Enhancing WordNet with Specialized Lexicons for Improved Detection of Culturally Sensitive Terms

Abstract

WordNet, a comprehensive lexical database for English, lacks specialization in culturally specific or sensitive lexicons, limiting its effectiveness in detecting words related to ethnicity, diaspora, slurs, or reclaimed terms. To propose and evaluate a method for integrating WordNet with specialized lexicons to improve detection and relationship mapping of culturally sensitive terms. We analyze the coverage of various term categories in WordNet and estimate the potential improvements when integrating with a custom lexicon. We propose new relationship mappings and evaluate their potential coverage. Integration of WordNet with specialized lexicons can significantly improve coverage across various categories of culturally sensitive terms, with overall coverage estimates ranging from 60 to 85 out of 100 terms, depending on the specific domain. The proposed integration method shows promise in enhancing WordNet's capabilities for detecting and mapping relationships between culturally sensitive terms, potentially improving its utility in natural language processing tasks related to cultural understanding and hate speech detection.

1 Introduction

WordNet stands as one of the most extensive lexical databases for the English language, encompassing approximately 117,000 synsets and over 150,000 unique words. Its primary focus lies in general English vocabulary, providing relationships such as synonyms, antonyms, hypernyms, hyponyms, and meronyms(Miller et al., 1990; Fellbaum, 1990; Fellbaum, 2014; Miller and Fellbaum, 2007). However, WordNet's broad coverage comes at the cost of specialization in culturally specific or sensitive lexicons, which limits its immediate effectiveness in detecting words of concern related to ethnicity, diaspora, slurs, or reclaimed terms. Many of these terms were annotated in SENTIWORDNET 3.0(Baccianella et al., 2010; Esuli and Sebastiani, 2006) and WordNet-Affect (Strapparava and Valitutti, 2004). In contrast, custom lexicons designed to detect ethnic, cultural, sensitive, or offensive terms are typically much smaller in size compared to WordNet, but they offer a high degree of specialization. Such lexicons might contain between 2,000 and 4,000 words, depending on their level of detail and focus. These specialized resources often pay attention to context, recognizing that certain words may become offensive only in specific situations or regions. We argue that adding new relations can help with improving the detection and relationship mapping of culturally sensitive terms

2 Offensive Language Detection

There are various innovative approaches to detect and mitigate hate speech in online environments, particularly on social media platforms. One study employed a hybrid deep learning approach using convolutional neural networks (CNN) and bidirectional gated recurrent units (Bi-GRU) for hate speech detection on Twitter. The researchers built four models: CNN, Bi-GRU, CNN+Bi-GRU, and Bi-GRU+CNN, utilizing term frequency-inverse document frequency (TF-IDF) for feature extraction and FastText for feature expansion. The best-performing model achieved an accuracy of 87.63%, demonstrating the potential of hybrid deep learning in comprehending sentences broken down

by hybrid n-gram types, specifically Unigram-Bigram-Trigram(Gde Bagus Janardana Abasan and Setiawan, 2024).

Detection and substitution algorithms were combined to address toxic content, define problematic text and suggest euphemistic alternatives to educate users about more inclusive language choices with an NLP classifier to promote self-awareness among users and target the issue at its source. Attention network visualization methods were proposed to improve hate speech detection and train embeddings through transfer learning, followed by synonym expansion of semantic vectors. Active learning cycles and entropy-based selection techniques were used to enhance the model's accuracy. This method achieved a receiver operating characteristic (ROC) of 0.91 and a precision-recall score of 0.90, while also providing visualizations to illustrate the rationale behind hate speech classifications. The use of semantic embedding and lexicon expansion played a crucial role in improving the model's performance (Ahmed and Lin, 2024).

One study focused on preprocessing techniques for Arabic offensive language classification, including emoji conversion, letter normalization, and hashtag segmentation. BERT-based models did not show significant improvements in covering broader domains and dialects to further refine these preprocessing(Husain and Uzuner, 2022) although BERT-based were successful in other sense delineations task(Tóth and Abdelzaher, 2023).

