4. Adoption of NLP in Various Sectors

Many industries employ NLP because it changes conventions. Healthcare practitioners can improve diagnosis accuracy and patient outcomes by using NLP to analyse massive amounts of unstructured clinical data. Marketing companies can use sentiment analysis algorithms to measure consumer preferences and adjust strategies to increase customer engagement and revenue. Cybersecurity threat detection and mitigation use NLP algorithms to find odd text patterns that may indicate intrusions. Legal contracts are summarised and analysed using natural language processing to save time and increase productivity. These and more domains are being optimised and innovated via NLP.

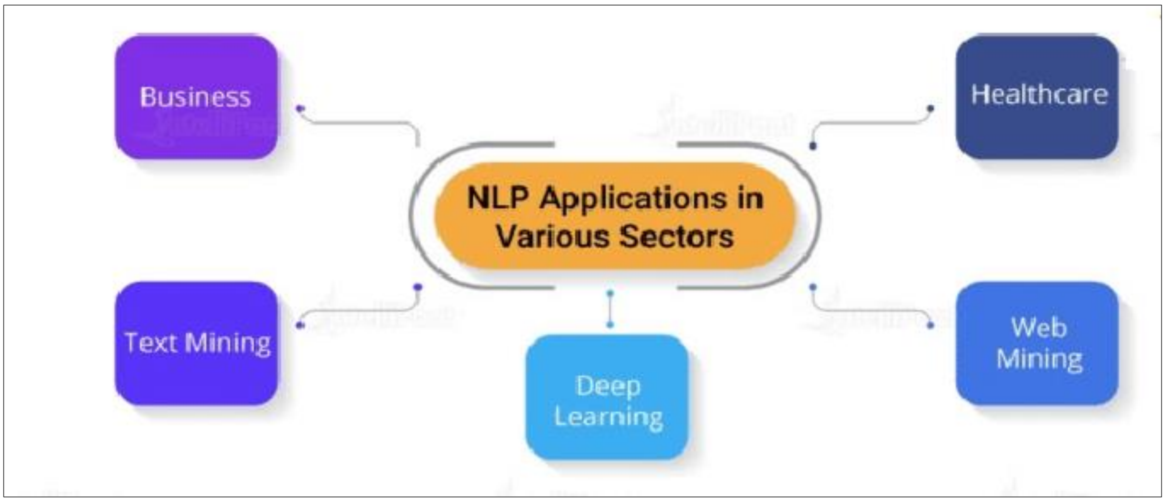

Figure 13 Application Sectors of NLP (Nilufar, 2024).

9.5. Collaboration Between Industry and Academia

NLP research and applications require academic-business collaboration. Industry stakeholders contribute statistics and practical difficulties, while academic institutions provide theoretical concepts and cutting-edge research. A symbiotic relationship makes industry-academia cooperation fruitful. They boost innovation, which enhances NLP solutions for many sectors. Collective research and knowledge exchange help bridge the gap between classroom learning and real-world application by teaching future experts to solve complex problems. Industry funding encourages academics to pursue high-impact research agendas, boosting multidisciplinary collaboration and natural language processing advancements (Clavel et al., 2022).

9.6. Research Funding and Grants

Future NLP depends on research grants and financing. Computational linguists rely on foundation and corporate funding. Academics can continue, to create, and solve major societal issues with proper funding. Targeted research funds also foster cross-disciplinary collaboration, which expands ideas and accelerates innovation. Academic challenge grants and prize contests spur NLP innovation by pushing the limits. NLP, research funds, and academic-business partnerships advance computational linguistics. This paper examines natural language processing breakthroughs and the need for synergistic interactions and continuing R&D from academic and industrial viewpoints. Academics, industry, and funding agencies must work together to better computational linguistics and society (Clavel et al., 2022).

10. Conclusion

Linguistic theory, deep learning, and machine learning have changed NLP in recent decades. This comprehensive assessment explored computational linguistics' advances, difficulties, and potential. NLP trains machines to understand human language. Advanced deep learning models, especially neural networks, enable more complicated and context-aware language processing. Transformer-based designs like BERT and GPT have improved sentiment analysis and machine translation. Despite the excitement about these new advances, NLP researchers still face challenges. Bias and unfairness in language models are major difficulties. Language and cognition theories are challenged by human-like language understanding. New NLP models excel at simple language tasks but lack pragmatics, context, and semantics. Linguistics, cognitive science, and AI research must fill language knowledge gaps.

NLP raises ethical, social, and technological challenges in addition to theoretical and technological ones. To make NLP more accessible, we must level the playing field so underrepresented groups can contribute to language technology development. Computational linguistics' future is bright and mysterious. When NLP, computer vision, multimodal learning, and human-computer interaction intersect, transdisciplinary research and innovation opportunities develop. Rapid technical advancement has raised concerns about social impacts, prompting requests for more openness, responsibility, and moral supervision in NLP system development and implementation.

