ISSN 2063-5346

VARIOUS SORTS OF AMBIGUITIES AND WORD SENSE DISAM-BIGUATION IN ENGLISH-SANSKRIT TRANSLATION: AN OVERVIEW AND ANALYSIS

***ARCHANA SACHINDEO MAURYA**
Research Scholar, Computer Science and Information Systems, Shri Ramswaroop Memorial University, India

BINEET KUMAR GUPTA
Department of Computer Science and Information Systems, Shri Ramswaroop Memorial University, India

***Corresponding Author:** Archana Sachindeo Maurya. Email: archanamaurya2308@gmail.com

Article History: Received: 01.02.2023 **Revised:** 07.03.2023 **Accepted:** 10.04.2023

Abstract

NLP is considered one of the most important areas of application in the artificial intelligence field. Natural languages are used to communicate with each other in the most common way. In a connected world, it is important to automate the translation from one language to another. Though, it is challenging to know all the languages. Therefore, it is necessary to translate one language into another through the translation process. When this translation process takes place through a machine, it is called "Machine Translation" (MT). MT is an advanced method of translation through a computer system. All MT processes face the challenge of various types of ambiguities. Ambiguity is an open challenge in the process of machine translation from the source language to target language. Here, we have addressed the most prominent NLP application, i.e. Machine Translation (MT). The most challenging issue in machine translation is the presence of various kinds of ambiguities, like lexical, syntactic, semantic, pragmatic and part-of-speech ambiguities. These ambiguities can be found at different levels of the translation process. In this paper, we have explained each type of ambiguity with examples with special focus on the English-to-Sanskrit translation. We have also discussed the significant approaches to solve the ambiguity challenges. Keywords: Ambiguity, Artificial Intelligence, Natural Language Processing, Part-of-Speech ambiguity, Word Sense Disambiguation

1. Introduction

Language is a phenomenon and a component that unites different communities as well as a means of transferring feelings and concepts that people try to pass on [**1**]. India is a multilingual country with only approximately three percent of the population understanding English [**2**]. Translation is one of the most fundamental, important, and sufficient ways in the advancement of civilization. Machine Translation (MT) plays an important role in NLP applications. MT is the process in which automatic translation takes place with the help of computers. MT is an important aspect of society. By using the MT process the social concepts can be transferred between two languages with the help of computers. Few barriers occur during this translation process. Ambiguity is one of them. MT is the automatic translation of the text from a source language (SL) to the text in the target language (TL) without any human intervention. The translation process is shown in **Fig. 1**.

Figure 1: Machine Translation Process

The primary goal of the MT process is to achieve an error-free translation, but it is more difficult to achieve a hundred-percent high-quality translation. The reason for incorrect translation is due to the presence of some stylistic and structural differences among languages. These stylistic and structural differences are of different types, like those due to word sense, word order, idioms, pronoun resolution, and ambiguity. The Structural and stylistic differences among various languages is shown in **Fig. 2**.

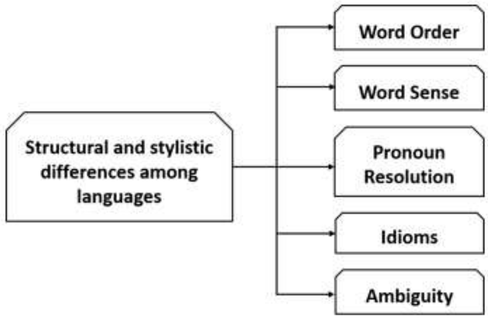

Figure 2: Structural and stylistic differences among languages

Each stylistic and structural difference exists among almost all the languages. Detailed explanation of each difference is shown in **Table 1** with respect to the English language.

Table 1: Detailed Explanation of Structural and Stylistic differences occurring during MT Processes

S. No.	Problem	Description	Example
1.	Word Order	The order of words varies from language to language, and this makes the translation more difficult. For example, we can see that, in the English language, the words are arranged in subject-verb-object order, while in Sanskrit, the sentences follow the subject-object-verb order.	English sentences follow S-V-O order while some target languages like Sanskrit language follow S-O-V order. **1.** Ram **(S)** plays **(V)** cricket **(O). (English sentence)** **2.** रामः **(S)** कन्दुकम्**(O)** क्रीडति **(V) (Sanskrit sentence)**
2.	Word Sense	The sense of a word in one language can be translated into a different sense in the target language. This issue can create a big problem for the selection of senses in the target language.	In the sentence "**This is a deep and dry well.**" When we translate the word "**well**" from English to Sanskrit, language, then well has three senses: 1. सम्यक्" (in a good or satisfactory way) 2. "अभुग्न" (free from disease) 3. "कूपः " (a deep hole full of water or oil) Therefore, when we translate from English-to-Sanskrit, then it is problematic to select the best sense from the available list of the senses.
3.	Pronoun Resolution	A correct resolution of pronominal reference is essential for accurate translation. Pronominal reference means an entity discussed previously	Consider the sentence "**Sita got good marks in examinations. She is very happy.**" In this sentence she is linked with

		will be linked to another textual part of the document.	Sita.
4.	Idioms	A sentence that contains the idioms does not have direct meaning. Idioms can be replaced by words in the target language.	Consider the idiom “**Cut to the chase.**” In the sentence is “**briefly explain the point, or you are wasting the time. So hurry up and tell me exactly what you are thinking**”. This phrase cannot be directly translated.
5.	Ambiguity	The words that have more than one meaning in the target language are called “multifunctional words or ambiguous words”, and the problem is called the “ambiguity problem.” During translation process it is necessary to eliminate this problem.	In the sentence “**The bank is next to my school.**” In this sentence, “**bank”** is an ambiguous word and when we translate it into the Sanskrit language, then it has necessary to solve this problem because it has two meanings exists in the Sanskrit language, i.e. “वित्तकोषः” and “नदीतीर: ”

There are lots of words present in almost all natural languages that have different meanings in various context. Words having multiple meanings are referred to as multifunctional words or ambiguous words and this phenomenon is known as ambiguity. Ambiguity is a multi-branched and complicated concept [**3**].

