



A COMPARATIVE ANALYSIS OF WORD SENSE DISAMBIGUATION METHODS FOR IDENTIFYING POLYSEMY AND HOMONYMY IN ENGLISH

Ergashova Shahodat Ilhom qizi

Master's Student, Alfraganus University

shahodat1506@gmail.com

Abstract. This study presents a comparative and empirical analysis of Word Sense Disambiguation (WSD) methods in identifying polysemy and homonymy in English. The research evaluates three major approaches—knowledge-based, statistical, and deep learning—under a unified experimental framework using a semantically annotated dataset derived from SemCor and WordNet. The performance of the models was assessed using standard evaluation metrics, including accuracy, precision, and recall. The results demonstrate that the knowledge-based method shows limited effectiveness in complex contexts due to its reliance on lexical overlap, while the statistical approach provides more stable performance by capturing contextual co-occurrence patterns. The highest results were achieved by the transformer-based BERT model, which consistently outperformed other methods by effectively modeling contextual semantics at multiple levels. The findings highlight that the ability to capture deep contextual information is a critical factor in resolving semantic ambiguity, particularly in cases of highly polysemous and homonymous words. In addition to quantitative evaluation, qualitative error analysis revealed that metaphorical and idiomatic usages remain challenging for all approaches, indicating the need for further improvements. The study contributes to the field by providing a systematic comparison of WSD methods under consistent conditions and offers practical insights for improving natural language processing applications such as machine translation and automated text analysis.



Keywords: Word Sense Disambiguation, polysemy, homonymy, semantic ambiguity, BERT, contextual modeling, computational linguistics.

Annotatsiya. Mazkur tadqiqot ingliz tilidagi polisemiya va omonimiyani aniqlashda soʻz maʼnosini ajratish (Word Sense Disambiguation, WSD) usullarining qiyosiy va empirik tahliliga bagʻishlangan. Tadqiqotda bilimga asoslangan, statistik va chuqur oʻrganishga asoslangan yondashuvlar yagona eksperimental muhitda baholandi hamda SemCor va WordNet asosida shakllantirilgan semantik belgilangan maʼlumotlar toʻplamidan foydalanildi. Modellar samaradorligi aniqlik, precision va recall kabi mezonlar orqali baholandi. Natijalar shuni koʻrsatdiki, bilimga asoslangan yondashuv lugʻaviy moslikka tayanishi sababli murakkab kontekstlarda past samaradorlikni namoyon qiladi, statistik model esa kontekstual bogʻliqliklarni hisobga olgani tufayli nisbatan barqaror natijalarni beradi. Eng yuqori natijalar transformer arxitekturasi asosidagi BERT modeli tomonidan qayd etildi, bu esa uning kontekstni koʻp qatlamli semantik darajada qayta ishlash qobiliyati bilan izohlanadi. Tadqiqot natijalari chuqur kontekstual tahlilning semantik noaniqlikni bartaraf etishda hal qiluvchi omil ekanligini tasdiqlaydi. Shuningdek, xatoliklar tahlili metaforik va idiomatik qoʻllanishlar barcha modellarda murakkablik tugʻdirishini koʻrsatdi. Ushbu tadqiqot WSD usullarini bir xil sharoitda qiyosiy baholab, tabiiy tilni qayta ishlash tizimlarida semantik aniqlikni oshirish uchun amaliy tavsiyalar beradi.

Kalit soʻzlar: soʻz maʼnosini ajratish (WSD), polisemiya, omonimiya, semantik noaniqlik, BERT, kontekstual modellashtirish, kompyuter lingvistikasi

Introduction

Polysemy and homonymy are fundamental linguistic phenomena that play a crucial role in the semantic structure of modern English. While polysemy refers to a single lexical item having multiple related meanings, homonymy involves words that share the same form but possess distinct and unrelated meanings. Although these



phenomena contribute to the richness and flexibility of language, they also introduce significant challenges for automatic language processing systems. In particular, semantic ambiguity arising from context-dependent meanings remains one of the major obstacles in natural language processing (NLP), negatively affecting the performance of applications such as machine translation, information retrieval, and automated text analysis (Loureiro et al., 2022; Camacho-Collados & Pilehvar, 2020).

