

Introduction to ProverbNet: An Online Multilingual Database of Proverbs and Comprehensive Metadata

Shreyas Pimpalgaonkar, Dhanashree Lele,
Malhar Kulkarni, and Pushpak Bhattacharyya
Indian Institute of Technology Bombay, India
shreyas.pgkr@gmail.com, dhanashreelele01@gmail.com
malhar@hss.iitb.ac.in, pb@cse.iitb.ac.in

Abstract

Proverbs are unique linguistic expressions used by humans in the process of communication. They are frozen expressions and have the capacity to convey deep semantic aspects of a given language. This paper describes ProverbNet, a novel online multilingual database of proverbs and comprehensive metadata equipped with a multi-purpose search engine to store, explore, understand, classify and analyse proverbs and their metadata. ProverbNet has immense applications including machine translation, cognitive studies and learning tools. We have 2320 Sanskrit Proverbs and 1136 Marathi proverbs and their metadata in ProverbNet and are adding more proverbs in different languages to the network.

1 Introduction

According to Norrick (2015), proverbs are self-contained, didactic linguistic units, and their content is usually fixed and poetic. Proverbs contain the same words as ordinary sentences, but are frozen expressions and have a non-compositional meaning, created by humans and handed down through generations (Meider, 2004). Proverbs can be used to present facts, refer to human socio-cultural events, give advice, and critique behavior (Lubis, 2019). Understanding of metaphorical ideas and a priori knowledge is essential for getting from the literal meaning of the proverb to its hidden and intended meaning. The crucial task of machine translation fails to translate proverbs because of the inability to detect proverbs automatically and incorrectness of direct translation. As proverbs are just like any other sentences, it is difficult to detect them in textual corpus automatically. There

are a few proverbs that have counterparts in other languages, and sometimes we can obtain them via direct translation, but many times, they are entirely different. Proverbs can be categorized into various classes and categories (Lauhakangas, 2001). They are usually static multi-word expressions having specific keywords. In most cases, we cannot change even a single word in the proverb. If done, then it becomes a simple sentence without an accepted non-compositional meaning.

For all languages in the Indo-European family in India, Sanskrit is considered to be linked to them historically. Sanskrit is one of the oldest documented members of the Indo European family (Woodard, 2008). Since it is an ancient language, the wisdom embedded in Sanskrit proverbs has been carried forward for thousands of years. However, nowadays the use of proverbs has reduced even in the native spoken languages, let alone the use of proverbs in the Sanskrit language. To preserve these wisdom-packed proverbs, we require a digital database with the analysis of the proverbs. Some compilations are found having internet base. However, the collections of Sanskrit proverbs found on the internet are comparatively small and limited only to the famous proverbs. People refer to these compilations when they are in want of a particular proverb, or the proverbs related to a particular theme or concept. For such reference, a mere compilation does not suffice. There should be an online database in a searchable form. As far as Indian languages are concerned, there are no databases of proverbs available and our tool removes this desideratum, and this ProverbNet may act as a single reference point.

2 Literature Survey

To the best of our knowledge, no online multilingual database of proverbs with comprehensive metadata focused on Sanskrit and other Indian Languages exists. We, therefore present a survey of related literature – the sources from which we compiled proverbs, proverbial analysis, cognition, multi-word expressions, wordnets, online proverb databases and possible applications.

There are many offline collections of proverbs. *Ābhānakajagannātha* by Jagannath (Jagannath, 2009) is a compilation of new proverbs in Sanskrit. Manwaring (1899) has collected, translated and classified a total of 1910 Marathi proverbs. *Bharatiya Kahavat Sangraha* by Naravane (1978) is a collection of proverbs in three volumes of proverbs in fifteen Indian Languages and can be called as the *Gītā* of Indian Proverbs. Bhosale (2016) has composed a collection of proverbs from Marathi language which may go into oblivion soon and has tried to preserve these proverbs. Such collections are essential in preserving the treasure of language.

A vast amount of literature exists on proverbial analysis. Several books (Honeck and Temple, 1994; Mieder et al., 1994; Honeck, 2016; Bhagwat, 1985; Brough, 1953) and articles (Bronkhorst, 2005; Temple, 1999; Swinney, 1979) provide deep insights into the journey of the meaning of a proverb from literal meaning to its suggestive meaning. Meider (2004) explores definitions, classification, origin, dissemination, collection, and various case studies of proverbs. Grzybek (2014) presents a semiotical study of proverbs in pragmatic, syntactical, and semantical dimensions. A psychological study suggests that the ability or inability to explain meanings of proverbs indicates the presence or absence of abstract thinking abilities, and a strong understanding of proverbs reveals essential insights into how people conceptualise metaphorical ideas (Gibbs and Beitel, 1995). According to Ferretti et al. (2007), people should be able to understand the non-compositional meanings of proverbs directly without having to first think about and reject their literal meanings.

Proverbs belong to a class of multi-word expressions. Therefore, research conducted

for the identification and analysis of multi-word expressions can be applied to the analysis of proverbs. Tsvetkov and Wintner (2014) present a survey on analysing multi-word expressions and provides a more in-depth understanding of tasks related to multi-word expressions like discovery, identification and translation.

When it comes to creating and populating proverb databases, automatic identification of proverbs from a textual corpus is beneficial. Quite a few ways of automatically identifying proverbs exist. Rassi et al. (2014) identify proverbs and their variants by creating a finite state automaton with thirteen commonly found syntactic patterns in proverbs and then test on Brazilian Portuguese journalistic corpus. Sidhu et al. (2010) describes a word-based algorithm that splits input text and known proverbs and to find proverb variants in a given text and uses this to perform machine translation. Garg and Goyal (2014) perform automatic extraction of idioms, proverbs and their variations from text using statistical approaches.