Organization of the Paper

This paper is organized as follows: In section 2, the review of literatures on various types of ambiguities and WSD are discussed. In section 3, we have explained and covered the various ambiguity challenges and it’s in English language. In next section, section 4 we have suggested discussed Word Sense Disambiguation and its various supervised, unsupervised and knowledge-based approaches. In Section 5, finally, we've covered the study's conclusion and its direction for future.

2. Review of Literatures

- WSD has been covered widely across the globe. Coverage includes various languages, including foreign and Indian languages. On the basis of the worked techniques, there are many sources of data sets proposed for the WSD disambiguation processes [**4**]. Analysis on the various literatures have been done related to the WSD process and methods. Many algorithms and approaches have been reported to be used. The findings are as below:
- In the Bayesian framework, a new algorithm is given and named as novel context clustering approach. The similarity between context pairings is the foundation of this algorithm. The developed model is trained by using the heterogeneous features that will reflect the probability distribution of context pair similarity [**5**].
- The Naïve Bayesian approach is a highly-featured approach in supervised learning. For selecting the ideal set of features, a selection technique named forward-sequential is used. This technique gives higher accuracy. Training data as well as testing data, might be tested to see the result and found that it is beneficial or not [**6**].
- Hybrid method is adopted for enhancing the performance and it is compared with the supervised and unsupervised techniques. They presented that the results of unsupervised, supervised and hybrid methods were sixty-three percent, seventy-six percent, and eighty percent, respectively. As a result, they conclude that a hybrid strategy improves accuracy. They said that the hybrid

method gives hundred percent accurate results if the ambiguous word is correctly disambiguated [**7**].

- In recent years, the most popular disambiguation technique is corpus-based technique. In this method, each word needs to be tagged, and these tagged words are used for the WSD. This approach is called as a supervised approach. This set of tagged texts is working as a training data set and helps to disambiguate the new test data. [**8**]. A variety of famous and well-known information-based techniques have been implemented to disambiguate word meanings. We've covered some commonly monitored WSD techniques here.
- A decision tree is used to assign a sense to an ambiguous word based on the bigram model. In this model the words can occur adjacent to each other [**9**].
- Various algorithms of supervised approaches were suggested in which class-based comparison has been used. Three different word related scores are: WordNet hypernym relations, classes of word similarity based on the clusters, and analysis of dictionary definitions. These scores are working for the collocation of the WorldNet [**10**].
- The IMS (It Makes Sense) uses the Support Vector Machines (SVM) classifiers. In this method the following characteristics are taken as features: Neighbor words, POS, and collocation of neighbor words [**11**].
- The decision tree procedure is the most widely used supervised WSD algorithm. This algorithm is used by many researchers. In this method, the data set is partitioned into the training data set and the testing data set. This partition takes place recursively. Classification rules are represented in the form of a tree called decision trees. In this decision tree the sense of any word is indicated by each leaf node [**12**].
- Word Embeddings are used by various researchers. A new semi-supervised method for WSD is proposed **[13, 14**]. Sense tagged data are automatically obtained from the English dictionary and the English dictionary contains one million sense tagged instances.
- A different supervised approach has been suggested. In this approach the Recurrent Neural Networks (RNN) have been used. This concept is designed with the help of the Long Short Term Memory (LSTM). The feature added by LSTM is the distributed word representations [**15**].
- The Topic Model is used in the disambiguation process. The most common used example of this model is LDA i.e. Latent Dirichlet Allocation. In this model, the total number of topics is used as the key features. Graphical model is also presented here. And this model is based on probability. This topic-based model is used for the supervised WSD method [**16**].
- Various WSD approaches, their advantages, and disadvantages are presented in the detailed review article. All these WSD approaches are successfully implemented in many Indian languages. A successful WSD algorithm was built when the same surrounding words have the same appearance. This is considered as the main factor of this approach [**17**].
- The knowledge-based system framework can be built up by using four components. These four components are: relation extraction, semantic space, semantic path exploration, and similarity calculation. Performance of this method is better for verbs as well as nouns in comparison to all other schemes and was tested on three different datasets [**18**].

- Complex problems can be solved by using the Dynamic Programming (DP). In this method the complex problem is broken down into smaller problems. DP is active to quickly identify the best subset, ad this subset contains the most definitely set for better accuracy than the previous one. Dynamic Programming Embedded Data (DPED) can be combined with boosting, and bagging algorithms for improving the efficiency of big-data and meticulous data, [**19**].
- A conceptual framework has been developed that raises situations of various types and their possible effects [**20**]. This approach is helpful for the decision-makers and stakeholders when they are facing various kinds of complex problems. The approach is a preliminary step for solving complex problems like the hydropower crisis.
- SensEmbert is a powerful technique for the multilingual WSD method. This works in both English and the other languages. Using sense embedding, the WSD produces high-quality silver data across many languages [**21**].
- Today, lots of information is available on the Web in the form of unstructured data. These lots of data are known as "big data." The review of literature addresses machine learning approaches for problems involving big data [**22**].