One of the core approaches to addressing this challenge is Word Sense Disambiguation (WSD), which aims to determine the correct meaning of a word based on its context. WSD has long been considered a central problem in computational linguistics due to its direct impact on semantic understanding (Navigli, 2009). Early research in this area primarily relied on knowledge-based methods, among which the Lesk Algorithm is one of the most prominent. This algorithm determines word meaning by measuring the overlap between dictionary definitions and contextual words. However, despite its conceptual simplicity, its effectiveness is limited by the quality of lexical resources and its inability to capture deeper contextual semantics.

Subsequently, statistical and probabilistic models were introduced, leveraging large corpora to infer word meanings based on contextual patterns. These approaches demonstrated improved performance compared to traditional methods by incorporating frequency and co-occurrence information (Jia et al., 2021; Wang et al., 2021). Nevertheless, statistical models often struggle with capturing deeper semantic relationships, especially in complex or less frequent contexts. In recent years, the emergence of deep learning techniques has significantly advanced the field of WSD. Transformer-based models, particularly Bidirectional Encoder Representations from Transformers (BERT), have shown remarkable performance by modeling contextual information at multiple semantic levels.



Despite these advancements, a critical gap remains in the literature. Most existing studies focus on individual approaches rather than conducting comprehensive comparative analyses under unified experimental conditions. Moreover, limited attention has been given to evaluating these methods specifically in the context of polysemy and homonymy, where semantic ambiguity is particularly pronounced. Additionally, the role of contextual depth in influencing model performance has not been systematically examined.

This study aims to address these gaps by conducting a comparative and empirical evaluation of knowledge-based, statistical, and deep learning approaches to WSD using a unified dataset. The primary objective is to assess the effectiveness of these methods in resolving semantic ambiguity in polysemous and homonymous words. By analyzing their performance through standard evaluation metrics, this research provides insights into their strengths and limitations from both computational and linguistic perspectives.

The novelty of this study lies in its integrated evaluation framework, where multiple WSD approaches are tested under identical conditions, enabling a fair and systematic comparison. Furthermore, the findings contribute to a deeper understanding of how contextual modeling affects semantic interpretation, offering practical implications for improving NLP applications such as machine translation and automated text analysis.

Literature Review

Word Sense Disambiguation (WSD) has been extensively studied as a core problem in computational linguistics, with a wide range of approaches proposed over the years. The complexity of WSD arises from the inherent ambiguity of natural language, particularly in cases of polysemy and homonymy, where the correct interpretation of a word heavily depends on its context (Loureiro et al., 2022).



Early approaches to WSD were predominantly knowledge-based, relying on lexical resources such as dictionaries and semantic networks. One of the earliest and most influential methods is the Lesk Algorithm, which identifies word meaning by measuring the overlap between the glosses of candidate senses and the surrounding context. While this method is intuitive and does not require training data, its performance is often constrained by the limited coverage and granularity of lexical databases, as well as its inability to capture contextual nuances.

To overcome these limitations, statistical approaches were introduced, utilizing large corpora to model contextual relationships between words. These methods rely on co-occurrence patterns and probabilistic inference to determine the most likely sense of a word in a given context. For instance, the TWE-WSD model incorporates topical word embeddings to improve semantic representation, while the approach proposed by Wang et al. (2021) enhances performance by dynamically adjusting semantic features. Despite their effectiveness, statistical models often fail to capture deeper semantic and pragmatic aspects of language, particularly in complex or low-frequency contexts.

The advent of deep learning has led to significant breakthroughs in WSD. Neural models, especially those based on transformer architectures, have demonstrated superior performance by capturing contextual information at multiple levels. The BERT model represents a major milestone in this regard, enabling context-sensitive word representations that significantly improve disambiguation accuracy. Similarly, contextualized embeddings introduced by Peters et al. (2018) have further enhanced semantic modeling capabilities. These approaches allow models to dynamically adjust word representations depending on context, making them particularly effective for resolving semantic ambiguity (Camacho-Collados & Pilehvar, 2020).



Recent studies have also focused on expanding evaluation frameworks and addressing more complex linguistic phenomena. For example, the XL-WSD framework provides a large-scale, cross-lingual benchmark for evaluating WSD systems. Additionally, research on idiomaticity and domain-specific language usage has highlighted the challenges posed by non-literal meanings and specialized contexts (He et al., 2024; Chopard et al., 2024). Optimization-based approaches, such as ADCSA-WSD, further demonstrate the potential of hybrid and adaptive techniques in improving WSD performance.