A significant amount of research has been done in the creation of lexical databases. Princeton WordNet (G. A. Miller, 1990) is a massive database of English lexicons. Sanskrit wordnet (Kulkarni et al., 2010) is more than just a dictionary and gives different relations between synsets which represent unique concepts. Indo wordnet (Bhattacharyya, 2010) is a wordnet in multiple Indian Languages. These wordnets are extensive, but their focus is on single words and not multi-word expressions or proverbs.

Online proverb databases usually attempt at aggregating proverbs of one or two languages. Lauhakangas (2013) created a multilingual database of proverbs in European Languages. Our database is focused towards Indian languages, contains more metadata, flexible categorisation, more features and addresses a different set of goals.

3 Main Contribution

We introduce ProverbNet, an intricate network of multilingual proverbs equipped with a multipurpose search engine that contains extensive metadata and detailed cognitive analy-

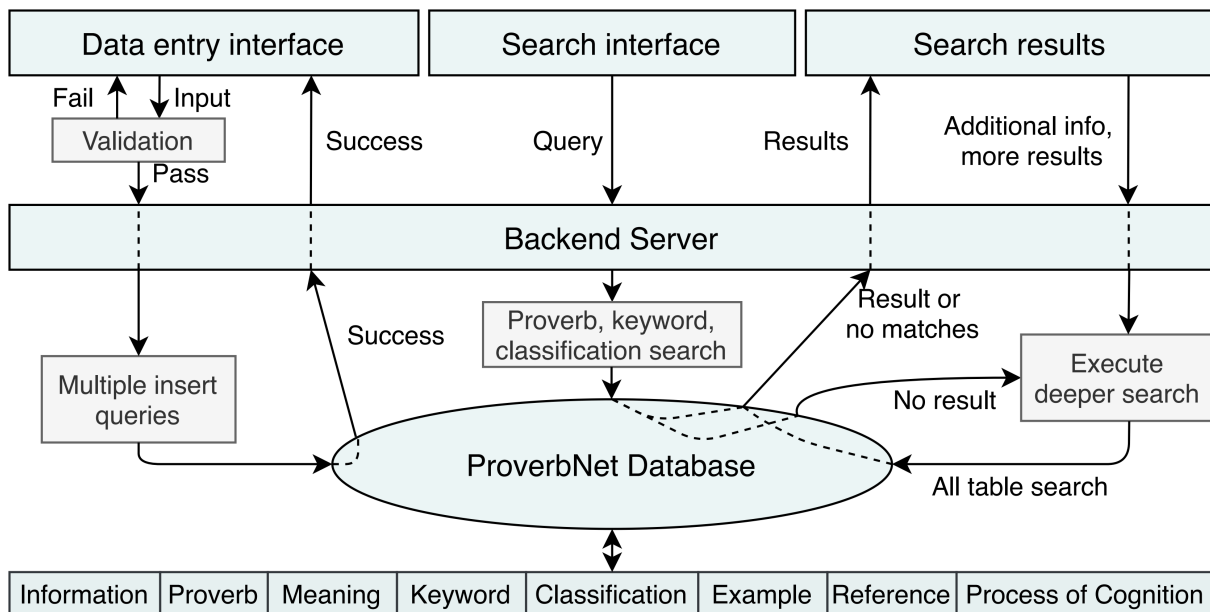


Figure 1: ProverbNet Block Diagram

sis. We gather inspiration from Sanskrit WordNet (Kulkarni et al., 2010) and Indo WordNet (Bhattacharyya, 2010) to design the interface. In order to populate this database and ameliorate the data entry process, we have designed an easy to use and elegant data entry interface. We have also designed a search engine to perform queries on this database. The block diagram of the system is shown in figure 1. We will now explain the system, and describe the details of various fields and their importance in the next section.

4 ProverbNet Database and Data Entry Interface

We have designed ProverbNet to contain comprehensive metadata about proverbs. We choose and present the data such that it benefits both, an amateur reader and a proficient researcher in Paremiology. The data entry interface (figure 2) is fast, responsive, and elegant. The interface is written in React, a JavaScript framework, and uses material-UI components. Currently, we have 2320 Sanskrit proverbs in the ProverbNet database. Amongst them, 1000 proverbs frequently occur in typical Sanskrit dialogues. Rest of the proverbs are from *Ābhāṇakajagannātha*, a book that comprises of 1320 new and rare Sanskrit proverbs. We also have 1366 Marathi proverbs, and we are adding proverbs of other

languages to the database.

The Data Entry Interface enhances the data entry experience by providing various features like language selectors for appropriate data fields, phonetic transliteration from English to any of the selected languages, and a virtual keyboard to type in the desired language. When the phonetic transliteration option is selected, and a language for a particular text box is chosen, one can type a word in English and press space to get the phonetic transliteration of the word into that language. If the transliterated word is incorrect, the user can press backspace to get more word recommendations, and they can replace their word accordingly. A virtual keyboard of the selected language appears under each text field when the option is enabled. The interface has different types of input fields depending on the type of data. On clicking the submit button, the data gets validated and then gets entered into the database. Now we explain the importance of the fields present in our database and the data entry interface.

Proverb: The user can enter a proverb in their language of choice.

Literal Meaning: The literal meaning of the proverb, preferably in English. It is what someone would say if they considered it as a typical sentence without any figurative meaning.

ProverbNet - सूक्तिबन्धः - Data Entry interface

Transliteration
 Virtual Keyboard

Data Fields

ALL

MEANINGS

TYPE OF PROVERB

KEYWORDS

CLASSIFICATION

OTHER INFORMATION

PROCESS OF COGNITION

EQUIVALENT PROVERBS

EXAMPLES

COMMENTS

Proverb:

Literal Meaning in English:

Secondary Meaning:

Suggestive Meanings:

+ -

Figure 2: Data entry interface

Secondary meanings: Most of the times, the literal meaning of a proverb does not disclose the actual meaning it wants to convey. Hence, after following the literal meaning of the proverb, we have to advance to its secondary meaning, lakṣaṇā.

