

# lakṣyārtha (Indicated Meaning) of Śabdavyāpāra (Function of a Word) framework from kāvyāśāstra (The Science of Literary Studies) in Samskr̥tam :Its application to Literary Machine Translation and other NLP tasks

Sripathi Sripada Anupama Ryali Raghuram Sheshadri  
School of Vedic Sciences, MIT-ADT University, India  
{sripathi.sripada, anupamaskt, raghuys}@gmail.com

## Abstract

A key challenge in Literary Machine Translation is that the meaning of a sentence can be different from the sum of meanings of all the words it possesses. This poses the problem of requiring large amounts of consistently labelled training data across a variety of unsages and languages. In this paper, we propose that we can economically train machine translation models to identify and paraphrase such sentences by leveraging the language independent framework of Śabdavyāpāra (Function of a Word), from Literary Sciences in Samskr̥tam, and its definition of lakṣyārtha ('Indicated' meaning). An Indicated meaning exists where there is incompatibility among the literal meanings of the words in a sentence (irrespective of language). The framework defines seven categories of Indicated meaning and their characteristics. As a pilot, we identified 300 such sentences from literary and regular usage, labelled them and trained a 2d Convolutional Neural Network to categorise a sentence based on the category of Indicated meaning and finetuned a T5 to paraphrase them. We then used these paraphrased sentences as input into Google Translate and compared this with Google Translate's translation before paraphrasing using BLEU scores against an expected reference translation. The BLEU scores improved significantly with the paraphrasing by the T5 trained on Indicated meaning sentences.

Keywords: Indicated meaning, Literary Machine Translation, language independent, T5<sup>1</sup>, Convolutional Neural Network, paraphrase.

<sup>1</sup> T5 is Google's state of the art text to text NLP model. T5 stands for Text-To-Text Transfer Transformer

## 1 Introduction

Consider a sentence from Rabindranath Tagore's *Gitanjali* "Drunk by the joy of singing, I forget myself", and its translation to various languages using Google Translate<sup>2</sup>. Refer to Table 1 for sample translations

Language	Google Translate's Translation
Hindi	गाने के आनंद के नशे में धुत मैं खुद को भूल जाता हूँ gaane ke aanand ke nashe mein dhut main khud ko bhool jaata hoon
Bengali	গান গাওয়ার আনন্দে মাতাল হয়ে আমি নিজেকে ভুলে যাই Gāna gā'ōyāra ānandē mātāla hayē āmi nijekē bhulē yā'i
Kannada	ಹಾಡುವ ಸಂತೋಷದಿಂದ ಕುಡಿದು, ನಾನು ನನ್ನನ್ನು ಮರೆತಿದ್ದೇನೆ Hāḍuva santōṣadinda kuḍidu, nānu nannannu maretiddēne
Telugu	పాడిన ఆనందంతో త్రాగి, నన్ను నేను మర్చిపోతున్నాను Pāḍina ānandantō trāgi, nannu nēnu marcipōtunnānu
Italian	Ubriaco dalla gioia di cantare, dimentico me stesso
German	Betrunken von der Freude am Singen vergesse ich mich selbst

Table 1: Sample translations by Google Translate of the example sentence "Drunk by the joy of singing I forget myself"

For the example above, while in some languages the usage could be appropriate, the translation is not very clear in quite a few languages. On the other hand, if we paraphrase the original sentence to "Overjoyed by singing, I forget myself", then

<sup>2</sup> Google's publicly available translation engine at <https://translate.google.co.in>

Google Translate’s translation is more consistent across multiple languages. Refer to Table 2 for sample translations of the paraphrased sentence

