A Hybrid Architecture for Metonymy Detection in Marathi

Pratibha Dongare
The English and Foreign Languages University
pratibhaphdlandp22@efluniversity.ac.in

Abstract

Metonymy, often considered as a figurative trope, is a frequently occurring linguistic phenomenon in which an entity is replaced by a semantically related entity. Named entities are commonly used to refer to associated concepts. For instance, in the sentence *India signed a treaty*, the geographical name *India* stands metonymically for the government rather than the physical location. This study develops a hybrid architecture to classify literal and metonymic usages of named entities in Marathi language using small data. The approach integrates Pustejovsky's Generative Lexicon framework with linguistic features, including part-of-speech tags, named entity labels, and lemmas. The model is evaluated on 890 sentences and achieved F1 scores of 66.98% and 71.97% for literal and metonymic instances, respectively. The study highlights the effectiveness of the features in capturing metonymic contexts, though precision remains a target for improvement. Ablation results confirm that the Formal and Constitutive Qualia roles are the most critical components for detecting metonymic shifts, while the Telic role introduces modest noise in the present corpus. This experiment shows the scope for developing hybrid models for learning non-literal language using small data, which could be beneficial for less-explored and low-resource languages.

Keywords: Metonymy, Generative Lexicon, Marathi language

1. Introduction

Named entity recognition models often classify entities as flat and surface level entities. While this helps in classifying NEs, not all entities convey their literal sense. Metonymy, a semantic phenomenon where an entity refers to something associated with it in the real world, occurs frequently in everyday language. Surface level classification overlooks metonymy, which is crucial for various NLP tasks. A location might represent an organization or an event, shifting its semantic role based heavily on the surrounding context. For example, in the sentence *India won the match*, *India* is referring to a sports team rather than a physical location. This is a metonymic instance. The present study examines metonymic senses in NEs using Marathi language data. Marathi is an Indo-Aryan language spoken in the central-western region of India. Although several NLP studies have addressed this language, metonymy detection remains under-explored. The limited availability of annotated datasets for metonymy detection in low-resource languages such as Marathi, necessitates approaches that perform effectively with small data.

This paper evaluates and explores the effectiveness of detecting metonymic shifts in Named Entities using a small dataset. By maintaining a highly controlled annotation environment, we investigate the extent to which a model can accurately learn to distinguish between literal and metonymic readings of NEs. The next section briefly overviews metonymy detection studies. The third section discusses the methodology employed in the present study, while the fourth section presents the experimental results and analysis of the proposed small-data approach. The fifth section discusses the limitations of the present study and finally the paper concludes with potential scope for future research.

2. Related works

Metonymy has been studied in various fields including rhetorics, cognitive linguistics and NLP. In natural language processing, understanding metonymy is critical for tasks such as information extraction and named entity recognition, often framed as a classification problem where a system identifies metonymic usage of a target word within a sentence (Ghosh and Jiang, 2025). Historically, approaches to metonymy resolution have largely depended on manually curated lexical resources, including parsers, taggers, and dictionaries (Gritta et al., 2017). Fass (Fass, 1988; Fass et al., 1997) proposed a method for identifying metonymy based on selectional restrictions, while later research introduced more sophisticated statistical and machine learning techniques. Some researchers have worked on unconventional metonymy (Ghosh and Jiang, 2025) while, some researchers have focused on conventional metonymy detection which accounts for metonymic interpretations of named entities (Poibeau, 2006). Early work in metonymy resolution often focused on selectional preferences, yet this approach frequently overlooked a significant number of metonymic readings (Meguellati et al., 2022). Few scholars have studied metonymy as a classification task (Markert and Nissim, 2002a,b, 2007). Markert & Hahn (Markert and Hahn, 2002) reported 17% of metonymic in-

Proceedings of Learning Non-Literal Expressions with Small Data @ LREC 2026, pages 88–92
11 May 2026.

stances in German magazines. Hence, metonymy is prevalent and is a crucial linguistic phenomenon for natural language understanding. (Gritta et al., 2017) introduced a novel dataset, RelocaR, specifically designed to address metonymy in geographical named entities, aiming to improve upon prior annotations. Additionally, they used a novel predicate window approach which used only a small, focused segment of the sentence surrounding the entity's head dependency for classification, thereby enhancing accuracy by minimizing irrelevant input. (Kupelioglu et al., 2016) used named entities, argument structures and wordnet for metonymy resolution. Their approach integrated these diverse linguistic features to capture the nuances of metonymic expressions, demonstrating a significant advancement in automated metonymy detection. The literature suggests that while extensive resources and sophisticated models have been developed for metonymy detection, there remains a critical need for effective strategies in low-resource settings.