Suggestive meanings: A proverb is not uttered in vacuum or isolation. A proverb is almost always a reactionary utterance to either refer to something that has happened or to refer to something that will probably take place. So knowing the context in which we use the proverb becomes vital in understanding the essence of a proverb. The same proverb can be uttered in multiple situations by different speakers. This subjectivity is not covered in secondary meaning (Lakṣaṇā) as secondary meaning limits itself to figurative meaning. Hence a third ṛtti is required to reach the essence of a proverb. This third ṛtti is termed vyañjanā, [the suggestive power (meaning)] in Sanskrit linguistics. The metaphorical meaning takes into consideration how a proverb yields itself to multiple situations. Understanding, the literal and figurative meanings of a proverb, is only a part of comprehending the proverb completely. The comprehension of a proverb as a whole (and not just the meaning) will be complete only when the listener can relate the proverb

to different situations. Therefore the suggestive meaning expresses the essence of the proverb. Literal, figurative, and suggestive meanings (i.e., Abhidhā, Lakṣaṇā, Vyañjanā) of a proverb are present in the ProverbNet. Proverbs are ‘frozen expressions’ with universally accepted fixed meanings hence the variations with reference to context will be minimum.

Cognition of the proverb: Most proverbs may allude to some universal truths or common human behaviours. Thus like morale of a story, cognition of proverb gives morale of the proverb wherever applicable.

Keywords and keyword categories: Lauhakangas (2001) describes classification of proverbs into different categories. For better clustering of proverbs and search, we enter certain keywords of those proverbs and define categories. We select the important words from the proverb or their synonyms as keywords. Further, we classify these keywords into emotions, animal names, body parts, and similar categories. This facilitates the search of proverbs containing specific words.

Type of proverb: The nature of proverbs are classified into five broad categories, viz. observation, criticism, comment, suggestion, advice. Based on the type of the proverb, the user may select zero or more of these types.

Observations: The user may enter any specific observations regarding the use of the proverb. Sometimes proverb and its literal meaning may lead to a positive outcome, but the action is aimed at the elimination of harmful things. For example, Mule Kuthār | The English translation of the proverb is ‘Nipping in the bud’, which means destroying a thing from its root. While we normally perceive destruction as bad, if the outcome is positive, the action may still be termed as good. Since the proverb suggests destruction, the observation for this proverb is ‘negative polarity’. Another example of negative polarity is Vināshkale Viparitbuddhi | Here the words have a negative tone, and the outcome is also negative. On the other hand, carāti carato bhaga:| is a proverb having neither a negative tone nor a negative outcome. The observation for this proverb is ‘positive polarity’.

Examples: The user may enter examples of the proverb in a contextual paragraph. The understanding of the use of proverb and its nuances grows with repeated exposure to a proverb. Therefore, examples would help the reader understand the use of proverb and show a possible usage of the proverb as well.

Reference: The user may enter the source of a proverb if available. In Sanskrit, many proverbs originate from Sanskrit Subhāṣitas. Many Subhāṣitas have a structure where the first line or first three lines provide an example, and the last line summarises the observation. The last line is commonly used as proverb rather than quoting entire shloka. Many of us are not even aware that a particular *sūkti* is part of a verse. If we understand the whole verse, it becomes easier to understand the essence of that *sūkti* on the backdrop of the thought/ idea conveyed in the whole verse. E.g. Maunm sarvārthsadhanm | keeping mum or remaining silent is beneficial. It is the last carāṇa, i.e. line of the following verse.

Ātmno mukhdosheṇ badhyante shuksārikāh |
Bakāstatra na badhyante Maunm sarvārthsadhanm ||

Parrots are caged because of their quality of imitating the human language in a sweet voice, but cranes are not caged because they do not speak. Therefore remaining silent is always beneficial (keeps you free). When we understand this *sūkti* on the backdrop of parrots being caged and the cranes remaining free, we

experience the specific shade of this proverb. The same is with a philosophical *sūkti* which has been drawn on the Bhagavadgītā, Upaniṣhad, Bhāgavatapurāṇa. Thus taking into consideration the background or the source of the *sūkti* helps us in extracting scent of the meaning from the proverb.

For some proverbs, there is reference to a story. Knowledge of the story is essential for complete cognition. A case from English language is ‘David v/s Goliath’. While a reader may comprehend it as a tussle between two persons, if the story of David and Goliath is known, one understands it as a fight of a small person with a mightier one. It is better to be aware of the story behind the shloka to comprehend the proverb completely.

For example, the complete shloka of the proverb Fatātopo Bhayankarh | is

Nirviṣeṇāpi sarpeṇ kartavyā mahati fanā |
Viṣamstu na vāpyastu fatātopo Bhayankarh ||

However, even understanding the śloka does not complete the comprehension until the story is told. Thus, the proverb may become part of the regular vocabulary, but the śloka may not. Here, we have given the complete subhāṣita/śloka and its source wherever available. Related context and references are also provided where necessary. This helps in better cognition of the proverb.

Parallels in other languages: One can enter proverbs that have equivalent or similar meanings in other languages. Historically, proverbs from Sanskrit can be said to have come down into modern Indian languages sometimes with variations. It is also observed that proverbs in different languages express a similar concept. Such similar proverbs may help a reader better understand the nuanced meaning of a proverb in Sanskrit. It will also throw light on how a proverb may have undergone change when it moved from Sanskrit to other language. An example is illustrated in figure 5. There is already an existing book of parallel proverbs in Indian languages spanning more than 1000 proverbs across around 10 Indian languages including Sanskrit (Naravane, 1978). We will be using it as a resource for building the database. However there is no metadata i.e., it is not searchable for similar proverbs, opposite proverbs, keyword search etc. Along with these features the ProverbNet explains

primary, secondary and suggestive meanings of the proverbs and it also contains the examples of the usage.