Language	Google Translate’s Translation
Hindi	गाते-गाते खुशी से झूम उठे मैं खुद को भूल गया gaate-gaate khushee se jhoom uthe main khud ko bhool gaya
Bengali	গান গেয়ে আনন্দিত, আমি নিজেকে ভুলে যাই Gāna gēyē ānandita, āmi nijekē bhulē yā'i
Kannada	ಹಾಡುವ ಮೂಲಕ ಹರ್ಷಗೊಂಡ ನಾನು ನನ್ನನ್ನೇ ಮರೆತಿದ್ದೇನೆ Hāḍuva mūlaka harṣagoṇḍa nānu nannannē maretiddēne
Telugu	పాడటం ద్వారా చాలా సంతోషించాను, నేను నన్ను మరచిపోయాను Pāḍaṭam dvārā cālā santōṣiṅcānu, nēnu nannu maracipōyānu
Italian	Felicissimo di cantare, mi dimentico di me stesso
German	Überglücklich vom Singen vergesse ich mich selbst

There are many such sentences across literary works where the sum of meanings of all the words in a given sentence, does not necessarily provide the meaning of the sentence. In all such cases, an appropriate paraphrasing should make machine

Table 2 : Sample translations by Google Translate of the paraphrased sentence "Overjoyed by singing, I forget myself"

translation more accurate. To train machine learning models for paraphrasing of such sentences before translation, we are faced with the challenge of creating large datasets for training across different types of usages, figures or speech etc., and across multiple languages.

**Recent related works:** Recent research in the applying machine translation models to literary works is broadly focused on:

<sup>3</sup> वक्तृ-बोद्धव्य-काकूनां सम्बन्धः। vaktr-boddhavya-kākūnām sambandhaḥ. The relationship between speaker, listener, tone.

- training the models to identify and paraphrase metaphors to their literal meanings (Jerry Lui, 2020) (Rui Mao, 2018) leveraging word embeddings
- modifications to existing machine translation models for classification of consistency, pronoun resolution, and tone/register error types to consider context of previous sentences or even the whole story, to improve quality of literary machine translation (Matusov, 2019)
- the role of referential cohesion to improve Literary Machine Translation (Rob Voigt, 2012)

However, to our knowledge, there is lack of a holistic approach that encompasses a variety of the challenges presented in Literary Machine Translation in a manner consistent across languages.

**Our approach:** To overcome this challenge, we seek inspiration from kāvyasāstra, the Science of Literary Works / Poetics, in Saṃskṛtam. Various texts in Saṃskṛtam in this domain, provide comprehensive and lucid frameworks to understand literary works. A variety of concepts discussed in these texts are language independent as well. Of many such concepts, kāvyasāstra lays much importance to a word and its meaning. It emphasises that a word and its meaning depend on the speaker, the listener, and the tone<sup>3</sup>. At times, it is understood with the context too. This framework of understanding the meaning is called Śabdavyāpāra (as explained in kāvyaprakāśa) and it categorises the word and its meaning broadly into three types<sup>4</sup>, namely,

- vācakaḥ (वाचकः - Expressive) word with vācyārthaḥ (वाच्यार्थः Expressed) meaning or literal / direct / primary meaning
- lākṣaṇikaḥ (लाक्षणिकः - Indicative) word with lakṣyārthaḥ (लक्ष्यार्थः - Indicated meaning)
- vyañjakaḥ (व्यञ्जकः - Suggestive) word with vyañgyārthaḥ (व्यङ्ग्यार्थः - Suggested) meaning.

The Indicated meaning from the above framework provides a very fundamental categorisation of words which covers a variety of

<sup>4</sup> स्याद्वाचको लाक्षणिकः शब्दोत्र व्यञ्जकस्तिथा । syādvācako lākṣaṇikaḥ śabdotra vyañjakastridhā . The words are of 3 types – Expressive, Indicative and Suggestive

figurative usages, metaphors, referential cohesion and other characteristics. Hence, it provides a more holistic approach, in comparison to techniques focusing on metaphors and figures of speech, to solving some of the key problems in Literary Machine Translation. We explain the Indicated meaning, its various types and their characteristics along with examples in section 2.

**The Hypothesis:** Our hypothesis is that if we train a state of the art NLP model to paraphrase based on the Śabdavyāpāra framework and then use a state of the art Machine Translation model to translate, the translation of Literary works across languages will be much more meaningful, and consistent. Moreover, since the framework is language independent and has a very structured definition of the various types of Indicated meaning and their characteristics, we should be able to achieve a very efficient training with smaller datasets and consistently across languages.