Despite these advancements, several limitations remain. Many studies focus on specific methods without providing comprehensive comparisons across different approaches. Furthermore, the interaction between linguistic complexity-particularly polysemy and homonymy-and model performance has not been sufficiently explored. The influence of contextual depth on disambiguation accuracy also remains an open research question.

This study builds upon existing research by providing a systematic comparison of knowledge-based, statistical, and deep learning approaches under consistent experimental conditions. By focusing specifically on polysemous and homonymous words, it aims to bridge the gap between computational modeling and linguistic analysis, offering a more comprehensive understanding of WSD performance in complex semantic contexts.

Methodology

This study adopts a comparative and experimental approach to evaluate the effectiveness of different Word Sense Disambiguation (WSD) methods in identifying polysemy and homonymy in English. The research design is based on applying multiple disambiguation techniques to a unified dataset and assessing their performance using consistent evaluation criteria. Such an approach enables an



objective comparison of different models and provides insight into their practical effectiveness in resolving semantic ambiguity.

The dataset used in this study was derived from two widely recognized linguistic resources: SemCor and WordNet. These resources were selected due to their reliability and extensive use in WSD research, as they provide standardized sense annotations and well-structured lexical information. From these sources, a set of polysemous and homonymous words was selected, including *bank*, *bat*, *light*, *head*, and *spring*. These lexical items were chosen because they exhibit multiple meanings across different contexts, making them suitable for evaluating disambiguation performance.

For each target word, a minimum of one hundred contextual sentences was extracted, resulting in a dataset of five hundred instances. The dataset was carefully balanced to include a variety of contexts, ranging from literal to more abstract and complex usages. To ensure a systematic evaluation, the data were divided into training, validation, and test subsets, allowing the models to be assessed under controlled and comparable conditions.

Three different WSD approaches were implemented in this study: a knowledge-based method, a statistical model, and a deep learning approach. The knowledge-based method was based on a modified version of the Lesk Algorithm, which determines word meaning by comparing dictionary definitions with the surrounding context. This method relies on lexical overlap and semantic similarity between glosses and contextual words.

The statistical approach was implemented using a Naive Bayes classifier based on contextual co-occurrence frequencies and probabilistic associations between neighboring words. This model determines the most likely sense of a target word by analyzing the distribution of contextual features within the dataset. This approach allows for a more data-driven interpretation of meaning compared to



knowledge-based methods, although it still depends on the availability and distribution of contextual information.

The deep learning approach employed the BERT model, which is based on transformer architecture and pre-trained on large-scale corpora. This model enables context-sensitive representations of words by considering both local and global contextual information within a sentence. The bert-base-uncased version of BERT was used in the experiments. The model was fine-tuned for three epochs with a batch size of 16 and a learning rate of $2e-5$ in order to optimize contextual semantic representation for the WSD task.

The performance of each method was evaluated using standard metrics, including accuracy, precision, and recall. These metrics provide a comprehensive assessment of model effectiveness by measuring not only overall correctness but also the quality and coverage of the predicted senses. All models were tested under identical experimental conditions, using the same dataset and evaluation procedures, ensuring a fair comparison of their performance.

In addition to quantitative evaluation, a qualitative error analysis was conducted to identify patterns of misclassification. Particular attention was given to cases involving metaphorical usage, idiomatic expressions, and contexts with limited semantic cues, as these are known to pose challenges for WSD systems. This analysis provided deeper insights into the strengths and limitations of each approach.

Despite its systematic design, the study has several limitations. Although the dataset size is relatively limited, the selected lexical items were intentionally chosen due to their high semantic complexity and frequent use in WSD research. Therefore, the dataset provides a suitable basis for comparative evaluation. Nevertheless, larger corpora and more diverse linguistic contexts may improve the generalizability of future findings. In addition, the performance of deep learning models may vary depending on computational resources and parameter configurations.



Results

The experimental evaluation was conducted to compare the effectiveness of knowledge-based, statistical, and deep learning approaches in resolving semantic ambiguity in polysemous and homonymous words. The results reveal clear performance differences among the applied methods, particularly as contextual complexity increases.