Specifics: Wherever applicable, the user can add specific references to Mythology, History or Culture. For example, Ishvarechā baliyasee | is related to Indian mythology due to reference to God Shankara. The proverb naro vā kuñjaro vā is related to history as it has reference to a story from Mahābhārata and we can say that the usage of this proverb has started after Mahābhārata. The proverb vasudhaiva kuṭumbakam has specific reference to Indian culture as this thought encompasses the essence of Indian philosophy.

Process of cognition through the process of Shābdabodha: The following kārīkā tells us the process of Shābdabodha.

Padjnyānm hi karaṇm dvāram tartar padārthdhi |
Shābdabodh falam tatra vṛttidhi sahakāriṇi ||

It means: (padaññāna) cognition of the word (maybe the sound or the alphabets) is an instrumental cause (the most effective means in Shābdabodh). Cognition of the meaning of the word (padārthdhi) is the gateway of the śābdabodha, vṛtti helps us to lead to the final goal that is śābdabodha, and the result is the verbal cognition. In order to cognise a proverb through the theory of śābdabodha, we analyse proverbs through each of these aspects. This theory is not only limited to Sanskrit, and we can apply this theory to other languages as well. Following are the steps in the process of cognition.

- Padaññāna - cognition of words of the sentence
- Padārthjnyānm – After identifying the words of the sentence by their sounds or by the sequence of alphabets, we understand their meanings.
- vṛtti - (ṣakti, lakṣaṇā, vyañjanā) to understand the meanings of the word we have to rely on the ṣakti or the primary meaning or the denotative power. However, when meanings are incompatible, we have to go to the secondary meaning that is lakṣaṇā. Vyañjanā is the suggestive meaning of the sentence. The literal meaning of an utterance is only a part of its whole meaning. The suggestive power

helps us to go beyond the primary and the secondary meaning.

- Saktigraha - We consider different ways of ṣaktigraha to understand the ṣakti.
- Akaṅkṣā, Yogyatā, Sannidhi - These are three conditions, meaning mutual expectancy of words, compatibility of words and proximity of words respectively. When these three conditions are fulfilled, cognition of the text is done successfully.

If a proverb yields to all these aspects, then we say that the proverb is cognisable through the theory of Śābdabodha. We illustrate this process in detail with the help of an example in Appendix A.

Classification: There are various ways to classify proverbs. Classification of proverbs by any criteria is a complicated question, and no comprehensive and standardized solution exists. Each type of classification has its validity, its practical uses, and also its drawbacks. [Kusi \(1972\)](#), a Professor of Finnish and Comparative Folk Poetry Studies and paremiologist found three aspects, i.e. the idea, structure, and basic core necessary for the classification of proverbs. He proposed that the proverbs with a common idea would be called synonymous proverbs, those with the common scheme would be called proverbs with shared structure and proverbs with images and phrases having the same meaning would be called proverbs with a common basic core. We use and extend his classification to adapt to Sanskrit and other languages. We classify proverbs in the following broad categories. Subcategories will be explained in Appendix B.

1. Proverb Structure: A proverb is an expression or a comment on a particular situation. The lesser the words, the stronger the essence. Thus the structure of the proverb has a great significance in conveying the expression effectively.
2. Proverbial Formulae: All languages possess certain structural formulae that exhibit a high degree of peculiarity towards the proverb as a linguistic form ([Coinnigh, 2015](#)). In Sanskrit proverbs, the following proverbial formulae are widely found: Itah.. tath, yatra.. tatra, yathā.. tathā,

yādr̥ṣm.. tādr̥ṣm, x vā.. y vā, x vā.. y na vā, na x na y, yah/yā.. sah/sā

3. Themes: Proverbs have a situational significance, and hence the circumstances for which we use proverbs show a wide range of themes. Therefore, theme becomes an essential basis for classification of proverbs. In 'International type system of proverbs', Matti Kussi provides a framework of specific themes for classificational international proverbs. Using Matti Kussi's framework as a base, this ProverbNet will explore and contribute to the thematic classification of proverbs

Now, after the data is recorded in the data entry interface, it gets added to the database after performing several validity checks. We have modelled the ProverbNet architecture as a relational database with separate tables for appropriate data elements. Every data entry has a distinct reference ID, is the primary key of the information table, and increments automatically. We put the single-valued entries, namely, literal meaning, secondary meaning, type of proverb, cognition, references, specific observations, historical specifics, social and cultural specifics, lingual specifics, and each of their languages in a single table called the information table. We store the remainder of the data elements, i.e. the ones that have more than one multiplicity, namely proverbs, suggestive meanings, the process of cognition, keywords and categories, classification, and examples in separate tables. This arrangement of data is a natural choice and benefits the search algorithm, which we will explain in the next section.

5 Search

We have designed a search interface (figure 3) to search for proverbs in the desired language. The search takes input a single or a multi-word query in a language specified by the user. The search interface has a transliteration option to type in English, and the text automatically gets transliterated to the selected language. It also has a virtual keyboard to type in the desired language. On submitting, the interface sends the search request to the back-end server for processing. On getting the response, we dis-

play the search results in a convenient format inspired by the Indo Wordnet (Bhattacharyya, 2010). The search interface has buttons to filter output proverbs by language. The proverbs are displayed as cards and have a button to expand them and display their stored metadata. It also has a feature to perform an in-depth search and get more results. Users can also specify the type of the query to search specific metadata.