We adopted a novel approach, based on Śabdavyāpāra framework's definition and characteristics of the Indicated meaning, to

- a) Train a 2d Convolutional Neural Network (CNN2d) to Identify the existence of an Indicated meaning, in a given sentence
- b) If Indicated meaning is present, then train a CNN2d to categorise the sentence based on the type of Indicated meaning, as per the framework
- c) Leverage the characteristics of the various types of Indicated meaning defined in the framework to finetune a Google T5 (Transformer NLP model) to paraphrase the sentence by elaborating the Indicated meaning such that the paraphrased sentence can be translated consistently by a model like Google Translate.

To do an initial validation of our hypothesis we created a dataset of 300 sentences from literary works, Śāstra works and common usage<sup>5</sup>. In Section 3, we describe the solution we adopted including the models we trained along with the

<sup>5</sup> We picked 300 sentences from a combination of Rabindranath Tagore's Gitanjali, kālidāsa's kumārasambhavam and śāstra texts of dhvanyāloka, kāvyaprakāśa.

<sup>6</sup> आकाङ्क्षा-योग्यता-सन्निधिवशाद् वक्ष्यमाणस्वरूपानां पदार्थानां समन्वये तात्पर्यार्थो विशेषवपुः अपथार्थोपि वाक्यार्थः | – ākāṅkṣā-

results we achieved in our pilot. In section 4, we conclude and highlight the other use cases of NLP where the identification and paraphrasing of Indicated meaning can be applicable. In Appendices we provide some examples of the seven categories of Indicated meaning.

## 2 Śabdavyāpāra (Function of a Word) and lakṣyārtha (Indicated meaning)

As stated above, according to Śabdavyāpāra framework meanings words convey are categorised as vācyārtha (Expressed meaning), lakṣyārtha (Indicated meaning) and vyaṅgyārtha (Suggested meaning). While 'Expressed meaning' is the straightforward sum of meanings of all the words in the sentence, in 'Indicated meaning' or 'Suggested meaning' the meaning of the sentence is not the sum total of the meanings of all the words and differ based on the various nuances of language, local culture etc.,

**Expressive:** That which denotes the direct conventional (or dictionary) meaning is the Expressive word. In ordinary parlance, a word denotes something by convention of the given language. Where the conventional denotation is not known, there is no comprehension of the meaning. Thus, when the conventional denotation is apprehended directly, without the intervention of any other agency, the word is said to be 'Expressive' of the denotation or meaning. In a sentence the words also need to satisfy three<sup>6</sup> conditions to be able to express the meaningful sentence. They need to have 'mutual requirement' as in all of them are needed, they need to be 'compatible' with each other and there needs to be 'proximity' meaning certain words need to be next to each other. Consider the sentence "*The student is studying mathematics*". It is very clearly understood what each word is denoting, hence each word is expressive. Moreover, they satisfy the three conditions of 'mutual requirement', 'compatibility' and 'proximity'; therefore, the sentence is a meaningful sentence. The meaning of such a sentence obtained by the meanings of the 'Expressive' words is called the 'Expressed'

yogyatā-sannidhivaśād vakṣyamāṇasvarūpānām padārthānām samanvaye tātparyārtho viśeṣavapuḥ apathārthopi vākyārthah When the denotations of different words become related together though 'mutual requirement', compatibility' and 'proximity' there appears in the shape of the 'meaning of the sentence' which is not expressed by any single word constituting the sentence.

meaning. It can also be referred to as the ‘Primary’ meaning of the sentence.