Overall, the knowledge-based approach achieved an average accuracy of 66%, while the statistical model reached 79%, and the deep learning approach based on BERT achieved the highest performance with an average accuracy of 88%. These findings indicate that models capable of capturing deeper contextual information significantly outperform those relying on surface-level lexical matching or frequency-based patterns.

Table 1 presents the accuracy scores for each target word across the three approaches. The results demonstrate that performance varies depending on the degree of semantic complexity associated with each lexical item. For instance, relatively straightforward contexts yield closer results across models, whereas highly polysemous words highlight more substantial performance gaps.

Table 1. Accuracy (%) of WSD methods across target words

Word	Lesk (%)	Statistical (%)	BERT (%)
bank	69	81	90
bat	61	76	87
light	64	78	87
head	68	80	89
spring	63	77	86
Average	66	79	88

The word-level analysis shows that the knowledge-based method performs reasonably well in less ambiguous contexts but struggles with words exhibiting



multiple abstract or context-dependent meanings, such as *light* and *spring*. In contrast, the statistical model demonstrates more stable performance across different contexts, benefiting from its ability to capture co-occurrence patterns. However, its effectiveness declines in cases where contextual signals are sparse or atypical. The deep learning model consistently achieves the highest accuracy across all target words, maintaining robustness even in semantically complex contexts.

A broader evaluation using standard performance metrics further supports these findings. As shown in Table 2, the deep learning model outperforms the other approaches not only in terms of accuracy but also in precision and recall, indicating both higher correctness and broader coverage of possible word senses.

Table 2. Overall performance of WSD methods

Method	Accuracy (%)	Precision (%)	Recall (%)
Lesk	66	67	65
Statistical	79	80	78
BERT	88	89	87

The statistical model achieves a balanced performance, with relatively high precision and recall values, while the knowledge-based approach exhibits lower recall, suggesting its limited ability to capture all relevant meanings. The superior performance of the deep learning model reflects its capacity to integrate contextual information at multiple levels, leading to more accurate disambiguation.

To provide a clearer comparison, Figure 1 illustrates the performance of all three approaches across the main evaluation metrics. The visualization highlights the consistent advantage of the deep learning model and the intermediate performance of the statistical approach.