User searches can vary from getting a proverb by entering a part of it to find the processes of cognition of proverbs related to a specific theme. One can search for proverbs by typing in the entire proverb (bhavitavyānām dvārāṇi bhavanti sarvatra |). If they do not know the entire proverb, they can type a few words of the proverb as well (bhavitavyānām bhavanti). If they want to find out a proverb with a given meaning, they can type the meaning of the proverb and get the proverbs that have the entered meaning. The meaning can be literal, secondary or suggestive (No one can change the destiny, or destiny is inevitable). Therefore, ProverbNet acts as a thesaurus of proverbs as well. One can search for proverbs containing specific keywords (bhavitavya, dvāra) or their categories (fate). To get complete śloka, or get proverbs written in specific reference, or get entire proverbs by a particular author, one can enter the reference name or the author name (Śhākuntalam, Kālidāsa). Users can get proverbs that have similar processes of cognition, classification, and other metadata.

The output of a sample query is shown in figure 3. The search currently uses a two-step algorithm to get the relevant output. We assume that the user will generally enter keywords, classification, or a part of the proverb as the search query. The first step of the algorithm is to match proverb prefixes (includes full proverb) with the search query, match exact keywords or categories, and classification. Make individual queries on each table and collate the results. If there are matches, return them. If no matches are present, execute the second step of the search. Another way to trigger the second step is to click the button called 'more results' provided in the search interface. The second step is to per-

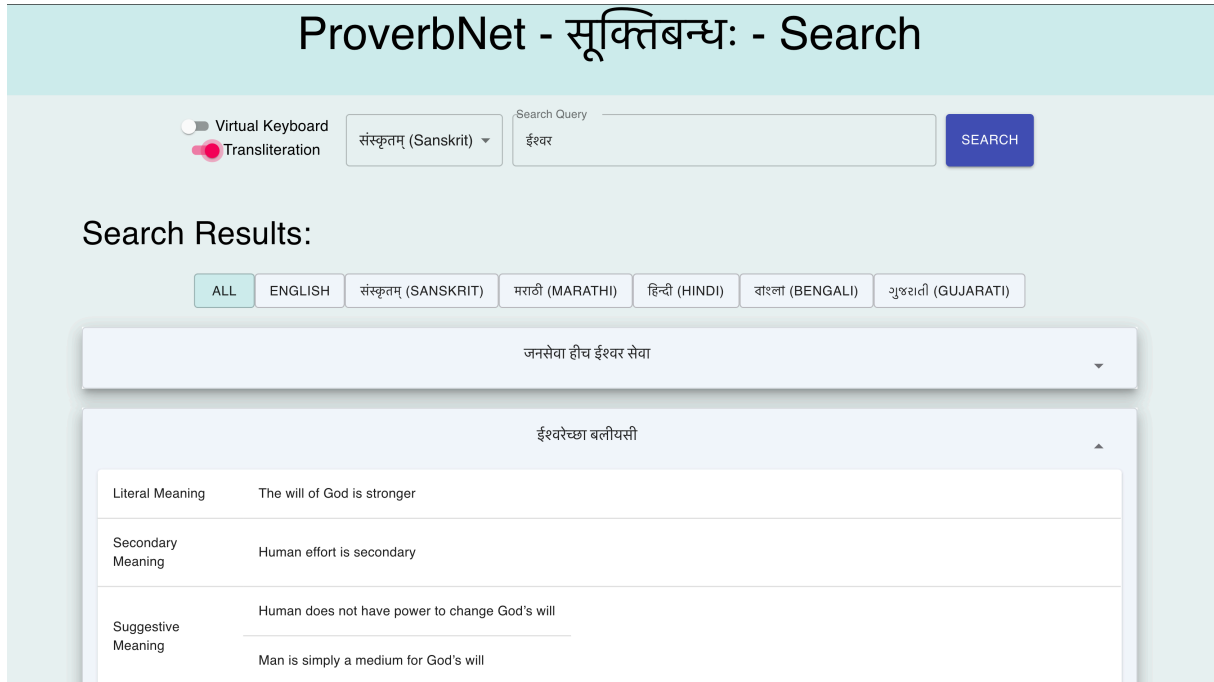


Figure 3: Search interface

form a broader search. We break the search query into individual words and remove certain words (e.g. a, and, the), search each word in all of the fields of the database, and return those proverbs not returned in the previous step. This step yields a more extensive collection of proverbs, and they are lesser relevant to the search query than those returned in the first step. We also have another selector in the interface that to specify the type of search query if needed (e.g. meaning, classification) to get more relevant and narrower search results. Selecting the option enables search in only those specific fields.

6 Applications

There are quite a few possible applications of ProverbNet. ProverbNet can be used as an online resource for authors and readers to understand proverbs, as the literal translation may not lead to a comprehensive understanding. Furthermore, The current state-of-the-art systems fail to translate proverbs correctly, and there is a need for such a multilingual resource for correct automatic translation of proverbs. Also, ProverbNet will help get closer to solving the problem of word sense disambiguation. ProverbNet can also become a learning resource as an educational applica-

tion. Finally, ProverbNet will be useful to get equivalent proverbs in different languages in one place.

7 Future Work

We are adding proverbs from other languages to ProverbNet. The design of ProverbNet ensures little changes to add newer languages. We plan to implement deep learning models to identify proverbs and use ProverbNet for automatic translation. We are also working towards clustering the different entries to create newer entries, so the equivalent proverbs in different data entries get clubbed together. For users having difficulty reading the words in the language present in the database, we plan to add automatic transliteration and translation of metadata in different languages.

