

DETERMINATION OF SEMANTIC RELATEDNESS

Anupama D. Sakhare

Department of Electronics and Computer Science,
R.T.M. Nagpur University
Nagpur, India
email: adsakhare616@gmail.com

Abstract

The present work focuses on Semantic automatics which is the implementation of semantic connectedness related to Karaniya Metta Sutta-Verse which is about the word [abode vihar] about the [state-of-mind] of an individual. Within the Pali language context, it is about the word vihar. WordNet® is used to bring about various semantic structures, glosses and major name-sense relations. Spectral ordering is used as a technique to reveal word connectedness as the patterns visualized are nodes and edges of a linguistic connectedness design. The obtained design has fine constructs such as – IS-A and HAS Relations which reflect the Innateness of the Pali language.

The following experiments have been performed related to the context of Karaniya Metta Sutta :

- i) Words Semantic Relatedness
- ii) Semantic Similarity of words and
- iii) Semantic compositionality in the Context of Karaniya Metta Sutta

Keywords :

Lexical Semantics, wordnet

1. INTRODUCTION

Words' semantic relatedness measures the degree to which two words are related.

The second part evaluates the word-computational phrase semantic similarity.

The third part returns the relations between words described by a set of predicates.

Contextual connectedness is related to semantics. 'semantic' usually refers to the meaning or interpretations of words or symbols. Connectedness typically implies to a relationship or linkage between entities. For example-Vihar is a Temple.

Combining such terms suggests a program or system that deals with the meaningful relationships between different elements such as words and concepts.

Lexical semantics plays a crucial role in Natural Language Processing (NLP), Computational linguistics (CL), and the development of resources such as Lexical Databases like the WordNet®

WordNet® is a large lexical database of English. In this database nouns, verbs, and adverbs are grouped into sets of cognitive synonyms together with lexical constructs from the synsets, each expressing a distinct concept. WordNet software is freely available. The structure of WordNet makes it a useful tool for CL and NLP. WordNet superficially resembles a thesaurus, however, there are some

important distinctions between WordNet and thesaurus.

First, WordNet interlinks not just word forms-strings of letters specific senses of words. As a result, words that are found near one another in the network are semantically disambiguated second, WordNet labels the semantic relations among words in a pattern, whereas the groupings of words in a thesaurus do not follow any explicit pattern other than meaning similarity.

Large Language models, such as GPT-3 (Generative Pre-trained Transformer) have demonstrated impressive capabilities in understanding and generating human-like response text. However, these models are built on the transformer architectures, which do not explicitly encode lexical semantics in the way as databases particularly lexical semantic databases like WordNet.

Prompt engineering is the process of structuring text that can be interpreted and understood by a generative AI model. A prompt is natural language text describing the task that an AI should perform Example prompts:

1) When we send a message such as [gloss of word itinerary] GPT gives the following output.

(The word itinerary refers to a detailed plan or schedule of a journey.)

Despite presenting Quality-based filtering and other features, the output is not satisfactory.

Thus a need is felt to find the contextual connectedness pattern of Pali words in English by making use of WordNet®

2. LITERATURE REVIEW:

In [1] R. Siblani and L. Kosseim used Clac to find semantic Relatedness of words and phrases.

The same authors in their paper [2], “Using a weighted Semantic Network for Lexical Semantic Relatedness” formed a semantic connection network and also evaluated – the sense and reverse-sense of two words.

Language changes due to many reasons. Some of the known reasons as mentioned in the linguistics study are due to invaders and rulers with the interaction of a particular language in a Region. The growth of English vocabulary is due to initial Celtic influence and the cultural life of the Celtic people who were among the earliest inhabitants of England. Celtic people were peace-loving people and their language has left its mark on English in some vocables such as “Devonian” and also in numerous place and palace names.

Latin, Germanic, Scandinavian, French, and Greek languages have made remarkable influence and contributions to the present English language.

Some important personal contributions were made by writers such as Thomas Browne the writer of *Religio Medici* who introduced words like

hallucination, insecurity, precarious incontrovertible, and antediluvian.

In his *Ortus Medicinæ* or *Rise of Medicine*, the Dutch chemist J.B. Van Helmont States that he coined the word [gas] on the model of the Greek word [Chaos].