3. Methodology

For this study, a dataset of 890 sentences (13807 tokens) was manually annotated for lemmas, part-of-speech, named entities, and senses. Person, Location, Organization, and Miscellaneous tags were used to annotate named entities. The data was sourced from Marathi news articles covering politics, sports, finance, and current affairs domains. Annotation was performed solely by the author, following guidelines from the larger dataset under development. NEs were tagged using BILOU encoding, with each assigned a literal or metonymic sense. The dataset contained approximately 11.9% of NE tokens, of which 54.4% were literal and 45.5% metonymic.

Qualia roles from the Generative Lexicon (Pustejovsky, 1991) were mapped to NEs via a 4-bit binary vector to encode additional potential meanings beyond literal senses. Qualia roles consist of Formal, Constitutive, Telic, and Agentive roles. These roles capture distinct aspects of an entity's semantic structure and help the model search for potential novel meanings a named entity may carry in a given context. Qualia roles were implemented as deterministic, rule-based feature functions rather than learned parameters. For each named entity token, a fixed 4-bit binary vector was computed at preprocessing stage. For example, a LOCATION entity receives Formal=1 (it has a definitional physical boundary), Constitutive=0 (it is not inherently part of a containing structure), Telic=0 (it has no canonical functional purpose), and Agentive=0 (it is not the result of a deliberate causal process). Table 1 summarises

Qualia Role	**Bit**	**Encodes**	**Example rule for Marathi NEs**
Formal	1	What the entity is (type/category)	LOCATION → physical boundary entity
Constitutive	2	What it is made of or part of	ORGANIZATION → comprises people and roles
Telic	3	Its purpose or function	ORGANIZATION → has institutional goal
Agentive	4	Its causal/agentive origin	PERSON → agent of volitional action

Table 1: Qualia role encoding

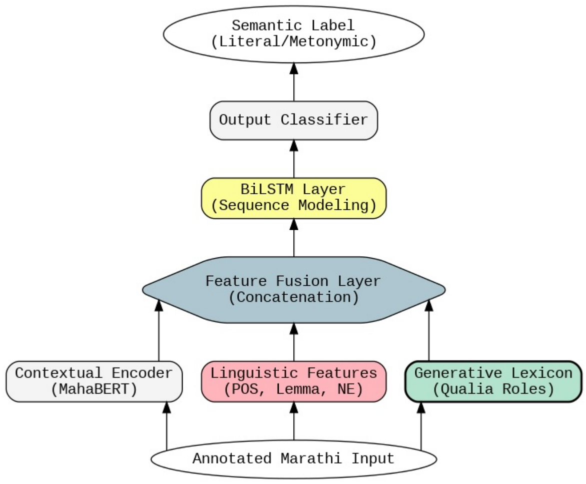

Figure 1: System architecture

the full role-to-bit mapping used in this study. The proposed system employs a BiLSTM-CRF architecture utilizing MahaBERT (Joshi, 2022) contextual embeddings, and was implemented using the PyTorch (Paszke et al., 2019) framework. figure 1 illustrates the system architecture. The annotated data is processed through the feature engineering where linguistic features and embeddings are used. Since Marathi uses a free word order, these features are important cues in detecting metonymy. The BiLSTM layer reads the data forward and backward before assigning labels such as literal or metonymic.