References

- V.B. Bhagwat. 1985. Paniniy vyakran ani bhasha tatvdnyan.
- Pushpak Bhattacharyya. 2010. Indowordnet. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

- D. T. Bhosale. 2016. Lopalelyā suvarṇamu-drā: Haravata cālalelyā arthapūrṇa mhaṇī. *Sakala Papers Private Li., Pune.*
- J. Bronkhorst. 2005. [Bhaṭṭoji dikshit on sphota](#). *Journal of Indian Philosophy*, page 3–41.
- J. Brough. 1953. [Some indian theories of meaning](#). *Transaction of the Philological Society*, 52(1):161–176.
- Marcus Coinnigh. 2015. [Structural Aspects of Proverbs](#). In *H. Hrisztova-Gotthardt, M. Aleksa Varga (Eds.), Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*. Berlin: de Gruyter., pages 112–132.
- Todd Ferretti, Christopher Schwint, and Albert Katz. 2007. [Electrophysiological and behavioral measures of the influence of literal and figurative contextual constraint on proverb comprehension](#). *Brain and language*, 101:38–49.
- C. D. Fellbaum D. Gross K. Miller G. A. Miller, R. Beckwith. 1990. Wordnet: An online lexical database. *Internal Journal of Lexicography*, 3(4):235–244.
- Chitra Garg and Lalit Goyal. 2014. Automatic extraction of idiom , proverb and its variations from text using statistical approach.
- R. W. Gibbs and D. Beitel. 1995. What proverb understanding reveals about how people think. *Psychological Bulletin*, 118(1):133–154.
- P. Grzybek. 2014. [4 Semiotic and Semantic Aspects of Proverb](#), in: *H. Hrisztova-Gotthardt, M. Aleksa Varga (eds)*.
- R.P. Honeck. 2016. *A Proverb in Mind: The Cognitive Science of Proverbial Wit and Wisdom*. Psychology Press.
- R.P. Honeck and J.G. Temple. 1994. [Proverbs: The extended conceptual base and great chain metaphor theories](#). *Metaphor and Symbolic Activity*, 9(2):85–112.
- S. Jagannath. 2009. Abhanakajagganatha. raghvendra matham mantralayam.
- M. Kulkarni, C. Dangarikar, I. Kulkarni, A. Nanda, and P. Bhattacharya. 2010. Introducing sanskrit wordnet. In *The 5th International Conference of the Global Word-Net Association (GWC-2010), 31st Jan-4th Feb.*
- M. Kuusi. 1972. *Towards an International Type-system of Proverbs*. Number nos. 211-218 in FF communications. Suomalainen tiedeakatemia.
- Outi Lauhakangas. 2001. The matti kuusi international type system of proverbs. *Folklore Fellows' Communications*, 275. Helsinki: Academia Scientiarum Fennica.
- Outi Lauhakangas. 2013. [The matti kuusi international database of proverbs](#). *Oral Tradition*, 28.
- Syahron Lubis. 2019. [The universality and uniqueness of proverb and its impact on translation](#). In *UNNES International Conference on English Language Teaching, Literature, and Translation (ELTLT 2018)*. Atlantis Press.
- Rev. Manwaring. 1899. Marathi proverbs translated and collected. *Clarendon Press, Oxford.*
- Wolfgang Meider. 2004. *Proverbs. A Handbook*. Westport, Connecticut: Greenwood Press.
- W. Mieder, W. Mieder, and A. Dundes. 1994. *The Wisdom of Many: Essays on the Proverb (Vol. 1994)*.
- Vishvanath Naravane. 1978. Bhārtiya vichārsādhanā prakāshan. *Bhārtiya Vichārsādhanā Prakāshan, Pune.*
- Neal R. Norrick. 2015. [1 subject area, terminology, proverb definitions, proverb features](#). *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies*, pages 7–27.
- Amanda P. Rassi, Jorge Baptista, and Oto Vale. 2014. [Automatic Detection of Proverbs and their Variants](#). In *3rd Symposium on Languages, Applications and*

Technologies, volume 38 of *OpenAccess Series in Informatics (OASISs)*, pages 235–249, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Brahmaleen K. Sidhu, Arjan Singh, and Vishal Goyal. 2010. Identification of proverbs in hindi text corpus and their translation into punjabi. *Journal of Computer Science and Engineering*, 2.

D.A. Swinney. 1979. [The access and processing of idiomatic expressions](#). *Journal of Verbal Learning and Verbal Behaviour*, page 523–534.

J.G. Temple. 1999. [Proverb comprehension: The primacy of literary meaning](#). *Journal of Psycholinguistic Research*, 28(1):41–70.

Yulia Tsvetkov and Shuly Wintner. 2014. [Identification of multiword expressions by combining multiple linguistic information sources](#). *Computational Linguistics*, 40(2):449–468.

Roger D. Woodard. 2008. *The Ancient Languages of Asia and the Americas.*, pages 1–2. Cambridge University Press.

Appendix A: Process of Cognition Example

Let us apply process of cognition to: Nirastapādape deshe eraṇdopi ḍṛmāyate |

1. Padjnyānm: Nirastapādape deshe eraṇdopi ḍṛmāyate

2. padārthjnyānm – After identifying all the words either by their sounds or by the sequence of alphabets, next step is to understand their meanings. The next step is understanding the vṛtti, i.e. the relation of word and word meaning. To understand the meanings of the word we have to rely on the śakti or the primary meaning or the denotative power. To understand the śakti let us consider the different ways of śaktigraha.

vyākaraṇa - Grammar. drumāyate - generally in Sanskrit verbs originate from roots like verb *gaccha* – root *gam*, verb *yaccha* – root *da*. However, there are certain verbs created from Nouns which are called nāmadhātū. They derive a root from a noun which they consider to be the most suitable, the most appropriate

to convey their thought. drumāyate: takes the place of a tree, pādapa: it is a compound, pādai: pibati | (upapada tatpuruṣa) one who drinks through their legs – a tree which drinks water through its roots.

koṣa: – Dictionary. *nirasta*– deprived or void of, *eraṇḍa*: castor bean plant. Thus the primary meaning of the proverb attained by the abhidhā śakti is: In a place where there is no vegetation, even a castor plant takes the place of a tree. This would have been called a mere description of a castor plant. But since we use the words *eraṇḍopi ḍṛmāyate*, even a castor plant takes the place of a tree indicates that this sentence wants to say something more than its literal meaning. Thus here abhidhā śakti does not suffice. We have to go to the next level, i.e. lakṣaṇā or figurative meaning. But before going to the secondary meaning let us examine whether the words in the given frozen expression satisfy the conditions of ākaṅkṣā, yogyatā, Sannidhi.