Compliance with ethical standards

Disclosure of conflict of interest

Author declares no conflict of interest.

References

[1] Abdallah, A., Eberharter, D., Pfister, Z., & Jatowt, A. (2024). Transformers and Language Models in Form Understanding: A Comprehensive Review of Scanned Document Analysis. arXiv preprint arXiv:2403.04080.

[2] Aminovich, T. J. (2024). MODERN DIRECTIONS AND CURRENT PROBLEMS OF LINGUISTICS. International Journal of Formal Education, 3(3), 25-27.

[3] Becker, B. A., Maley, M., & Gallagher, D. (2023). Computing Crossroads. ACM Inroads, 14(4), 17-20.

[4] Choi, S. R., & Lee, M. (2023). Transformer architecture and attention mechanisms in genome data analysis: a comprehensive review. Biology, 12(7), 1033.

[5] Chen, P., Xu, H., Zhang, C., & Huang, R. (2022, July). Crossroads, buildings and neighborhoods: A dataset for fine-grained location recognition. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3329-3339).

[6] Clavel, C., Labeau, M., & Cassell, J. (2022). Socio-conversational systems: Three challenges at the crossroads of fields. Frontiers in Robotics and AI, 9, 937825.

[7] Chopan, M. A. (2023). REFLECTIONS UPON LITERATURE AT CROSSROADS IN THE AGE OF INTERACTIVE ARTIFICIAL INTELLIGENCE.

[8] Ding, Z., Xia, R., & Yu, J. (2020, July). ECPE-2D: Emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 3161-3170).

[9] Karakanta, A. (2022). Experimental research in automatic subtitling: At the crossroads between machine translation and audiovisual translation. Translation spaces, 11(1), 89-112.

[10] Legallois, D., & Koch, S. (2020). The notion of motif at the crossroads of disciplinesfolkloristics, narrativity, bioinformatics, automatic text processing, linguistics.

[11] Lenci, A., & Padó, S. (2022). Perspectives for natural language processing between AI, linguistics and cognitive science. Frontiers in Artificial Intelligence, 5, 1059998.

[12] Nilufar, N. (2024). THE RELATION OF COMPUTATIONAL LINGUISTICS TO NATURAL SCIENCES. International Multidisciplinary Journal for Research & Development

[13] Nahli, O., Sanna, A., Bandini, M., & Boschetti, F. (2022). Commerce Numérique: traffic signals for the crossroads between cultures. International Journal of Information Science and Technology, 5(3), 36-45.

[14] Oyewole, A. T., Adeoye, O. B., Addy, W. A., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Automating financial reporting with natural language processing: A review and case analysis. World Journal of Advanced Research and Reviews, 21(3), 575-589.

[15] Pirozelli, P., & Câmara, I. (2022). Natural language at a crossroads: formal and probabilistic approaches in philosophy and computer science. Manuscrito, 45, 50-81.

[16] Shokhrukh, J. B., & Abror, S. H. (2022). THE DEVELOPMENT TENDENCIES OF COMPUTATIONAL LINGUISTICS IN UZBEKISTAN: NLP, MACHINE TRANSLATION, CORPUS LINGUISTICS AND AUTOMATIC TEXT EDITING. The American Journal of Social Science and Education Innovations, 4(10), 01-05.

[17] Smidt, G. R., De Graef, K., & Lefever, E. (2024). At the Crossroad of Cuneiform and NLP: Challenges for Fine-grained Part-of-Speech Tagging. In The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. European Language Resources Association.

[18] Satibaldieva, N. (2024). CHALLENGES AND STRATEGIES FOR TERMINOLOGICAL CLARITY IN COMPUTER LINGUISTICS. ОБРАЗОВАНИЕ НАУКА И ИННОВАЦИОННЫЕ ИДЕИ В МИРЕ, 38(1), 166-168

[19] Ive, J. (2022). Natural language processing: A machine learning perspective by yue zhang and zhiyang teng.

[20] Xudaybergenova, Z., qizi Muxammedaliyeva, G. K., & Kdirbaeva, G. (2024). CURRENT PROBLEMS OF MODERN LINGUISTICS. World of Scientific news in Science, 2(3), 239-243.

[21] Yu, H., Fan, L., Li, L., Zhou, J., Ma, Z., Xian, L., ... & Ma, X. (2024). Large Language Models in Biomedical and Health Informatics: A Bibliometric Review. arXiv preprint arXiv:2403.16303.

[22] Yenduri, G., Srivastava, G., Maddikunta, P. K. R., Jhaveri, R. H., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2023). Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. arXiv preprint arXiv:2305.10435.

[23] Zendaoui, N., Bouchemal, N., & Zendaoui, Z. (2023, December). Revolutionizing Pandemic Management: A Comprehensive Review and Exploration of Technological Innovations. In 2023 Third International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS) (pp. 1-8). IEEE.

[24] Zock, M. (2020). AI at the Crossroads of NLP and Neurosciences. Journal of Cognitive Science, 21(1), 1-14