**Indicative**<sup>7</sup> : When the ‘Primary’ (or ‘Expressed’) meaning does not make sense (because of incompatibility), another meaning, which is in close affinity to what the word is denoting, is implied by the word. Such a meaning is called the ‘Indicated’ meaning that such a word is the ‘Indicative’ word in the given sentence. Consider the sentence “*Drunk by the joy of singing, I forget myself*”. Here when we put together the ‘Primary’ meanings of all the words we see there is incompatibility as joy is not a physical drink that one can get drunk on. The word drunk here implies the meaning overtaken or completely filled with. Using this implied meaning of the word drunk, we arrive at the meaning “Overjoyed by singing, I forget myself”. This is called the ‘Indicated’ meaning and the word drunk is the ‘Indicative’ word in this sentence. This process of implying the ‘Indicated’ meaning is called ‘Indication’. The ‘Indicated’ meaning of such a sentence makes the import of the sentence much clearer and is also very easily translatable by a machine learning model to any other language

**Suggestive**: Where the ‘Primary’ meaning is clear, there can also exist a ‘Suggested’ meaning. Such a word is called the ‘Suggestive’ word. The ‘Suggested’ meaning can also exist along with the ‘Indicated’ meaning. Since the focus of this paper is on the ‘Indicated’ meaning and its application, we do not go into the details of this category.

## 2.1 Various types of ‘Indication’

**‘Usage’ and ‘Special Purpose’ Indication**<sup>8</sup> :The process of imposing the ‘Indicated’ meaning is done either based on ‘Usage’ or for a ‘Special Purpose’. and as such these are the 2 categories of ‘Indication’.

Example of ‘Indication’ on the basis of ‘Usage’: Consider the sentence “*Do not beat around the bush when expressing your viewpoint*”. Here the primary meaning of words ‘do not beat around the bush’ are incompatible with the words ‘expressing your viewpoint’. However, it is common usage that means ‘do not waste time by giving lengthy and

cyclical explanations. This ‘Indicated’ meaning conveys the meaning of the sentence appropriately. A lot of idioms in English language, for example, fall into this category of ‘Indication based on Usage’

Example of ‘Indication’ on the basis of ‘Special Purpose’: Consider the sentence “*Her face had blooming smiles at the thought of meeting her lover*”. Here the ‘Primary’ meaning of the word ‘blooming’ is to be flowering and this is incompatible with the sense of the sentence which is describing the expression of a person’s face. The word ‘blooming’ is implying the ‘Indicated’ meaning in excess / lot of / big, which is in affinity with its ‘Primary’ meaning. Read with the ‘Indicated’ meaning, the sentence means that “Big smiles appeared on her face at the thought of meeting her lover”. An appropriate paraphrased sentence could be ‘She had big smiles on her face at the thought of meeting her lover’. Moreover, the implication of the ‘Indicated’ meaning also has a ‘Special Purpose’ of referring to the beauty, radiance etc. in an excessive way that appeared on her face at the thought of meeting her lover.

While ‘Indication’ on the basis of ‘Usage’ has no further sub-categories, ‘Indication’ on the basis of ‘Special Purpose’ has 6 sub-categories.

**Six sub-categories of ‘Indication’ on the basis of ‘Special Purpose’**: ‘Indication’ on the basis of ‘Special Purpose’ is further categorised into two<sup>9</sup>, namely, ‘Pure’ and ‘Qualitative’ Indications. When the ‘Indication’ relies upon similarity / similitude it is called ‘Qualitative’ Indication and when it is based upon other kinds of relationships (like cause-effect and not on similarity / similitude) it is called ‘Pure’ Indication.

‘Qualitative’ Indication: Consider the sentence “*Her eyes are lotus petals*”. In this sentence, the qualities of lotus petals are being imposed upon the eyes of the person and this is to show the similarities in their qualities, for example this lady has big eyes and in the shape of lotus petals. Here the imposed meaning is the quality of the lotus petal that is being imposed upon the eyes of the lady. This is an example of ‘Qualitative’

<sup>7</sup> मुख्यार्थबाधे तद्योगे – mukhyārthabādhe tadyoge - When there is incompatibility in the Primary meaning and the other meaning has affinity with the Primary meaning  
<sup>8</sup> रूढितोऽथ प्रयोजनात् – rūdhito’tha prayojanāt - The Indication is of 2 types, based on Usage and based on Special Purpose

<sup>9</sup> भेदाविमौ च सादृश्यात्सम्बन्धान्तरस्तथा bhedāvimau ca sādṛśyātsambandhāntarastathā. These 2 are different. One is by similarity and other by other relationships

Indication. An appropriate paraphrased sentence could be 'Her eyes are big and beautiful like lotus petals'.