Since words are occurring in immediate sequence, condition of Sannidhi (phonetic contiguity) is fulfilled. In this sentence, we have a verb, related kartā to that verb and the other related words also tell us their connection with the verb. So the condition of ākaṅkṣā syntactic expectancy) is also fulfilled. In this expression eraṇḍ ḍṛmāyate these two words are incompatible. Because the plant of eraṇḍ grows up to two to four feet, it is a plant with a small number of leaves and not the tree. Still, it is called a tree. Thus these are incompatible with each other. Though there is no compatibility of words in this expression, this expression is not considered meaningless. It is generally observed that many of the proverbs hold incompatible words and still express deep meaning. The deliberate use of incompatible words in proverbs makes us more attentive towards its meaning. Now let us move to the secondary meaning.

3. The primary meaning attained by the denotative power does not seem meaningful, so we consider the next level, i.e. the secondary meaning or the lakṣaṇā. The secondary meaning of the proverb/ meaning attained by the lakṣaṇā is : When everyone around is of mediocre capacity, one with even a few accolades becomes the hero. The secondary mean-

Theme		Sub-theme		Proverb	Sentence type		Sentence function		Proverbial formulae	Marathi	Hindi ...
D	Concept of morality	d	Noble-wicked	शठे शाठ्यं समाचरेत्	I	Simple (+)	a	Advisory	-	घटासी असावे घट	
B	Cause - effect	c	(-) cause (-) effect	विनाशकाले विपरीतबुद्धिः	I	Simple (+)	e	Informative/ observation	-	बुडत्याचा पाय खोलात	
C	Relativity	a	better than	बधिरात् मन्दकर्णः श्रेयः	I	Simple (+)	c	Comparative	-	वासरात लंगडी गाय शहाणी	अंधों में काना राजा
B	Cause- effect	a	Cause- effect	यथा बीजं तथा अङ्कुरः	II	complex	e	observation	3	यथा- तथा	खाण तशी माती
B	Cause- effect	b	(+) cause (+) effect	सा विद्या या विमुक्तये	II	complex	d	complement	8	या – सा	
D	Morality	e	Bad deed	बुभुक्षितः किं न करोति पापम् ?	I	Simple (-)	b	Rhetoric	-		
G	Human Nature	a	Condemnation	गुणैः विहीनाः बहु जल्पयन्ति	I	Simple	f	criticism	-	उथळ पाण्याला खळखळाट फार	अधजल गगरी छलकत जाय
J	temporal	d	Modern, contemporary	यत्र नेटवर्क विद्यते रमन्ते तत्र युवकाः	II	Complex	e	informative	2.	यत्र – तत्र	
G	Human Nature	b	praise	अनिर्वेदः श्रियो मूलम्	I	Simple	e	Informative	-	उद्योगाचे घरी देवता लक्ष्मी वास करी	
J	temporal	b	Epic Ramayana								

Figure 4: Classification Example

ing tells us the implied meaning of the proverb. The same proverb can be uttered in multiple situations by different speakers. Lakṣaṇā does not cover subjectivism as lakṣaṇā limits itself to figurative meaning. Hence we have to go further and find out the suggestive meaning (vyanjanā) expressed through this proverb. There can be many suggestive meanings depending upon the situation, the intention of the speaker, the way the listener receives it.

4. Suggestive Meanings: It is quite natural to compare oneself to his surroundings and feel a sense of accomplishment. Calling oneself better than those around, aids only in boosting the ego. If the general average of the company is kept low, one may not even know there are bigger things out there in the world. People should raise the bar of self-assessment by keeping the company of noble people.

5. Cognition of the Proverb: Be an achiever in harsh surroundings than being the best amid trivial situations.

Appendix B: Classification of Proverbs

An example of classification is shown in figure 4. Following is the detailed explanation of

classification that we use in ProverbNet.

B.1 Structural Classification

1. Sentence Type: Proverbs appear in a variety of different sentence types; from a syntactic perspective, these sentences may be classified into simple, and non-simple (compound, complex) (Coinnigh, 2015). Simple sentences are typically simple, declarative, non-oppositional, and they do not contain any stylistic markers. E.g. ati sarvatra varjayet (affirmative) , amantram akṣram nāsti (negative). compound/complex sentence are sentences containing one or more dependent clauses in addition to the main clause. E.g. yādṛṣm vapate beejam tādṛṣm labhte falam, yah kriyāvān s pandith |

2. Sentence Functions: We classify sentences based on the function they perform in the communication. Proverbs may belong the following categories.