In the present work, semantic relations and name sense relations of the Pali words are considered for finding the following relationships.

1. Hyponymy
2. Synonymy
3. Antonyms
4. Polysemy
5. Homonymy
6. Meronymy
7. Troponymy and
8. Pertainymy

3. METHODOLOGY

Online experiments have been performed on :

- i) Words Semantic Relatedness
 - ii) Semantic Similarity and
 - iii) Semantic Compositionality
- Using Clac Software.

Spectral ordering is the technique which is used to reveal patterns of some given words.

In Natural language processing (NLP) spectral ordering can be applied to various tasks that involve analyzing the connectedness or relationships between words, documents or entities within a linguistic structure.

Spectral ordering can be used for

- (1) Document clustering
- (2) Word Embeddings and similarity
- (3) Modeling Language to represent language
- (4) Named entity Recognition and
- (5) Topic Modeling

The following functional approach is used in the paper. This functional approach is based on Margaret Berry's diagram fig. 1 is about contextual connections.

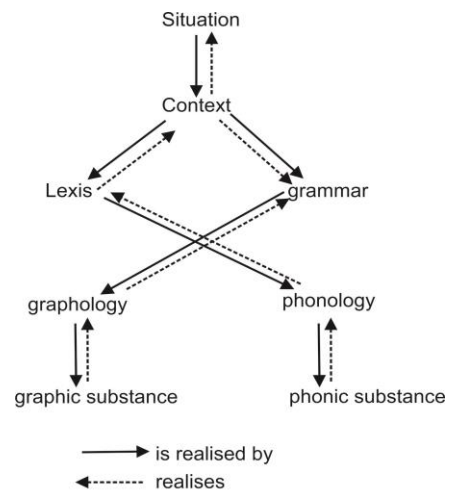


Fig. 1 Margaret Berry's contextual connections diagram

Using this approach we illustrate a sketch of semantic connectedness. As shown in Fig. 2 is related to the Pali word vihar.

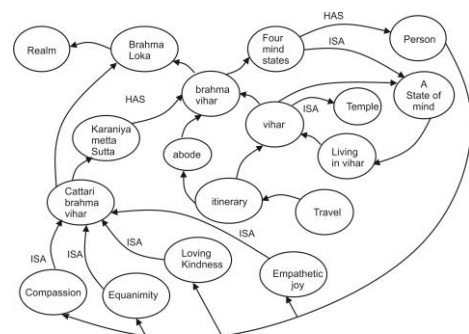


Fig. 2 Sketch of Semantic Connectedness about Pali word Vihar

Table : Semantic Connections

Sr. No	Order of Semantic Structure
1.	{Brahma loka, Realm}
2.	{brahma vihar, Brahma loka}
3.	{Brahma vihar, four mind states, person}
4.	{person, joy, loving kindness, equanimity, compassion }
5.	{cattari brahma vihar, karaniya metta sutta}
6.	{ empathetic joy, cattari brahma vihar}
7.	{loving kindness, cattari brahma vihar}
8.	{equanimity, cattari brahma vihar}
9.	{vihar, temple}
10.	{Travel, itinerary, abode, vihar}
11.	{Travel, itinerary, vihar}
12.	{Travel, itinerary, vihar}
13.	{State of mind, living in vihar, vihar}

The aforementioned table gives the semantic connections applied in the semantic automatics

Program for Homonymy

```
import nltk
nltk.download('wordnet')
from nltk.corpus import wordnet
def get_homonyms(word):
homonyms = set()
for synset in wordnet.synsets(word):
for lemma in synset. Lemmas():
homonym = lemma.name()
if homonyms != word:
homonyms.add(homonym)
return homonyms
def print_homonyms(word):
homonyms=get_hononyms(word)
if not homonyms: .print(f "No homonyms
found for 'word' ")
```

else:

```
print (f "Homonyms of '(word)' :
(',' .join (homonyms))")
```

#Example

```
Word_to_lookup = "bye"
```

Print homonyms (word to lookup)

```
[ 'bye', bye-bye, Chario, au_revoir, pass,
goodby, so_long, sayonara]
```

4. RESULTS

The verses of Karaniya Metta Sutta (English Version) have be selected to determine Semantic Relatedness of two words.