According to the discussed literatre review, there are numerous supervised, unsupervised and knowledge-based methodologies are available for the disambiguation of an ambiguous word. A word in any sentence has many senses in various contexts. Our aim is to disambiguate the meaning of that target word in that context.

3. Ambiguity: An Open Challenge During Machine Translation Process

Over the last decade, the ambiguity problem is an open challenge in almost all NLP applications. Ambiguity is still a great challenge for computational morphology and computer scientists [**23**]. Ambiguity has been a critical issue in the interaction between humans and computers. The words that have more than one meaning or way of interpretation present in the sentence are called "**multifunctional words**" or "**ambiguous word**", and this situation is known as "**ambiguity**". Table 1 shows some ambiguous words with their respective meanings.

3.1 Definition of Ambiguity

It may appear that there isn't much to say about ambiguity because "**it is a well-defined phenomenon in which words and sentences can have many meanings**" [**24**]. A sentence is said to be ambiguous whenever "**it can be associated with two or more different meanings"** [**25**]. Ambiguity can be defined as the **"quality of being able to be interpreted in more than one way"** [**26**]. All the authors agree that ambiguous words, phrases, or sentences can have several meanings. It takes a certain context to figure out the exact meaning or message given by ambiguous words or sentences. The first type of ambiguity happens when an expression has two or more meanings that are not related.

From the above definitions, we can define that "**ambiguity can exist when any word, phrase, or sentence has more than one possible meaning or way of understanding it.**"

Example: "This gift box is very **light.**"

In this example, the word "**light"** is an ambiguous word and it has **three** different meanings:

1. **Light**: light weight of something, not a heavy gift box
2. **Light**: a gift box that has an electric lamp
3. **Light**: a shiny box

When we translate the word "light" into any other target language, like Hindi, Sanskrit, or any other language, it is necessary to locate the actual meaning of an ambiguous word.

3.2 Ambiguity Classification

Ambiguities can be broadly classified into two categories: word-level ambiguity and sentence-level ambiguity. This categorization is shown in **Fig. 3**.

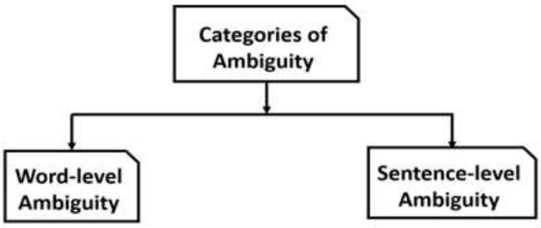

Figure 3: Two categories of Ambiguity

Word-level ambiguity: A word's level of ambiguity occurs in a sentence when a single word has more than one possible meaning. The word-level ambiguity is also called "lexical ambiguity", or "semantic ambiguity."

Sentence-level ambiguity: The sentence-level ambiguity occurs within a sentence due to the structure of the sentence, which leads to more than one possible meaning. This type of ambiguity is also known as "structural or grammatical ambiguity."

Ambiguity can be broadly classified as semantic, pragmatic, lexical, syntactic, and language errors. When a word has multiple meanings, it is said to have lexical ambiguity. Syntactic ambiguity also termed as structural ambiguity. Structural ambiguity in any phrase or sentence can be arise when the sentence can be described in multiple ways. This results in more than one well-formed structure and each structure also provides different meanings. When the predicate logic of any statement can be interpreted in many ways, then that phrase is termed as semantic ambiguity. Language ambiguity refers to grammatically incorrect constructions that are still understood, though in different ways, as a result of Berry's error [**27**]. Pragmatic ambiguity is frequently caused by human uncertainty about common sense knowledge and context knowledge. Some authors have further divided the ambiguities into "intentional and unintentional" categories [**28**], Intentional category of ambiguity exists when the writer has left an ambiguity by planning, while unintentional ambiguity occurs when ambiguity left by mistake. Further, ambiguity can be classified as "nocuous and innocuous" [**29**], where readers can interprets a requirement either in a single interpretation or in a multiple interpretation. Nocuous ambiguity can be divided into "acknowledged and unacknowledged" categories [**30**]. We can classify the ambiguity into the following five distinct classes, which are shown in **Fig. 4**.

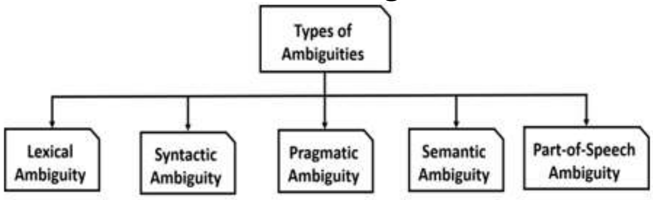

Figure 4: Different Categories of Ambiguities

Table 2, provides detailed information on the various kinds of ambiguities each with an example and an explanation.