- (a) Advisory/potential – in Sanskrit the potential (vidhyartha) form is used to denote advisory function. E.g. ati sarvatra varjayet| kaṅtakenaiva kaṅtakam ud-dharet |

- (b) Rhetoric - A question asked to create a dramatic effect or to make a point rather than to get an answer. E.g. *kr̥she kasyasti sauhṛdam? andhasya deepen kim?*
- (c) Comparative – Denotes comparison between two objects. In Sanskrit the suffixes we use *tara*, *īyasa*, *varam* to denote comparison. E.g. *Janani janmabhumi-cha swargādpi gariyasi | Varam sarpo n durjanah |*
- (d) Complement (*x is y*) - It is a word, phrase, or clause that is necessary to complete the meaning of a given expression. In Complement sentence type of proverb, we use one object to denote another object. E.g. *yah kriyāvān sah pandith| guṇāh pujāsthanam|*
- (e) Informative sentence – a sentence which provides general information. E.g. *jalbindu nipaten kramaṣah puryate ghath| sāhase shree prativasti|*
- (f) criticism - proverbs that criticise things, people, actions, behaviour. E.g. *kr̥śe kasyāsti sauhṛdam|*
- (g) observation - proverbs that contain general observations. E.g. *Fatātopo Bhayankarh |*
- (h) comment - proverbs that provide general comments, these proverbs can sometimes also be classified as observations. E.g. *Maunm sarvārthsadhanm|*
- (i) suggestion - It is an advice, suggestion. E.g. *Mule Kuthār|*

B.2 Proverbial Formulae

If applicable, we classify proverbs according to the following formulae:

1. *Itah.. that – ito vyāghrh tato tati*
2. *yatra.. tatra – yatra yatra dhumo tatra tatra vahni*
3. *yathā.. tathā - yathā rājā tathā prajā*
4. *yādṛṣm.. tādṛṣm - yādṛṣm vapate beejam tādṛṣm labhte falam*

5. *x vā.. y vā - naro vā kunjaro vā*
6. *x vā y.. na vā - dātā bhavti vā na vā*
7. *na x.. na y - kāmāturaṇām na bhayam na lajjā*
8. *yah/yā.. sah/sā - yah kriyāvān sah pandith*

It is seen that structurally Marathi proverbs are comprised of two distinct parts. Very often a distinct rhyme scheme is seen in these two parts of a proverb. There is no specific theory that explains the reason for such a rhyme scheme based structure. However since the proverbs have their roots in colloquial (oral) use of language, one may conclude that owing to the oral nature of proverbs and its brief, concise and crisp structure rhyme scheme might have been incorporated into proverbs. We classify proverbs as rhyming or not rhyming, and following are some examples of rhyming proverbs.

- *ati udār to sadā nadār*
- *sāpalyamadhye wāgh sāpade, bāyakā mule mārīti khade*
- *dev tāri tyālā kon māri*
- *khāin tar tupāshi nāhitar upāshi*
- *āiji rāṇi bāyji rāṇi, toṇd dhuvāyalā kon deil pāṇi*

B.3 Thematic Classification

- (A) Natural Elements: human behaviour explained through natural elements
- (a) water, fire, earth, sea, soil, flora
 - (b) animals
 - (c) weather
- (B) Cause-Effect Relationship
- (a) cause-effect
 - (b) positive cause leading towards positive effect
 - (c) negative cause leading towards negative effect
 - (d) negative cause (contribution) leading towards a positive output
- (C) Relativity

Sanskrit	Marathi	Hindi	Gujrati	Bengali
न नश्यति तमो कृतया दीपवार्तया	बोलाची कडी बोलाचा भात	नापे सौ गज, फाड़े न एक गज	नहि लेवा, नहि देवा, वडोदरे विवाह	कथाय चिंड़े भेजे ना
अत्युन्नतिः पतनहेतुः	चढेल तो पडेल	चढेगा सो गिरेगा	भणे ते भूले ने चडे ते पडे	-
राजा हरति सर्वस्वं शरणं कं ब्रजेत्तदा ?	कुंपणाने शेत खाल्ले तर तक्रार कुणाकडे करायची ?	कुतिया चोरों से मिले तो पहारा कौन दे ?	राजा पोते लूटी ले तो प्रजा कोणने कहे ?	रक्षके भक्षण करे, के बाँचाते पारे तारे ?
पिण्डे पिण्डे मतिर्भिन्ना	व्यक्ती तितक्या प्रकृती	जितने मुँह उतनी बातें	माणस माणसमां आंतरो, कोई जवेर कोई कांकरो	नाना मुनीर नाना मत

Figure 5: Equivalent proverbs in different languages

- (a) one is better than other
(b) one is bigger than other
- (D) Concept of Morality
(a) good – bad
(b) pride – humility
(c) youth – old age
(d) noble – wicked
(e) sin – good deed
(f) virtue – vice
- (E) Group Behavioural Observations (subject to morality of the respective community)
(a) quality having a positive connotation
(b) quality having a negative connotation
- (F) Human Attitude
(a) optimistic
(b) pessimistic
- (G) Human Nature
(a) condemnation: laziness, greed, ego, selfishness, cowardice, treachery, misery, foolishness,
(b) praise: hardworking, bravery, patriotism, politeness, kindness, generosity, cleverness, truthfulness, dexterity, enterprise, intelligence,
- (H) Social Position: Based on,
(a) class: wealth, poverty, money
(b) caste: hierarchy
(c) gender: man-woman relationship
(d) power : strength (powerful/privileged) – weakness (powerless/unprivileged)
- (I) Universal Truth
- (J) Temporal - As most of the Sanskrit proverbs have their origin from various verses and various ancient scriptures, this aspect of classification is essential. We can have four subdivisions under this them. They are:
(a) from Vedic scriptures: Four Vedas, Upanishads, Brāhman, Āraṇyakas
(b) from epics: Rāmāyaṇa and Mahābhārata
(c) from classical literature: plays, poems of renowned poets, Panchatantra, vidurniti, Kautilya arthashastra, triṣatak by Bhṛṅhari, subhāshitaratna-bhāndāgāram and others
(d) modern/ contemporary proverbs: ābhānakajagannātha
- (K) Relationship :
(a) relatives
(b) friends
(c) social relationship
(d) relationship with animals, plants, nature