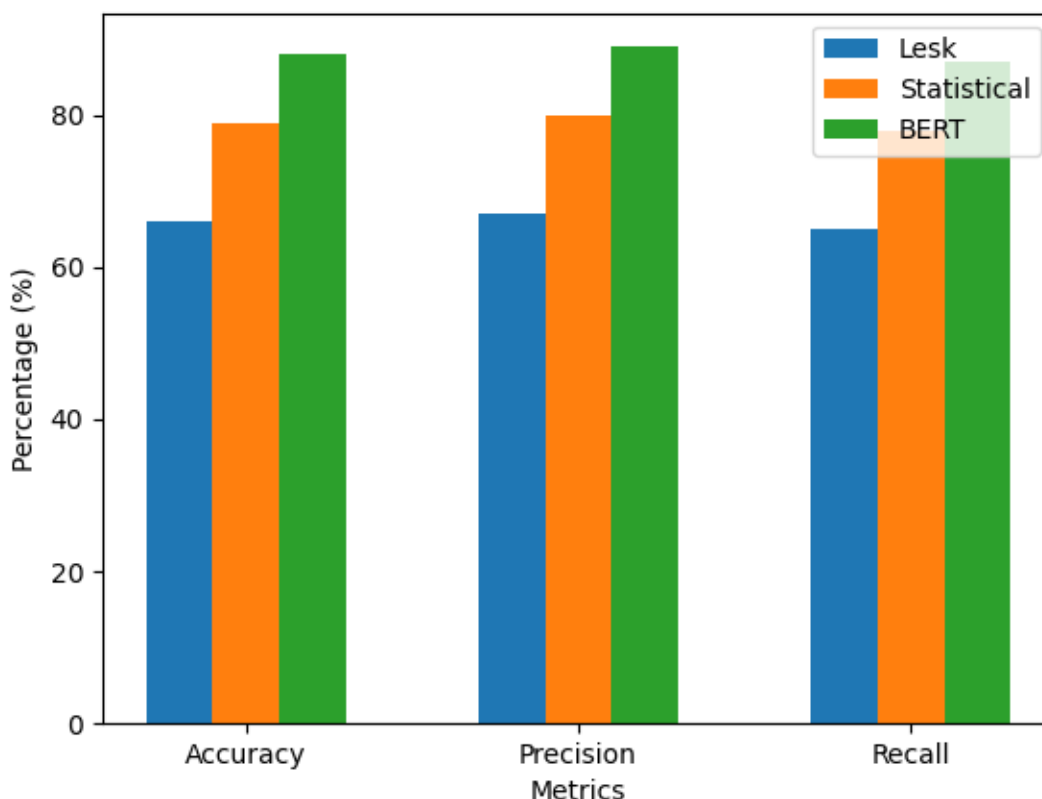


Figure 1. Comparison of WSD methods across evaluation metrics

Error analysis revealed that the most challenging cases for all models involved metaphorical expressions, idiomatic usage, and contexts with limited semantic cues. These findings suggest that while modern models significantly improve disambiguation performance, certain linguistic phenomena remain difficult to capture fully.

Discussion

The findings of this study provide important insights into the relationship between contextual modeling and semantic disambiguation. The results clearly demonstrate that the effectiveness of WSD approaches is strongly dependent on their ability to incorporate contextual information at varying levels of depth. As shown in Table 1, the performance gap between models becomes more pronounced as the semantic complexity of the target words increases, highlighting the limitations of traditional approaches in handling polysemy and homonymy.



The relatively low performance of the knowledge-based approach can be attributed to its reliance on lexical definitions and surface-level matching. Although methods based on the Lesk Algorithm offer interpretability and simplicity, they fail to capture deeper semantic relationships required for resolving ambiguity in complex contexts. This limitation is particularly evident in cases where meanings are context-dependent or abstract, as reflected in the lower scores for words such as *light* and *spring* in Table 1.

The statistical model demonstrates improved performance by leveraging contextual co-occurrence patterns, as reflected in both Table 1 and Table 2. Its relatively balanced precision and recall indicate that it can generalize better than knowledge-based methods. However, its dependence on frequency patterns limits its effectiveness in handling rare or unconventional contexts. This observation aligns with previous research suggesting that statistical models struggle with sparse data and lack semantic depth.

In contrast, the deep learning approach based on BERT achieves consistently superior performance across all evaluation metrics, as shown in Table 2 and visualized in Figure 1. The model's ability to capture both local and global contextual information allows it to effectively resolve ambiguity even in highly complex linguistic environments. These findings are consistent with recent studies emphasizing the advantages of contextualized language models in semantic tasks (Loureiro et al., 2022; Camacho-Collados & Pilehvar, 2020).

The visualization in Figure 1 further reinforces this conclusion by clearly illustrating the performance differences among the three approaches. The deep learning model maintains a noticeable advantage across all metrics, while the statistical model occupies an intermediate position and the knowledge-based approach lags behind. This pattern suggests that incorporating deeper contextual representations is essential for achieving high-performance WSD systems.



The error analysis provides additional insight into the limitations of current approaches. Across all models, the most frequent errors were associated with figurative language, idiomatic expressions, and contexts lacking sufficient semantic cues. These findings indicate that even advanced models struggle with pragmatic and discourse-level aspects of language, which are not always explicitly encoded in textual data. Although the deep learning model performs better in these cases, it is not immune to such challenges, suggesting that further improvements are needed.

From a linguistic perspective, the results confirm that semantic ambiguity cannot be effectively resolved without considering broader contextual structures. Polysemy and homonymy are inherently context-driven phenomena, and their accurate interpretation requires models that go beyond surface-level analysis. This explains the superior performance of transformer-based models, which are designed to process contextual dependencies more effectively.

At the same time, the findings suggest that traditional approaches should not be entirely disregarded. Knowledge-based and statistical methods still demonstrate reasonable performance in simpler contexts and offer advantages such as interpretability and lower computational cost. Therefore, future research may benefit from developing hybrid models that combine the strengths of different approaches.

Despite the robustness of the results, certain limitations must be acknowledged. The dataset size, although sufficient for comparative analysis, may limit the generalizability of the findings. Additionally, the performance of deep learning models is influenced by computational resources and parameter settings. Nevertheless, the consistency of the results across multiple evaluation metrics and their alignment with existing research provide strong evidence for the effectiveness of contextual models in WSD tasks.