Table 2.Kinds of ambiguities with detailed explanation

S. No.	Ambiguity Types	Definition	Example	Explanation
1	Lexical Ambiguity	It happens when just a single word has more than one importance, or when a word can be deciphered in multiple sense [**31- 34**].	“My salary is directly transferred to my bank account.”	In the example word bank, there is an ambiguous word that has two meanings: (i) "An organization or building where individuals or corporations can spend, borrow, convert money into another currency, etc., or a building where these services are provided." (ii) "Sloping raised land, often near riverbanks."
2	Syntactic ambiguity	Syntactic ambiguity occurs in a particular sentence when the sentence has two or more possible meanings due to the different sentence structures [**35, 36**]. The other names for this type of ambiguity are structural ambiguity or grammatical ambiguity.	I saw a black horse with spectacles.	: In this sentence, saw is an ambiguous word, and this sentence has two meanings due to the sequence of words. This sentence means either the black horse was wearing spectacles or I was wearing spectacles to see the horse.
3	Pragmatic ambiguity	Pragmatic ambiguity implies that there are multiple interpretations in a sentence [**37, 38**]. It means there can be more than one way of understanding a sentence.	“The chicken is ready to eat.”	: This sentence has a multiple interpretation, like: a) The chicken is ready to eat its breakfast. b) The cooked chicken is ready to be served.
4	Semantic ambiguity	The semantic ambiguity occurs when the denotation of the words can be taken in different ways [**39, 40**]. In other words, we can say that semantic ambiguity exists when a sentence contains an unclear word or phrase.	“The car hit the scooter while moving.”	This sentence has two ways of understanding it:’ a) The car, while moving, hit the scooter. b) The car hit the scooter while the scooter was moving.
5	Parts-of-Speech (POS) ambiguity	The part of speech ambiguity occurs when a single word may be a noun or a verb, an adjective or an adverb, singular or plural, etc. [**41**].	“Comb your hair.”	a) Comb is a noun whose meaning is “a strip of plastic, wood, or metal having narrow teeth, used for arranging the hair”. a) Comb can also be a verb whose meaning is “comb through the hair to arrange it.”

3.3 Ambiguity: A Common Problem in English Language

All Indian languages bear various kinds of ambiguities. English is expected to be the leading primary language in the world due to its foundation as the maternal language in many countries [**42**]. English is a language that has plenty of various kinds of ambiguities. There are unlimited words within the English language that have different meanings in different backgrounds. These words are termed as "ambiguous words." These ambiguous words need to be disambiguated properly for the correct translation in the objective language translation process from English to any other India language. Some ambiguous words in the English language are represented in **Table 3**.

Table 3.Some ambiguous words and their meanings in the English language with respect to their part of speech

S. No.	Ambiguous Word	Meaning1 (Noun)	Meaning2 (Verb)	Meaning3 (Adjective)	Meaning4 (Adverb)
1.	Fly	A small insect with two wings; the opening with a pair of pants that covers the zipper or buttons	To move or travel through the air; control something in the air	-	-
2.	Form	The visible shape or configuration of something.	Join or combine parts together to create something. Parts or combine to create (something)	-	-
3.	Double	A thing that is twice as large as usual or is made up of two standard units or things.	Double in size or increase in number.	Consisting of two equal, identical, or similar parts or things.	At or to twice the amount or extent.

Table 3, shows that many terms in the English language are ambiguous, having two or more meanings` that are recognized by the dictionary. These words must be translated correctly for their intended meaning in the objective language when we translate them from English language to Sanskrit language.

4. Word Sense Disambiguation (Wsd)

Word sense disambiguation (WSD) is the procedure of selecting the most appropriate denotation of an ambiguous word. The complete process was completed by considering its special situations and best emotions within that word's context. WSD can be defined as "**a method for picking a word's precise meaning out of a list of previously defined meanings** [43, 44]."

4.1 Disambiguation Techniques

There are several disambiguation techniques that can be used to reduce ambiguity during MT. The different types of WSD techniques and their further classification are shown in **Fig. 5.**

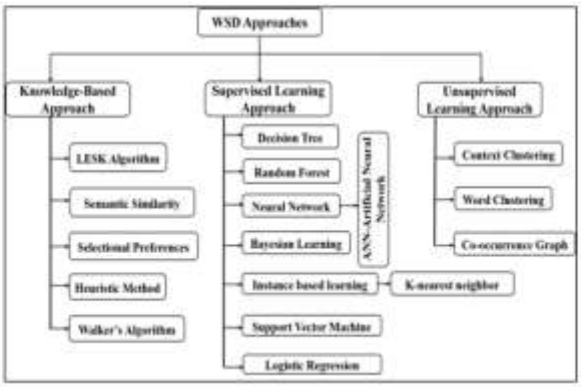

Figure 5: Various WSD approaches.

Word sense disambiguation (WSD) plays an important role in detecting the exact meaning of an ambiguous word. Hard work is still continuing to solve the problem using machine translation (MT). MT is one of the most important applications through which we can convert a source language into a target language with the help of machines [**45**]. Different approaches used in the WSD processes are explained in detail below.

4.1.1 Knowledge-Based Approach

This approach is totally depends on knowledge for example machine readable dictionaries, thesaurus, vocabularies, catalogues, etc. In this field, the most commonly used dictionaries are WordNet [**46**]. Four algorithms based on the knowledge-based approach are: The LESK algorithm, selectional preferences, semantic similarity, and Heuristic Method. These methods are explained in detail:

