Some Strategies to Capture Kāraka-Yogyatā with Special Reference to apādāna

Swaraja Salaskar[1], **Diptesh Kanojia**[1,2,3], **Malhar Kulkarni**[1]

[1]IIT Bombay, [2]Monash University,
[3]IITB-Monash Research Academy,
{swaraja.salaskar77, dipteshkanojia}@gmail.com, malhar@hss.iitb.ac.in

Introduction

- Language by itself has an ontological structure, epistemological pinning and grammar.
- Ambiguity is a feature of natural language; In layman's terms, 'ambiguous' means 'having more than one meaning'.
- Meanings understood by human beings are based on context, background knowledge, tonal and gestural basis.
- There are mainly three types of ambiguities i.e., structural, lexical and semantic.
  - If ambiguity is present in a single word, it is known as lexical ambiguity.
  - Semantic ambiguity means the presence of multiple meanings for the same word.
  - Structural ambiguity, on the other hand, is the presence of two or more possible structures within one single sentence.
- In our paper, we observe that the sentences can be interpreted in multiple ways with the help of examples.
- In Sanskrit language every word has kāraka role to fulfil the meaning of the sentence; a single word cannot have more than one kāraka role in the same sentence.
- Similarly, each dhātu (root word) does have its own expectancy of various kārakas to complete the meaning of the sentence.

Contributions

- We create an annotation tool which allows a lexicographer/annotator to mark kāraka-yogyatā relations within the tool for a dhatu word with another word. We ensure the tool allows the annotator to delete and create new entries, along with the facility of viewing these entries.
- We study the special case of apādāna kārakas and present which play a very important role in disambiguation of Sanskrit concepts.

Background and Related Work

- In order to get rid of the preconceived notion of the yogyatā in question, Ogawa (1997) proposed that yogyatā is a notion which is originally formed in the framework of kāraka theory.
- Ramanuja Tatacharya (2006) described a collection of theories of śābdabodha as an assembly view of different sastras (nyāya, mīmāsā, vyārakarṇa, vedāmta etc.) and examines theories and subjects.
- Kunjunniraja (1968) discusses Indian theories of meanings of different schools which find yogyatā as a necessary condition for Verbal Cognition.
- We extend it further not just for the śābdabodha, but also use a database as a solution to some problems, as discussed in our paper.
- Huet (2003) report the progress in the field of computational linguistics for the Sanskrit language,and propose a solution to the tagging of verb phrases which correctly handle the non-associativity of external sandhi arising the treatment of preverb a.
- For English language, Pedersen (2006) provide a detailed description of WSD as a computational problem and describe the classical methodologies to help solve it. They detail various methods such as supervised, unsupervised and semi-supervised. They, also, describe various knowledge sources for WSD, domain specific WSD and use of WSD in various NLP applications.
- Navigli (2009) provide a comprehensive survey of the algorithms which can be used to solve WSD for NLP. They provide a detailed description of the clustering algorithms which can also be used to solve WSD for the English language.
- Khapra et al. (2008) propose iterative-WSD for English, Hindi, and Marathi in a domain specific setting.
- A projection of this work based on corpus and WordNet parameters was later performed by Khapra et al. (2009).

Methodology

- We aim to extend the ontological tag-set presented by Nair and Kulkarni (2010) and provide an exhaustive set of ontologies.
  - For *e.g.,* the current ontological tag for yānaṁ is acalanirjīva, but the proposed ontological tag in context of the root word *gam* for yānaṁ should be gamana-sādhana.
- We extend the tag-set by providing more such categories using our methodology and the tool we created.

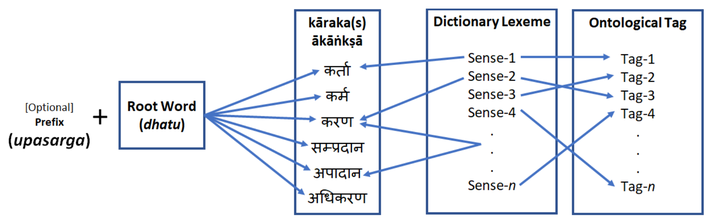

Figure 1: *Methodology Depiction*

Markup Process

- We choose a root word.
- We look for the expectation for various kārakas of the root word.
- We choose a lexeme from the lexicon.
- We tabulate various senses of the lexeme, and check for kāraka yogyatā relation of the senses with the root word.
- We mark the lexeme and its senses with kāraka yogyatā relations and store them in our database.
- We mark up the lexicon available to us with kāraka yogyatā relations between:
  - dhātu and Word,
  - dhātu and a different sense of the word, and
  - Prefix - dhātu *i.e.,* changed sense of the resultant dhātu with all senses of a word.