Qualitative Indication can be of two types, namely, 'Super-imponent Qualitative' Indication and 'Intro-susceptive Qualitative' Indication, based on how the imposed qualities and that which they are being imposed upon are expressed in the sentence.

1. 'Super-imponent'<sup>10</sup> Qualitative' Indication: When what is being imposed and that which it is being imposed upon are mentioned separately in the sentence it is called a 'Super-imponent'. The sentence "*Her eyes are lotus petals*" is an example of 'Super-imponent Qualitative' Indication as what is being imposed (lotus petals) and that which it is being imposed upon (eyes) are mentioned separately. (akin to a simile)

2. 'Intro-susceptive'<sup>11</sup> Qualitative' Indication: When what is being imposed consumes (takes within itself) that which it is being imposed upon it is called 'Intro-susceptive'. Both are not mentioned separately in the sentence and only what is being imposed is mentioned. Considering the same example of the lady with big and beautiful eyes, if someone were to look at the lady's beautiful eyes and say "*They are lotus petals*", then this becomes an example of 'Intro-Susceptive Qualitative' Indication (akin to a metaphor). Here that which is being imposed (lotus petals) has consumed that which it is being imposed upon (eyes). A paraphrased sentence will be "Her eyes, which are big and beautiful, appear to be lotus petals themselves."

#### 'Pure'<sup>12</sup> Indication

'Pure' Indication is of four types, namely, Inclusive Indication, Indicative Indication, Super-imponent Pure Indication and Intro-susceptive Pure Indication.

3. 'Inclusive Pure' Indication: When the implication of the 'Secondary' meaning is for the sake of completing the 'Primary' meaning itself, it is called Inclusive Indication. Consider the sentence "*Your pizza is on its way*". Here, the pizza that has been ordered cannot be travelling on its

own, there is an unwritten actor present in the sentence, the pizza delivery person. The word pizza without losing its 'Primary' meaning is implying an actor to complete the 'Primary' meaning itself. This is 'Inclusive' Indication. A paraphrased sentence elaborating the Indicated meaning could be "The pizza delivery boy, along with your pizza, is on his way"

4. 'Indicative Pure' Indication: When the 'Primary' meaning is replaced by the 'Secondary' meaning, it is called Indicative Indication. Consider the sentence "*She jumps to conclusions*". Here, the Primary meaning of the word jumps is replaced by as Secondary meaning 'to form quickly'. Hence this is an Indicative Indication. A paraphrased sentence elaborating the Indicated meaning will be "She forms conclusions very quickly"

5. Super-imponent Pure Indication: When the Indication is based upon a relationship like cause-effect (and not similarity / similitude) between the imposed and what it is being imposed, and both are stated separately in the sentence it is 'Super-imponent Pure' indication. Consider the sentence "*Knowledge is power*". Here there is a cause-effect relationship between Knowledge (that which it is being imposed upon) and Power (imposed). Moreover, both are being stated clearly in the sentence. Hence it is a 'Super-imponent Pure' Indication. An appropriate paraphrased sentence will be "Knowledge gives power"

6. 'Intro-susceptive Pure' Indication: This is like the 'Super-imponent Pure' Indication but the imposed and that which it is being imposed upon are not stated separately in the sentence. When someone described a knowledgeable person and says "*He has the power*", it is an example of Intro-susceptive Pure Indication as Power (imposed) consumes the word Knowledge (that which it is being imposed upon) and both the words are stated not stated separately in the sentence. An appropriate paraphrased sentence will be "He has the power of knowledge"

<sup>10</sup> सारोपान्या तु यत्रोक्तौ विषयी विषयस्तथा । sārōpānyā tu yatroktau viṣayī viṣayastathā . Super-impotent is one where imposed and that which it is being imposed are stated / said separately.