- **The LESK algorithm:** This algorithm was first introduced by Michel Lesk in 1986 [**47, 48**]. This algorithm is built around vocabulary. This algorithm identifies the correct meaning of an ambiguous word. This algorithm is faster and reduces the complexity in terms of computation time.
- **Semantic Similarity:** This algorithm is used to find the smallest distance between two semantically related words. This approach is knowledge-based and is used for solving WSD problems. This algorithm determines the relationship between two words [**49**]. Semantic similarity can be used to check the patterns and quantify the doubtfulness [**50**].
- **Selectional preferences:** For collecting information related to the potential relationship between various categories of words, the selection preferences method is used. This method is based on the senses of available information.
- **Heuristic Approach:** For determining the correct senses of an ambiguous word, the heuristic method is used. There are three popular heuristic methods:
- Most frequently used senses search for all possible meanings of an ambiguous expression**.** This approach is used to uncover information about possible connections between various word categories and also refers to the familiar feeling based on the information's source. These partialities are described using semantic classes rather than a single
- According to the principle of one sense per dialogue, meaning of any word will be maintained throughout all instances in a particular text. It affects categorization likelihood, but if there is strong local evidence, it can also be ignored.
- One sense per association estimation is comparable to one sense per discourse estimation. The exception to the hypothesis is that nearby words provide stronger and more unique messages for the same word.
- **Walker's algorithm**: The thesaurus is the foundation of this algorithm. This algorithm works by calculating the results for each sense while looking up synonyms of an ambiguous word. If the meaning of the synonym is similar to the words, then it will add one. The usage of synonyms in this technique produces the best results.

Benefits: The main benefits of various algorithms of knowledge-based approach are that the improved LESK algorithm is much faster than the actual LESK algorithm and also has a low level of computational complexity. The distance between two words which is smallest is semantically related to each other. Preferences for selection reduce the timing related to the manual tagging required by humans. The heuristic method helps in investigating potential problems after the testing process.

Drawbacks: The main drawback of the LESK algorithm is that it requires lots of knowledge sources. The uniform distance problem occurs in the semantic similarity method when two concepts along the same pathway will have the same semantic similarity. In the preferred selection method, it may be challenging to determine the grammatical link between particular words or phrases. Both knowledge and experience are required in the heuristic method. Additionally, it costs more for designers.

4.1.2 Supervised Learning Approach

This learning approach is a probability-based approach. In this approach, the data set is divided into the training data set and the test data set. To train the system, supervised techniques can be applied to a well-labeled dataset. After the training process, the newest test dataset is applied to the trained system to produce the results. The supervised learning approach uses the two methods i.e. Classification and regression.

Classification: Using supervised learning classification techniques, we can divide the text into two or more classes. These techniques may be neural networks, support vector machines (SVM), Naïve Bayes, logistic regression, decision tree, random forest, K-nearest neighbor etc.

Regression: This method shows the relationship between the dependent and the independent variables. This method is also used to forecast data. There are two methods, i.e. Logistic and linear regression. Some supervised algorithms are explained here in detail:

- **Decision Tree:** A decision tree is one of the most significant and popular supervised learning approaches. Another name for decision tree learning is the tree-based learning method. This algorithm helps in taking quick decisions. This algorithm offers excellent reliability and predictability for the test data. These trees can be utilized to make decisions and put those decisions into action by implementing systems [**51**].

Example of a decision Tree: Figure 6 shows the decision tree of a situation of rain condition for playing indoor or outdoor game.

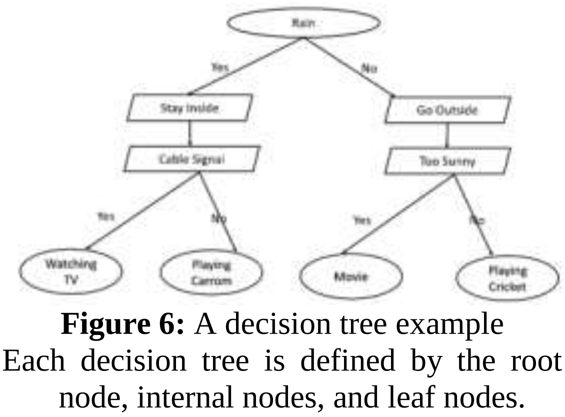

Figure 6: A decision tree example

Each decision tree is defined by the root node, internal nodes, and leaf nodes.

- **Random Forest:** One of the most popular and effective techniques for information analysis, is the modeling of data and the prediction of new data. The classifier uses an ensemble approach in which more than one decision tree is grouped together to make the final decisions. This method outperforms the approach and gives higher accuracy in comparison to the single decision tree. We can create a random forest using decision trees from the same data set, but we cannot relate the trees. The result of this algorithm is a tree, built from the results of individual decision trees [**52**]. This algorithm reduces the chance of overflow, and the accuracy is much higher than a single decision tree. In this algorithm, the decision trees work in parallel, and no bottleneck problem occurs. The visual representation of the Random Forest algorithm is shown in **Fig. 7**.

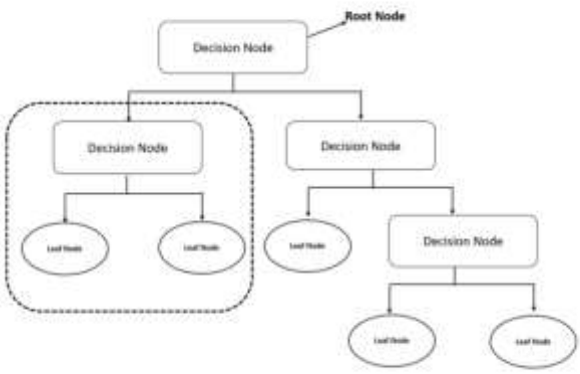

Figure 7: Visual representation of Random Forest Approach

- **Naïve Bayes':** This is the popularly used supervised WSD algorithm and this algorithm is based on the probability value. The algorithm used the Bayes theorem [**53**]. This approach is very popular for segmentation problems. The training dataset contains a large number of variables, and all these variables are not related to each other. These unrelated variables are called independent variables and are referred to as "features." With the help of Bayes' theorem, we can calculate the likelihood of specific features in a given class [**54**]. The following formula is used to calculate the probability value:

Bayes' theorem states that if there are two events occurring, then the probability can be calculated with the help of the following formula:

$$P(\frac{a}{b}) = \frac{P(b/a)*P(a)}{P(b)} \quad (1)$$

Here,
a and b are two events.
P(a | b) : probability of event A occurring after the event B has occurred.
P(a) : prior probability that the event will occur.
P(b | a) : likelihood probability of event B following the occurrence of evidence of event X.
P(b) : the marginal probability
Bayes' rule is used to determine the conditional probability characteristics of a given class [**55-57**]. The conditional probabilities of each term's value and characteristic in a given sentence are calculated with the aid of this algorithm. The highest value will produce the best result.

- **Neural Network:** In neural networks, data are processed by artificial neurons [**58**]. Artificial neural networks (ANN) comprise one of the most crucial and significant characteristics of neural networks. These neurons have the ability to learn like a human brain [**59**]. Neurons are the processing units of the neural networks. A neuron has dendrites as its input units and synapses or axons as its output units. Three layers, the input layer, the hidden layer, and the output layer, make up the fundamental structure of an ANN. The middle hidden layer receives information from the first layer, i.e. Input layer which contains neurons. After doing some calculations on the data, the hidden layer sends the results to the output layer. The following formula is given for calculating the input for the general model of ANN [**60**].

$$\mathrm{Yin} = \mathrm{x1.w1} + \mathrm{x2.w} \ldots\ldots\ldots \mathrm{xm.wm} \quad (2)$$

i.e. total input $\mathrm{yin} = \sum_{i}^{m} \mathrm{xm.wm}$ (3)

The result can be determined using the stimulation function on the total input value.

$$\mathrm{y} = \mathrm{F(yim)} \quad (4)$$

- **K-nearest neighbor:** This is a straightforward, however effective, approach that uses a non-parametric grouping method [**61**]. This is a very popular technique for classifying data, and it is frequently used with unlabeled data. The workings of this algorithm are based on the votes of its k-neighbors. Where k will be any value like 1, 2, 3 …etc. The two key concepts used by this method are:
  (i) The first idea is founded on calculating the distances between the training data and testing datasets for two similar features. In this idea, the calculation for the value of k is completed first, after which the test results are used to determine which category the neighbors belong to [**62**].
  (ii) Choose the value of K in the second notion first. The number of neighbors utilized in the calculation depends on the value of k [**63**].

It is recommended that the value of k in this algorithm is between 0 and 1. Under-fitting and overfitting are two terms used to describe this issue.
Following is a formula for calculating the Euclidean distance between two points:

$$= \sqrt{((P_2 - P_1)^2 + (Q_2 - Q_1)^2)} \quad (5)$$

The test word's coordinates in this instance are P1 and Q1.
The coordinates of the matching feature are P2 and Q2.

- **Support Vector Machine:** SVM is the popular supervised machine learning model that is working on the classification problem. The primary objective of support vector machines is to find the best highest margin distinguishing the hyperplane for the two given classes in the training data set. In this algorithm, if the data points are closer in the other class, then it is not acceptable for the hyperplane. The hyperplane closest to the data points belong to the other class for better generalization [**64**]. Hyperplane that exist away from the data points for each category are not selected. Support vectors are those points that lie nearest to the margin of the classifier [**65**].

- **Logistic Regression:** This is a probability-based text classification method. This is a predictive algorithm that is applied to the prediction of test data. The approach is used for the unconditional datasets, and the output will be in binary form. The classification problem that is based on the binary output is called a "binary classification problem" [**66**].

4.1.3 Unsupervised WSD Approach

This is a probability-based text classification method. This is a predictive algorithm that is applied for the prediction of test data. The approach is used for the unconditional datasets, and the output will be in binary form. The classification problem that is based on the binary output is called a "binary classification problem" [**67**].
Unsupervised Learning Approach: An unsupervised learning strategy doesn't always call for a training corpus and a lot of computation effort

- **Clustering:** Data points can be split into two or more groups via clustering, depending on the characteristics of the groups. Clustering is of two types:
- **Context Clustering:** In this method, groups are represented using context vectors or similarity matrices, depending on the grouping strategies used. These are grouped together to form clusters, which are then utilized to determine a word's meaning. When there is no class to forecast, but the inputs may be categorized into natural groupings, this strategy can be used.
- **Word Clustering:** Using this technique, words are grouped together based on the semantic similarity of their attributes. These words' shared characteristics can be used to determine how similar they are. Words in the same category that are strikingly similar share the same characteristics. Then, using the clustering process, it is possible to distinguish between the senses.
- **Association:** Numerous variables of dataset are related to each other are related to one another using the association rule learning method.

5 Conclusion

This paper we have outlined the challenging issues in a machine translation process termed "ambiguity". It is the presence of various kinds of ambiguous words found at the different levels of the translation process. Further, we have done a detailed literature review of the work on the ambiguity and WSD approach. We have also presented the classification of ambiguities i.e. Word level and sentence level ambiguity and their types, like lexical, syntactic, semantic, pragmatic, and part of speech ambiguity. These ambiguities are addressed with reference from English to Sanskrit translation. Ambiguity in English language with focus on part of speech ambiguity is discussed.