4. Results and Discussion

To evaluate the hybrid architecture, we examined token-level predictions. The model achieved a Literal F1 of 66.88 % (Precision: 52.99, Recall: 91.03) and a Metonymic F1 of 71.97% (Precision: 58.50, Recall: 93.48), yielding a Macro F1 of 69.42% across 170 NE instances. Table 2 shows the results. The recall is very high for both classes,

Class	Precision	Recall	F1
Literal	52.99	91.03	66.98
Metonymic	58.50	93.48	71.97
micro avg	55.87	92.35	69.62
macro avg	55.74	92.25	69.42
weighted avg	55.97	92.35	69.68

Table 2: Performance metrics of the model

Features	Precision	Recall	F1
Qualia (all)	59.9	61.2	60.5
Formal	59.9	59.3	59.6
Constitutive	66.7	60.3	63.5
Telic	68.9	70.9	69.9
Agentive	58.8	62.6	60.7
PoS	59.0	63.8	61.8
Lemma	64.1	63.7	63.9
NE	64.9	71.0	67.9

Table 3: Feature wise ablation results.

while precision is comparatively low. This asymmetry indicates that the model is broadly inclusive. It successfully recovered the vast majority of metonymic and literal instances, but overpredicted both labels. Of the NE instances evaluated, 86 were correctly identified as metonymic (true positives) and 71 as literal (true negatives), with only 3 false positives and 5 false negatives at the NE level. Notably, 114 non-NE tokens were predicted as either metonymic or literal by the model contributing to the low overall precision. Figure 2 shows the performance of the model.

4.1. Ablation Study

To assess the individual contribution of each feature, a systematic ablation study was conducted. Each feature was removed independently and the model was retrained; Macro F1 was compared against the full model baseline (Macro F1 = 61.4). Table 3 summarises the results. Several patterns can be seen from this test. Among the Qualia roles, the Formal role had the largest detrimental effect when removed (Metonymic F1 dropped by 0.063), followed by the Constitutive role (−0.053). Removing POS tags and lemmas produced only modest drops (−0.018 and −0.019, respectively), suggesting that surface morphological features play a secondary supporting role relative to the Qualia-based semantic features. The most notable finding is the behaviour of the Named Entity feature: its removal increased Metonymic F1 by 0.054. This counter-intuitive result suggests that NE labels may introduce a classification bias, causing the model to rely on entity type rather than contextual semantic cues. Metonymy in Marathi appears to be more reliably signalled by the surrounding morpho-syntactic and semantic features than by entity category alone.

The Telic role condition produced the highest Macro F1 (69.9%) and Metonymic F1 (70.9%) across all ablation conditions, indicating that removing the Telic role improves performance. Telic role encodes the purpose or function of an entity. This role introduced classification noise among the metonymic expressions in the current dataset. 114

Token	NE Tag	POS	Predicted Label	Contextual Trigger
राष्ट्राध्यक्ष (President)	O	NOUN	Literal	Trump (PER)
यांनी (By / Agentive marker)	O	PRON	Literal	Trump (PER)
हे (This / He)	O	PRON	Metonymic	Zohran (PER)
प्रचार (Campaign)	O	NOUN	Metonymic	Republican Party (ORG)
चर्चचा (Church's)	O	NOUN	Metonymic	institutional context

Table 4: Examples of non-NE marked as literal or metonymic.

non-NE tokens were tagged as either metonymic or literal. Table 4 shows five representative examples of non-NE tokens assigned a literal or metonymic label by the model, along with the contextual NE trigger that likely drove the prediction.

In the first two rows, the presence of a known Person entity (*Trump*) in the immediate context leads the model to assign a Literal label to adjacent non-NE tokens. The model correctly recognises that these tokens are functioning in their literal roles (referring to an actual individual), but their labelling as non-NE items increases the false positive count for the Literal class. The last three rows illustrate that metonymic labels are assigned to nouns and pronouns that are contextually adjacent to an Organisation or Person entity whose presence signals an institutional or representative reading. The pronoun हे (this/he) is labelled Metonymic because it is the pronominal trace of *Zohran*, a PER entity used in a representative context. प्रचार (campaign) is marked Metonymic because it co-occurs with Republican Party, an ORG entity that conventionally activates the Telic and Agentive Qualia roles. Finally, चर्चचा (church's) is labelled Metonymic in an institutional context where the genitive marker implies organisational agency. These examples explain the low precision. The model's contextual window for Qualia feature activation is wide enough to capture genuine metonymic triggers, but it does not apply strict NE-boundary. Every token inside the influence zone of a metonymically active NE becomes a candidate for labelling, regardless of its PoS category. This is a feature when it enables the model to detect metonymic effects that extend beyond the head entity and a liability when it affects the precision.