Conclusion

This study investigated the effectiveness of different Word Sense Disambiguation (WSD) approaches in identifying polysemy and homonymy in



English through a comparative and experimental framework. The results demonstrate that the performance of WSD methods is strongly influenced by their ability to capture contextual information. Knowledge-based approaches showed relatively limited performance due to their reliance on surface-level lexical matching, while statistical models provided more balanced results by incorporating contextual co-occurrence patterns. The highest performance was achieved by the deep learning approach based on BERT, confirming its superiority in modeling complex semantic relationships.

The findings further indicate that as the degree of polysemy increases, the performance gap between traditional and contextual models becomes more pronounced. In semantically complex contexts, transformer-based models maintain higher accuracy and stability, whereas traditional approaches struggle to resolve ambiguity effectively. This highlights the crucial role of deep contextual representation in semantic disambiguation tasks.

From a theoretical perspective, this study contributes to the understanding of how different computational approaches handle semantic ambiguity, particularly in relation to polysemy and homonymy. It provides empirical evidence that contextual modeling is a key factor in improving WSD performance. From a practical standpoint, the results have important implications for enhancing natural language processing applications, including machine translation, information retrieval, and automated text analysis systems.

At the same time, the study acknowledges several limitations. The relatively limited size of the dataset may affect the generalizability of the findings, and the performance of deep learning models remains dependent on computational resources and parameter configurations. Future research should focus on expanding the dataset, exploring multilingual scenarios, and developing hybrid models that integrate linguistic knowledge with deep learning techniques.



Overall, the findings provide empirical evidence that transformer-based contextual models are significantly more effective in resolving semantic ambiguity in English polysemous and homonymous words than traditional knowledge-based and statistical approaches.

References

1. Abdelaali, A., Bekhouche, S., & Saadi, T. (2022). ADCSA-WSD: Adapted discrete crow search algorithm for word sense disambiguation. *Revue d'Intelligence Artificielle*, 36(1), 131–138. <https://doi.org/10.18280/ria.360115>
2. Jia, L., Wang, H., & Li, J. (2021). TWE-WSD: An effective topical word embedding based word sense disambiguation. *CAAI Transactions on Intelligence Technology*, 6(1), 72–79. <https://doi.org/10.1049/cit2.12006>
3. Pasini, T., Raganato, A., & Navigli, R. (2021). XL-WSD: An extra-large and cross-lingual evaluation framework for word sense disambiguation. *Proceedings of AAAI*, 13648–13656. <https://doi.org/10.1609/aaai.v35i15.17609>
4. Loureiro, D., Jorge, A., & Camacho-Collados, J. (2022). Advances in word sense disambiguation: A survey. *Computational Linguistics*, 48(2), 357–403. https://doi.org/10.1162/coli_a_00435
5. Camacho-Collados, J., & Pilehvar, M. T. (2020). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, 67, 743–788. <https://doi.org/10.1613/jair.1.11703>
6. He, W., Vieira, T. K., Garcia, M., Scarton, C., Idiart, M., & Villavicencio, A. (2024). Investigating idiomaticity in word representations. *Computational Linguistics*, 50(2), 505–555. https://doi.org/10.1162/coli_a_00546
7. Wang, W., et al. (2021). Word sense disambiguation based on stretchable semantic features. *Mathematical Foundations of Computing*, 4(2), 1–15. <https://doi.org/10.3934/mfc.2020022>



8. Kulkarni, D. S., & Rodd, S. F. (2022). Word sense disambiguation for lexicon-based sentiment analysis in Hindi. *Webology*, 19(1), 592–600. <https://doi.org/10.14704/web/v19i1/web19042>
9. Chopard, D., et al. (2024). Word sense disambiguation of acronyms in clinical narratives. *Frontiers in Digital Health*, 6, 1282043. <https://doi.org/10.3389/fdgth.2024.1282043>
10. Laukaitis, A. (2024). Reversal of the word sense disambiguation task using optimized frameworks. *Applied Sciences*, 14(13), 5550. <https://doi.org/10.3390/app14135550>
11. Peters, M. E., et al. (2018). Deep contextualized word representations. In *Proceedings of NAACL-HLT* (pp. 2227–2237). <https://doi.org/10.18653/v1/N18-1202>
12. Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186). <https://doi.org/10.18653/v1/N19-1423>