<sup>11</sup> विषयन्तःकृतेन्यस्मिन् सा स्यात्साध्यवसानिका । viṣayantahkrtenyasmin sā syātsādhyaśānikā . That which the imposed consumes that which it is being imposed upon is Intro-susceptive

<sup>12</sup> स्वसिद्धये पराक्षेपः परार्थ स्वसमर्पणम् । उपादानं लक्षणं चेत्युक्ता शुद्धीव सा द्विधा । svasiddhaye parākṣepaḥ parārthaḥ svasamarpanam . upādānaḥ lakṣaṇaḥ cetyuktā śuddhauva sā dvidhā . Pure is of 2 types Inclusive and Indicative. Inclusive implies another actor to achieve its Primary meaning. Indicative gives up its Primary meaning to take on the Secondary meaning.

**Seven categories of Indication:** The two types of ‘Qualitative’ Indication and four<sup>13</sup> types of ‘Pure’ Indication make up the six categories of Indication based on Special Purpose. Along with the ‘Usage based’ Indication there are in total seven categories of Indication. The categorisation helps understanding the sentences and also provides a distinctive way for paraphrasing the sentence for each of the categories. The characteristics of the seven categories discussed above have been summarised into a flow chart presented in Appendix I. A few more examples of the seven categories of Indication are provided in Appendix II.

### 3 Application of the lakṣyārtha concept to Machine Learning models

If we notice the paraphrasing of the sentences with Indicated meaning, there are patterns that are correlated to the category of the Indicated meaning, in most cases except in the ‘Usage based’ Indication. At a high level the patterns of paraphrasing are summarised in Table 3 below.

Category of Indication	Typical Pattern of Paraphrase
Super-imponent Qualitative	Typically ‘like’ or an equivalent word is added in the sentence as the Indication is based on comparison
Intro-susceptive Qualitative	Similar to above along with the addition of the words that are left out in the sentence. This needs context in which the sentence as the speaker would leave out some of the words
Inclusive Pure	A related word(s) are added to explicitly mention the unspoken actor
Indicative Pure	A secondary meaning of the word replaced the word in the sentence where this indication exists. This secondary meaning is typically very closely related to the primary meaning of the word

<sup>13</sup> Commentators of kāvyaprakāśa also explain that Inclusive and Indicative Indications are further divided into Super-imponent and Intro-susceptive each, giving rise to the 4 Pure Indications. Either as per this categorisation or as per

Super-imponent Pure	Words are added to show the relationship between the imposed and this which it is being imposed. Typically this relationship between the words is quite commonly used
Super-imponent Pure	Same as above along with the addition of the words that are left out in the sentence. This needs context in which the sentence as the speaker would leave out some of the words
Usage based	Does not have any pattern as it is based on widely accepted usage in the given language.

Inspired by the correlation between the paraphrasing and the category of the Indication, we embarked on the pilot of training the Machine Learning models to do this paraphrasing before translation by Google Translate. Given the comprehensiveness and the fundamental nature of the categorisation, we believe that the training can be achieved with relatively small datasets. Hence, we attempted the pilot with a very small dataset. We broke the pilot down into three steps

Table 3: Typical patterns that can be observed in Paraphrasing sentences, based on the category of Indication

**Step 1: Identify the existence of an Indicated meaning in the sentence.** This means that the model needs to identify the incompatibility between the words in a sentence. To achieve this, we trained a multi-layer perceptron of 3 layers and a 2d Convolutional Neural Network (CNN2d) with filter sizes of 3, 4, 5, 6, 7 and 8 for binary classification on a dataset of 400 example sentences (100 without Indicated meaning and 300 with Indicated meaning). We used 320 of these sentences for training and 80 sentences for testing. While the multi-layer perceptron trained to 70% test accuracy, CNN2d achieved 78% test accuracy. This was on expected lines as the existence of Indicated meaning is identified based on incompatibility between words (refer footnote 4). The CNN2d is comparing groups of adjacent words of length 3, 4, 5, 6, 7 and 8 in its filters to

what has been explained in Section 2, there are 4 types of Pure Indication. We took the choice of the categorisation that we think is most appropriate for the Machine Translation.

map the incompatibility. Sample classification results by the CNN2d trained on our dataset of 400 sentences are presented below in Table 4.