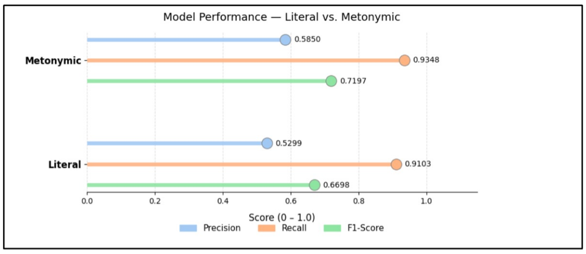

Figure 2: Performance of the model on literal and metonymic categories

5. Limitations

This study has several limitations, which are addressed in ongoing and future research.

- **Corpus size and single annotator:** The corpus size remains restricted at 890 sentences. Additionally, the dataset was annotated solely by a single domain expert; inter-annotator agreement metrics have not yet been established to measure human baseline performance on Marathi metonymy.
- **Noise:** The ablation results indicate that the Telic role introduces noise in the current corpus. Future work will investigate whether a domain-adaptive role assignment, can recover the expected benefit.
- **Integration of case markers and Selectional Preferences:** A particularly promising direction currently under active development by the author but outside the scope of this paper is to formally integrate case markers and Selectional Preference (SP) with Qualia role vectors. Marathi uses several case markers frequently. Such integration would enable the model to jointly reason over morphosyntactic evidence and structured semantic constraints during inference.
- **Named entity label bias:**The finding that removing NE labels improves Metonymic F1 suggests that flat NE type categories introduce classification bias. Future work will explore fine-grained entity tagging and entity-context interaction modelling to address this.

6. Conclusion

This study showed that metonymy detection in Marathi named entities is achievable with small datasets. The hybrid model used deterministic Qualia role vectors derived from Pustejovsky's Generative Lexicon alongside transformer embeddings to detect metonymy in Marathi text. It achieved a high recall for both literal and metonymic cases. Ablation test confirmed that all features are influential, while the NE feature introduced an unexpected classification bias. Although limitations remain, this experiment showed encouraging results. Future work will expand the Marathi metonymy corpus, integrate Selectional Preferences and case markers with Qualia-informed type coercion, and broaden the role assignment schema to improve the performance of the system.

7. Bibliographical References

References

Dan Fass. 1988. An account of coherence, semantic relations, metonymy, and lexical ambiguity resolution. In *Lexical Ambiguity Resolution*, pages 151–177. Elsevier.

Dan Fass, Allan Lesgold, and Vimla Patel. 1997. *Processing metonymy and metaphor*, volume 1. Ablex Publishing Corporation Greenwich, CO.

Saptarshi Ghosh and Tianyu Jiang. 2025. ConMeC: A dataset for metonymy resolution with common nouns. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6493–6509, Albuquerque, New Mexico. Association for Computational Linguistics.

Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. 2017. Vancouver welcomes you! minimalist location metonymy resolution. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1248–1259.

Raviraj Joshi. 2022. L3cube-mahacorpus and mahabert: Marathi monolingual corpus, marathi bert language models, and resources.

H Burcu Kupelioglu, Tankut Acarman, Tassadit Amghar, and Bernard Levrat. 2016. Recognition of metonymy by tagging named entities. *WSEAS Transactions on Computer Research*, 4:81–85.

H Burcu Kupelioglu, Tankut Acarman, Bernard Levrat, and Tassadit Amghar. Helping metonymy recognition and treatment through named entity recognition.

Katja Markert and Udo Hahn. 2002. Understanding metonymies in discourse. *Artificial intelligence*, 135(1-2):145–198.

Katja Markert and Malvina Nissim. 2002a. Metonymy resolution as a classification task. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 204–213.

Katja Markert and Malvina Nissim. 2002b. Towards a corpus annotated for metonymies: the case of location names. In *LREC*.

Katja Markert and Malvina Nissim. 2007. Semeval-2007 task 08: Metonymy resolution at semeval-2007. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 36–41.

Muhammad Elyas Meguellati, Rohana Binti Mahmud, Sameem Binti Abdul Kareem, Assaad Oussama Zeghina, and Younes Saadi. 2022. Feature selection for location metonymy using augmented bag-of-words. *IEEE Access*, 10:81777–81786.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library.

Thierry Poibeau. 2006. Dealing with metonymic readings of named entities. *arXiv preprint cs/0607052*.

James Pustejovsky. 1991. The generative lexicon. *Computational linguistics*, 17(4):409–441.