Word Sense Disambiguation is presented as the solution for the ambiguity challenge. There are various machine learning approaches like knowledge-based approach, supervised and unsupervised approaches are explained in detail in reference to resolve ambiguity.

Funding Statement: The authors did not receive any special funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Author Contributions: All authors have substantially contributed to the manuscript and have approved the final submitted version.

References

1. Genç, B., & Bada, E. (2010). English as a world language in academic writing.\
2. Sinha, R. M. K., & Jain, A. (2003). AnglaHindi: an English to Hindi machine-aided translation system. In Proceedings of Machine Translation Summit IX: System Presentations.
3. Song, L. (2010). The role of context in discourse analysis. Journal of Language Teaching and Research, 1(6), 876.
4. Haroon, R.P., 2011. Word sense disambiguation-A survey. Proceedings of the International Colloquiums on Computer Electronics Electrical Mechanical and Civil, (EMC' 11), ACEEE, pp: 58-60. DOI: 02.CEMC.2011.01.582

5. Niu, C., Li, W., Srihari, R. K., Li, H., & Crist, L. (2004, July). Context clustering for word sense disambiguation based on modeling pairwise context similarities. In Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (pp. 187-190).
6. Le, C. A., & Shimazu, A. (2004, December). High WSD accuracy, using Naive Bayesian classifier with rich features. In Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation (pp. 105-114).
7. Saktel, P., & Shrawankar, U. (2013). An Improved Approach for Word Ambiguity Removal. ArXiv preprint arXiv:1304.7282.
8. Banerjee, S., & Pedersen, T. (2002, February). An adapted Lesk algorithm for word sense disambiguation using WordNet. In International conference on intelligent text processing and computational linguistics (pp. 136-145). Springer, Berlin, Heidelberg.
9. Pedersen, T. (2001). A decision tree of bigrams is an accurate predictor of word sense. ArXiv preprint cs/0103026.
10. O’Hara, T., Bruce, R., Donner, J., & Wiebe, J. (2004, July). Class-based collocations for word sense disambiguation. In Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (pp. 199-202
11. Zhong, Z., & Ng, H. T. (2010, July). It makes sense: A wide-coverage word sense disambiguation system for free text. In Proceedings of the ACL 2010 system demonstrations (pp. 78-83).
12. Singh, R. L., Ghosh, K., Nongmeikapam, K., & Bandyopadhyay, S. (2014). A decision tree based word sense disambiguation system in Manipuri language. Advanced Computing, 5(4), 17.
13. Taghipour, K., & Ng, H. T. (2015, July). One million sense-tagged instances for word sense disambiguation and induction. In Proceedings of the nineteenth conference on computational natural language learning (pp. 338-344).
14. Yuan, D., Richardson, J., Doherty, R., Evans, C., & Altendorf, E. (2016). Semi-supervised word sense disambiguation with neural models. arXiv preprint arXiv:1603.07012.
15. Popov, A. (2017, September). Word sense disambiguation with recurrent neural networks. In Proceedings of the Student Research Workshop associated with RANLP 2017 (pp. 25-34).
16. Chaplot, D. S., & Salakhutdinov, R. (2018, April). Knowledge-based word sense disambiguation using topic models. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).
17. Ahmed H. Aliwy and Hawraa A. Taher, 2019. Word Sense Disambiguation: Survey Study. Journal of Computer Science, Accepted July 2019, Iraq.
18. Wang, Y., Wang, M., & Fujita, H. (2020). Word sense disambiguation: A comprehensive knowledge exploitation framework. Knowledge-Based Systems, 190, 105030.
19. Alzubi, O. A., Alzubi, J. A., Alweshah, M., Qiqieh, I., Al-Shami, S., & Ramachandran, M. (2020). An optimal pruning algorithm of classifier ensembles: dynamic programming approach. Neural Computing and Applications, 32(20), 16091-16107.
20. Gaudard, L., & Romerio, F. (2020). A Conceptual Framework to Classify and Manage Risk, Uncertainty and Ambiguity: An Application to Energy Policy. Energies, 13(6), 1422.
21. Scarlini, B., Pasini, T., & Navigli, R. (2020, April). SensEmBERT: Context-enhanced sense embeddings for multilingual word sense disambiguation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 05, pp. 8758-8765).
22. Bineet Kumar Gupta, Rajeev Kumar et.al, “Analyzing the Big Data Security

through a Unified Decision-Making Approach”, in Intelligent Automation & Soft Computing (IASC), ISSN 10798587, 2326005X. pp- 1071-1088, Volume 32(2), September 2021, DOI:10.32604/iasc.2022.022569,