To compute Semantic Relatedness in the 2-lexicons of first sections of Karaniya Metta sutta

First word: Calm

Second word: Gentle

Semantic Relatedness: 91.00%

Sense : [1.01]

Assuage, Gruntle, Appease, Pacify, Conciliate, Lenify, Mollify, Placate, Gentle

Hypernum : [0.50]

Calm, Lull, still, calm-down, tranquilize, quieten, tranquilize

Table -1 Semantic Relatedness

First word	Second Word	Semantic Relatedness
Calm	Humble	78%
Calm	Gentle	91.00%

Reverse Sense : [1.01] calm

Total cost = 2.52

5. DISCUSSION

- 1) The two words (Lexicons) word-1 and word 2 must be carefully chosen

2) Network Representation of semantic relationships can be obtained.

Table – 2 Percentage of Semantic Relatedness example :-

Section	Word-1	Word-2	Percentage of Relatedness
1	Ship	Boat	90%
2	Ship	Car	69%

6. CONCLUSION

The article has made use of WordNet and Wordnets for implementing various lexical constructs using the English language lexical Database. We have worked upon the spectral ordering which is used to reveal interesting patterns found in Pali words. The connections of the Pali words with the known English terms are identified in the form of Word sense, Gloss, Node identity, ISA, HAS and PART-WHOLE relations. This article is anticipated to serve the researchers of Pali Computational Linguistics in order to develop better Pali Wordnet.

7. LIMITATIONS

1) The conclusion may prove to be misleading if the results are not interpreted properly.

2) Semantic Evaluation may be incorrect at times.

8. FUTURE SCOPE

Results of the NLP experiments performed on wordNet and ClaC software may be compared in order to develop new software.

REFERENCES

- [1] Reda Siblimin and Leila Kosseim, "Using a weighted Semantic Network for Lexical semantic Relatedness", in the proceedings RANLP2013.
- [2] <http://www.aclweb.org/anthology/S13-2019>
- [3] Fellbaum, Christiane, "WordNet and wordnets." In: Brown Keith et al. (eds.), Encyclopedia of Language and Linguistics, Second Edition, Oxford : Elsevier, 665-670. (2005)
- [4] Anupama Devendra Sakhare et al, J Phys. : conf. Ser. 1913 012 097 "Pali Lexicography using lexonomy" 2021.
- [5] M. Mechura, "Contributions to e-Lexicography".
- [6] B.A. y Arcas, "Do large language models understand us ?" Daedalus Vol. 151, No. 2, pp-183-197, 2022.
- [7] "Prompt Engineering Guide", 2022.
- [8] W.X. Zhao, K. Zhou, J.Li. T. Tang, X. Wang, Y. Hou, Y. Min. b. Zhang, J. Zhang, Z. Dong et al., "A survey of large language models", arXiv : 2303. 18223, 2023.
- [9] R. Sennrich, B. Haddow, and A. Birch, "Neural Machine translation of rare words with subword units", in proceedings of the 54th Annual Meeting of the

- Association for Computational Linguistics (Volume I : Long Papers), 2016 pp. 1715-1725.
- [10] S.J. Mielke, Z. Alyafeai, E. Salesky, C. Raffel, M.Dey, M. Galle, A. Raja, C.Si, W.Y. Lee, B. Sagot et al. "Between words and characters : A brief history of open-vocabulary modeling and tokenization in nlp", arXiv : 2112.10508, 2021.
- [11] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick barnes, Ajmal Mian,"A Comprehensive Overview of Large Language Models", arXiv : 2307.06735v7 [cs.CL] 27 Dec 2023.
- [12] Malhar Kulkarni, Abhishek Nanda, Chaitali Dangarikar, Irawati Kulkarni, Pushpak Bhattacharyya."Introducing Sanskrit Wordnet" Proceedings on the 5th Global wordnet conference (GWC 2010), Narosa, Mumbai 2010.
- [13] S Walawalikar, S. Desai, R Karmali, S. Naik irgu.unigoa.ac.in, "Experiences in building the Konkani wordnet using the expansion approach" unigoa.ac.in, 2010.