The impact of text normalization

on hate speech detection, particularly for out-of-vocabulary (OOV) words with repeated letters could detect offensive language, combining rule-based patterns and the SymSpell spelling correction algorithm. This approach reduced OOV words by 8% and improved the F1 score of the detection model by 9-13% compared to existing methods, demonstrating the value of effective text normalization in enhancing hate speech detection. The model applied multiple rules regarding the position of repeated letters in a word, considering whether they appeared at the beginning, middle, or end of the word and the repetition pattern (Mansur et al., 2024). However, researchers could consider developing relations that capture the emotional intensity of words, the cultural context of potentially offensive terms, or the historical evolution of language used in discriminatory contexts.

3 Derogatory Exonyms, Caconyms and Endonym

Synonym relations are often employed in NLP and wordNet applications to address hate speech and covert offensive language, transforming them into more inclusive alternatives. By identifying and replacing harmful terms with neutral or positive equivalents, these systems can help reduce the prevalence of discriminatory language in various contexts(Petiwala and Siva Sathya, 2011). This approach leverages the semantic relationships between words to find suitable substitutions that preserve the intended meaning while removing offensive connotations.

Exonyms are names used by outsiders to refer to a place, group of people, or cultural entity, different from the name used by the people or group themselves. For instance, "Germany" is an exonym used in English, whereas Germans refer to their country as "Deutschland." While exonyms are common and often neutral, they can sometimes take on negative connotations, especially when they reflect colonial history or outdated, foreign views of a group(Vidović, 2022; Nick, 2020; Jordan, 2023). In certain cases, exonyms are used in a dismissive or offensive way, reinforcing cultural otherness, or stereotypes, such as calling the Indigenous peoples of the Americas "Indians," a misnomer from colonial times.

Caconyms are incorrect or improper names used for people, places, or things. They often arise from linguistic misunderstandings or historical errors. Caconyms can range from being mildly inaccurate to highly offensive, particularly when they perpetuate outdated, incorrect, or derogatory representations. For example, the term "Eskimo" is considered a caconym for the Inuit people, as it originates from an external misunderstanding of their culture and has pejorative overtones. Mislabeling with caconyms can carry unintended disrespect or reinforce harmful narratives, especially when tied to colonialism, racism, or ignorance(2023.)

Endonyms are names used by a group of people to describe themselves, their land, or their cultural practices. These are self-referential terms, like "Suomi" for Finland or "Roma" for the Romani people. Using endonyms tends to reflect respect and recognition of a group’s self-identification. However, problems arise when outsiders ignore or refuse to use endonyms, opting instead for

offensive exonyms or caconyms. While endonyms generally carry neutral or positive connotations, their omission in favor of external terms can become offensive, especially if the exonym or caconym has a history of derogatory usage or condescension. In all three cases, terms that are incorrectly applied or misused can become offensive depending on historical, cultural, or political context. What begins as an exonym can evolve into a slur or derogatory term if used to demean or "other" a group. Similarly, caconyms, though often unintended in their offense, can perpetuate ignorance and harm.

3 Proposing New -nyms Towards Inclusive Languaging

The coverage of different categories of terms varies considerably. For general ethnic and national terms, WordNet's inherent coverage is estimated to be high, between 70 and 85 out of 100 terms. However, for more specific ethnonyms and endonyms, WordNet's coverage drops to between 30 and 50 out of 100 terms. The representation of slurs and derogatory exonyms in WordNet is particularly limited, with coverage estimated at only 10 to 25 out of 100 terms without a specialized lexicon. Integration with a custom lexicon can raise this coverage to between 70 and 80 out of 100 terms for offensive and derogatory language.

Reclaimed terms and euphemisms present a particular challenge for WordNet, with initial coverage estimated at less than 10 out of 100 terms. The addition of a specialized lexicon could potentially increase this coverage to between 60 and 75 out of 100 terms, depending on the depth of reclaimed and euphemistic language in the custom database. For culturally sensitive terms related to diaspora and mixed ethnicity, WordNet's coverage is moderate, at about 40 to 60 out of 100 terms, but this could be significantly enhanced to between 75 and 85 out of 100 terms with a custom lexicon.