Yogyata Relations Tool

- We develop a tool to manually annotate a Sanskrit dictionary with such rules, and store them separately into a database.
- Our tool is an online web interface which simultaneously shows the annotator a list of prefixes, a list of dhātu (one dhātu at a time), a list of yogyatā relations, and a list of words from the Monier-Williams Dictionary (one word at a time).
- The tool requires an annotator creates rules for a pair of words, one of which is a dhātu which may or may not be perpended with a prefix.
- We call this resultant word L-word.
- On the other side, a word from the Monier-Williams dictionary is displayed which we refer to as the R-word.
- The tool has some unique features as described below:
  1. The rule to be created by an annotator requires them to mark every pair of L-word and R-word with a kāraka-yogyatā relation.
  2. We have an added functionality of appending comments along with the rule for the annotators to justify the rule, if needed.
  3. The changed semantics of the dhātu along with the prefix which results in the formation of L-word can also be submitted along with.
  4. They can also manually enter the sandhi of the dhātu and prefix *i.e.,* the final L-word in the space provided.
  5. For the annotators ease, we provide a Transliteration API on the interface so that romanized typing can be facilitated.
  6. The tool also provides the functionality to view the rules created for a particular L-word and R-word pair.
- The tool is a PHP based interface which utlizes Javascript for front-end rendering, and MySQL, as back-end database, for storing the rules created.

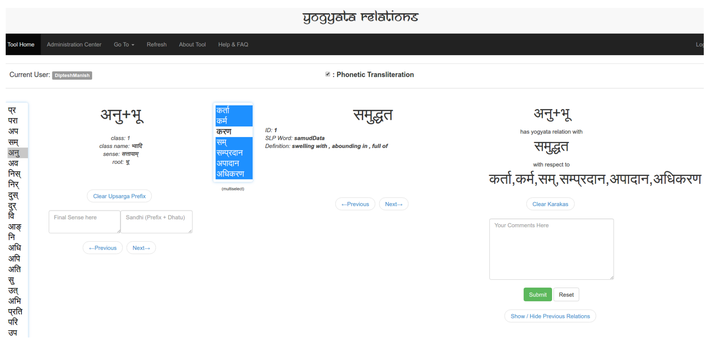

Figure 2: *A screenshot of our tool*

Conclusion and Future Work

- In this paper, we come up with a methodology for marking lexemes with karakā-yogyatā relations with a dhātu word.
- We also study the use of ontological tag-sets as a solution for the problem of WSD in NLP, and extend the tag-set previously proposed by others.
- We develop a tool for marking the Sanskrit lexicon with kāraka-yogyatā relations with root words, which stores these relations in a way they can be utilized later for resolving sense disambiguation.
- Our work proposes to resolve the issue by pruning the number of senses which are available for a lexeme and also via pruning the ontological categories which have the expectancy of a kāraka relation with a root word.
- In future, we would like to analyze and extend the ontological tag-set previously proposed by Nair and Kulkarni (2010) and mark the kāraka yogyatā relations among them.
- We also aim to annotate more dhatu-word pairs with kāraka yogyatā relations and form a database which can be utilized for solving the problem of WSD and thence for helping NLP applications such as Machine Translation for Sanskrit to other languages and vice versa.
- We also aim to use Cognitive Psycholinguistics and for verifying if yogyatā is an absolutely necessary condition for verbal cognition.
- With this, we aim to improve the state of Computational Linguistics for the Sanskrit language with the hope that this impacts other languages as well.

References

Huet, G. (2003). Towards computational processing of sanskrit. In *International Conference on Natural Language Processing (ICON)*.

Khapra, M., Bhattacharyya, P., Chauhan, S., Nair, S., and Sharma, A. (2008). Domain specific iterative word sense disambiguation in a multilingual setting. In *Proceedings of International Conference on NLP (ICON 2008), Pune, India*.

Khapra, M. M., Shah, S., Kedia, P., and Bhattacharyya, P. (2009). Projecting parameters for multilingual word sense disambiguation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 459–467. Association for Computational Linguistics.

Kunjunniraja, K. (1968). Indian theories of meaning.

Nair, S. S. and Kulkarni, A. (2010). The knowledge structure in amarakosa. In *Sanskrit Computational Linguistics*, pages 173–189. Springer.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)*, 41(2):10.

Ogawa, H. (1997). Paniniyas on yogyata and sakti. *JOURNAL OF INDIAN AND BUDDHIST STUDIES (INDOGAKU BUKKYOGAKU KENKYU)*, 46(1):508–503.

Pedersen, T. (2006). Unsupervised corpus-based methods for wsd. *Word sense disambiguation*, pages 133–166.

Ramanuja Tatacharya, N. (2006). Sabdabodhameemamsa-an inquiry into indian theories of verbal cognition part i. *Pondicherry-Rashtriya Sanskrit Sansthan, New Delhi*.

This paper draws insights from an unpublished original Sanskrit work on Yogyatā by Malhar Kulkarni; Presented at the 1st edition of the Sanskrit and Other Indian Languages Technology (SOIL-Tech), January 2019, New Delhi (India)