23. Jusoh, S. (2018). A Study on NLP Applications and Ambiguity Problems. Journal of Theoretical & Applied Information Technology, 96(6).
24. Navigli, R. (2009). Word sense disambiguation: A survey. ACM computing surveys (CSUR), 41(2), 1-69.
25. Kempson, R. M. (1977). Semantic theory. Cambridge University Press.
26. Tuggy, D. (1993). Ambiguity, polysemy, and vagueness.
27. V. Gervasi and D. Zowghi, "On the role of ambiguity in RE," in Requirements Engineering: Foundation for Software Quality, ed: Springer, 2010, pp. 248-254.
28. V. Gervasi and D. Zowghi, "Reasoning about inconsistencies in natural language requirements," ACM Transactions on Software Engineering and Methodology, vol. 14, pp. 277-330, 2005
29. D. M. Berry, "Ambiguity in natural language requirements documents," in Innovations for Requirement Analysis. From Stakeholders' Needs to Formal Designs, Springer, 2008, pp. 1-7.
30. D. M. Berry and E. Kamsties, "Ambiguity in requirements specification," in Perspectives on software requirements, ed: Springer, 2004, pp. 7-44.
31. F. Chantree, B. Nuseibeh, A. De Roeck, and A. Willis, "Identifying nocuous ambiguities in natural language requirements," in 14th International Conference Requirements Engineering, 2006, pp. 59-68.
32. Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. Cognition, 122(3), 280-291.
33. Fera, A. (2019). Lexical Ambiguity Compared to Syntactic/or Structural Ambiguity. Knowledge-International Journal, 30(5), 1113-1116.
34. https://www.researchgate.net/publication/324440596_A_study_on_nlp_applications_and_ambiguity_problems
35. Dayal, V. (2004). The universal force of free choice any linguistic variation yearbook 4: 15-40. Retrieved from http//www.ingentaconnect.com
36. Cruse, D. (1986). Lexical Semantics. Introducing lexical relations. Cambridge: Cambridge University Press.
37. Yadav, A., Patel, A., & Shah, M. (2021). A comprehensive review on resolving ambiguities in natural language processing. AI Open, 2, 85-92.
38. Elena, B. (2019). Ambiguity matters in linguistics and translation. Слово. ру: балтийский акцент, 10(3), 81-93.
39. Kadłub, M. (2017). Sources of ambiguity in language.
40. Zelta, E.N. (2014). Ambiguity. In Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu.R. Navigli
41. https://stanford.library.sydney.edu.au/entries/ambiguity/
42. Cruse, D. (1986). Lexical Semantics. Introducing lexical relations. Cambridge: Cambridge University Press.
43. V. Gervasi and D. Zowghi, "On the role of ambiguity in RE," in Requirements Engineering: Foundation for Software Quality, ed: Springer, 2010, pp. 248-254.
44. Mincă, A., & Diaconescu, Ş. (2013, November). An Approach to Reduce Part of Speech Ambiguity Using Semantically Annotated Lexicon Definitions. In 2nd International Conference
45. Navigli, R. (2009). Word sense disambiguation: A survey. ACM computing surveys (CSUR), 41(2), 1-69.
46. D. L. a. P. Pantel, “Discovering Word Senses From Text,” in ACM, 2002
47. Bahadur, P., & Chauhan, D. S. (2014, August). Machine translation—A

journey. In 2014 Science and Information Conference (pp. 187-195). IEEE.

48. Banerjee, S., Pedersen, T.,(2002) "An adapted Lesk algorithm for word sense disambiguation using WordNet", In Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City, February.
49. Lesk, M.,(1986) "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone", Proceedings of SIGDOC.
50. Jiang, J.J. and D.W. Conrath, 1997. Semantic similarity based on corpus statistics and lexical taxonomy. Proceedings of the 10th Research on Computational Linguistics International Conference, Aug. 5-7, Taipei, Taiwan, 19-33.
51. Ahmed H. Aliwy and Hawraa A. Taher, 2019. Word Sense Disambiguation: Survey Study. Journal of Computer Science, Accepted July 2019, Iraq.
52. http://link.springer.com/article/10.1023/A%3A1002674829964#page-1
53. Cuong Anh Le and Akira Shimazu, “High WSD accuracy using Naive Bayesian classifier with rich features”, PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo, pp. 105-114Venkatesan, N., & Priya, G. (2015). A study of random forest algorithm with implementation using weka. International journal of innovative research in computer science and engineering, 1(6), 156-162.
54. Szepesvári, C. (2010). Algorithms for reinforcement learning. Synthesis lectures on artificial intelligence and machine learning, 4(1), 1-103.
55. Nyein Thwet Thwet Aung, Khin Mar Soe, Ni Lar Thein, “A Word Sense Disambiguation System Using Naïve Bayesian Algorithm for Myanmar Language”, International Journal of Scientific & Engineering Research Volume 2, Issue 9, September-2011, pp. 1-7
56. Cuong Anh Le and Akira Shimazu, “High WSD accuracy using Naive Bayesian classifier with rich features”, PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo, pp. 105-114.
57. https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/
58. Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science, 3-21.
59. Nyein Thwet Thwet Aung, Khin Mar Soe, Ni Lar Thein, “A Word Sense Disambiguation System Using Naïve Bayesian Algorithm for Myanmar Language”, International Journal of Scientific & Engineering Research Volume 2, Issue 9, September-2011, pp. 1-7.
60. G.E.a. D. Radev. “Lexrank: Graph Based Lexical,” Artificial Intelligence Research, vol 22 pp.457-479, 2004.
61. Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5), 717-727.
62. Haroon, R. P. (2010, December). Malayalam word sense disambiguation. In 2010 IEEE International Conference on Computational Intelligence and Computing Research (pp. 1-4). IEEE.
63. Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003, November). KNN model-based approach in classification. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 986-996). Springer, Berlin, Heidelberg.
64. Cuong Anh Le and Akira Shimazu, “High WSD accuracy using Naive Bayesian classifier with rich features”, PACLIC 18, December 8th-10th, 2004, Waseda University, Tokyo, pp. 105-114.

65. S.N. Jeyanthi, "Efficient Classification Algorithms using SVMs for Large Datasets," A Project Report Submitted in partial fulfillment of the requirements for the Degree of Master of Technology in Computational Science, Supercomputer Education and Research Center, IISC, BANGALORE, INDIA, Jun. 2007