The generation of new relationships through the integration of WordNet and specialized lexicons offers significant potential for improved coverage. Mapping WordNet hypernyms and hyponyms to specific ethnic and cultural categories through the custom lexicon could achieve coverage of 85 to 90 out of 100 terms in this domain. For antonyms and endonyms, coverage could improve from 40-60 out of 100 terms to 70-80 out of 100 terms with additional resources focused on self-referential names. Historical and diachronic analysis, which is not a focus of WordNet, could see coverage increase from 10-20 out of 100 terms to 65-75 out of 100 terms through integration with a lexicon designed to detect shifts in meaning over time.

In conclusion, the integration of WordNet with specialized lexicons shows promise in significantly enhancing coverage across various categories of culturally sensitive terms. The overall estimated coverage for detecting and generating new relationships between words and ethnic/cultural terms ranges from 60 to 85 out of 100 terms, depending on the specific domain and complexity of the relationships being mapped. This improved coverage has the potential to enhance WordNet's utility in natural language processing tasks related to cultural understanding and hate speech detection.

The methodology adopted for mapping pre-existing WN's relations to the new ones is summarized as follows:

```
function mapWordNetToNewRelations(word):
    // Initialize new relation mappings
    newRelations = {}
    // Fetch existing WordNet relations for the word
    synsets = WordNet.getSynsets(word)

    for synset in synsets:
        // Check for hypernyms and map to ethnonyms if applicable
        hypernyms = synset.getHypernyms
        for hypernym in hypernyms:
            if isEthnicTerm(hypernym):
                newRelations["ethnonym"] = hypernym
        // Check for hyponyms and map to specific ethnic terms
        hyponyms = synset.getHyponyms
        for hyponym in hyponyms:
            if isEthnicTerm(hyponym):
                newRelations["specific_ethnonym"] = hyponym
        // Check for antonyms and map to potential endonyms
        antonyms = synset.getAntonyms
        for antonym in antonyms:
            if isEndonym(antonym):
                newRelations["endonymic"] = antonym
        // Check for holonyms and map to cultural context
        holonyms = synset.getHolonyms
        for holonym in holonyms:
            if isCulturalContext(holonym):
                newRelations["cultural_context"] = holonym
        // Check for meronyms and map to specific cultural attributes
        meronyms = synset.getMeronyms
        for meronym in meronyms:
            if isCulturalAttribute(meronym):
                newRelations["cultural_attribute"] = meronym
    // Add custom relations based on lexical analysis
    if isOffensive(word):
        neutralTerm = findNeutralAlternative(word)
        newRelations["endonymic"] = neutralTerm
    if hasHistoricalContext(word):
        historicalUsage = getHistoricalUsage(word)
        currentUsage = getCurrentUsage(word)
        newRelations["diachronic"] = (historicalUsage, currentUsage)
    if hasCulturalVariation(word):
```

```
            variations = getCulturalVariations(word)
            newRelations["culturally_specific_endonymic"] = variations
        if isMixedEthnicityTerm(word):
            appropriateTerms = findAppropriateTerms(word)
            newRelations["mixed_ethnicity_descriptor"] =
appropriateTerms
        if isDiasporicTerm(word):
            homelandTerms = findHomelandTerms(word)
            newRelations["diasporic_variant"] = homelandTerms
        if isSlur(word):
            severityLevel = assessSeverity(word)
            newRelations["offensiveness_level"] = severityLevel
        if hasEuphemisticAlternative(word):
            euphemism = findEuphemism(word)
            newRelations["euphemistic_substitute"] = euphemism
        if isReclaimedTerm(word):
            reclaimingGroup = findReclaimingGroup(word)
            newRelations["reclaimed_usage"] = reclaimingGroup
        if hasCrossCulturalEquivalent(word):
            equivalents = findCrossCulturalEquivalents(word)
            newRelations["cross_cultural_equivalence"] = equivalents
        return newRelations
    // Helper functions (to be implemented)
    function isEthnicTerm(term): ...
    function isEndonym(term): ...
    function isCulturalContext(term): ...
    function isCulturalAttribute(term): ...
    function isOffensive(term): ...
    function findNeutralAlternative(term): ...
    function hasHistoricalContext(term): ...
    function getHistoricalUsage(term): ...
    function getCurrentUsage(term): ...
    function hasCulturalVariation(term): ...
    function getCulturalVariations(term): ...
    function isMixedEthnicityTerm(term): ...
    function findAppropriateTerms(term): ...
    function isDiasporicTerm(term): ...
    function findHomelandTerms(term): ...
    function isSlur(term): ...
    function assessSeverity(term): ...
    function hasEuphemisticAlternative(term): ...
    function findEuphemism(term): ...
    function isReclaimedTerm(term): ...
    function findReclaimingGroup(term): ...
    function hasCrossCulturalEquivalent(term): ...
    function findCrossCulturalEquivalents(term): ...
```

Our study proposes new relationship mappings to enhance WordNet's capabilities in this area. These include ethnonyms, specific ethnonyms, endonymic relations, cultural context, and cultural attributes. Additionally, the research suggests incorporating relations for offensive terms, historical context, cultural variations, mixed ethnicity descriptors, diasporic variants, offensiveness levels, euphemistic substitutes, reclaimed usage, and cross-cultural equivalence.

The methodology for mapping these new relations involves analyzing existing WordNet relations and extending them with custom relations based on lexical analysis. This approach aims to create a comprehensive understanding of culturally sensitive terms within the lexical database.

The research also highlights the importance of understanding the distinctions between exonyms, caconyms, and endonyms in addressing potentially offensive language. It emphasizes that the misuse or ignorance of these terms can lead to unintended offense or perpetuation of harmful narratives, particularly in contexts related to colonialism, racism, or cultural misunderstanding.

4 Conclusion

The integration of WordNet with specialized lexicons and the proposed new relationship mappings show significant potential for improving natural language processing tasks related to cultural understanding and hate speech detection. This approach could enhance the ability of NLP systems to detect, understand, and appropriately handle culturally sensitive terms.

Acknowledgments

None.

References

Michael Adams. 2013. From Elvish to Klingon: Exploring invented languages. *Linguistic Typology*, 17(2):338–356.

Usman Ahmed and Jerry Chun Wei Lin. 2024. Deep Explainable Hate Speech Active Learning on Social-Media Data. *IEEE Transactions on Computational Social Systems*, 11(4).

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*.

Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation, LREC 2006*.

Christiane Fellbaum. 1990. English verbs as a semantic net. *International Journal of Lexicography*, 3(4):278–301.

Christiane Fellbaum. 2014. Large-scale lexicography in the digital age. *International Journal of Lexicography*, 27(4).

I. Gde Bagus Janardana Abasan and Erwin Budi

Setiawan. 2024. Empowering hate speech detection: leveraging linguistic richness and deep learning. *Bulletin of Electrical Engineering and Informatics*, 13(2).

Peter Jordan. 2023. Exonyms as parts of the cultural heritage. *Onoma*, 58.

Zainab Mansur, Nazlia Omar, Sabrina Tiun, and Eissa M. Alshari. 2024. A normalization model for repeated letters in social media hate speech text based on rules and spelling correction. *PLoS ONE*, 19(3 March).

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4).

George A. Miller and Christiane Fellbaum. 2007. WordNet then and now. *Language Resources and Evaluation*, 41(2).

I. M. Nick. 2020. Socio-Onomastics: The Pragmatics of Names. *Names*, 68(1).

A J Petiwala and S Siva Sathya. 2011. A multi-agent system to learn literature ontology: An experiment on English Quran corpus. In *IAMA 2011 - 2011 2nd International Conference on Intelligent Agent and Multi-Agent Systems*, pages 46–51.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: An affective extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004*.

Ágoston Tóth and Esra Abdelzaher. 2023. Probing visualizations of neural word embeddings for lexicographic use. In M. Medveď, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček, and S Krek, editors, *Proceedings of Electronic Lexicography in the 21st Century Conference*, volumes 2023-June, pages 545–566. Brno: Lexical Computing CZ s.r.o.

Domagoj Vidović. 2022. The treatment of demonyms, ktetics and exonyms in the more recent printed and online sources of the Institute of Croatian Language and Linguistics. *Studia lexicographica*, 16(30).

Caconym 2023. caconym, n. In *Oxford English Dictionary*.

Appendices

None

A Supplementary Material

None.