Sl No	Sentence	Classification – Indication Exists (Yes/No)	
		Trained CNN2d	Actual
1	My heart spreads its wings	Yes	Yes
2	Master is knowledge	Yes	Yes
3	He is speaking the truth	Yes	No
4	Truth is bitter to swallow	No	Yes
5	I can run fast	No	No
6	On this stormy night the sky groans	Yes	Yes
7	It is raining heavily today	No	No
8	He is a walking encyclopaedia	Yes	Yes
9	He stole her heart	Yes	Yes
10	The water is blue in colour	No	No

**Step 2: Identify the category of Indicated meaning in a sentence.** We labelled the 400 sentences with eight labels (one label for Expressive and one each for category of Indication per the framework explained in Section 2) and trained the CNN2d for multi-classification. 320 sentences were used for training and 80 were used for testing. The model achieved 72% test accuracy. Sample classification results by the CNN2d trained

Table 4: CNN2d classification of whether an Indicated meaning exists in the given sentence

on our dataset are presented below in Table 5.

Sl No	Sentence	Category of Indication	
		Trained CNN2d	Actual
1	He is stretching the truth	Indicative Pure	Indicative Pure

2	He is a walking encyclopaedia	Usage Based	Super-imponent Qualitative
3	On this stormy night the sky groans	Inclusive Pure	Inclusive Pure
4	This is a magnificent new shirt	Expressive	Expressive
5	Health is wealth	Super-imponent Qualitative	Super-imponent Qualitative
6	His radiance was visible from far	Indicative Pure	Indicative Pure
7	The bus is arriving late	Inclusive Pure	Inclusive Pure
8	Master is knowledge	Super-imponent Qualitative	Super-imponent Pure
9	My heart spreads its wings	Super-imponent Qualitative	Intro-susceptive Qualitative
10	Truth is bitter to swallow	Expressive	Indicative Pure
11	Your pizza is on its way	Expressive	Inclusive Pure
12	Time heals everyone	Intro-susceptive Qualitative	Intro-susceptive Qualitative

**Step 3: Paraphrase the sentence with elaborating the Indicated meaning based on the category of Indicated meaning.** We finetuned a pre-trained Google’s T5 model to paraphrase sentences with our custom dataset of sentences with Indicated meaning. We used the patterns described in Table 3 to create our custom dataset. Our dataset contained 250 training sentences and 50 testing sentences. We refer to this finetuned T5

Table 5: CNN2d classification of a given sentence based on the type of Indicated meaning it contains

model as the T5-I.

We then used the T5-I paraphrased sentences as input to Google Translate for translation to various



languages. We then used BLEU<sup>14</sup> (Bilingual Evaluation Understudy) score to compare the translation with and without T5-I paraphrasing. For the purpose of this pilot we used translation to three Indian languages – Telugu, Hindi, Kannada. We used a typical human translation of the sentences in these 3 languages as reference for calculation the BLEU scores. Here, we present the comparison of translation with and without the paraphrasing by T5-I for a few validation sentences along with the respective BLEU scores.

<b>Original Sentence:</b> She was showered with blessings			
<b>Expected Translation:</b> ఆమె చాలా ఆశీస్సులు పొందింది. / उसको बहुत सारे आशीर्वाद मिले/ ಅವಳಿಗೆ ಬಹಳಷ್ಟು ಆಶೀರ್వాదಗಳು ದೊರಕಿದವು			
<b>Paraphrased Sentence (by T5-I) :</b> She received lots of blessings			
<b>Google Translate's Translation Of</b>			
Original Sentence	BLEU Score	Paraphrased Sentence	BLEU Score
ఆమె ఆశీస్సుల తో ముంచెత్తింది	0.54	ఆమె చాలా దీవెనలు పొందింది	0.70
वह आशीर्वाद के साथ नहाया गया था	0	उन्हें बहुत आशीर्वाद मिला	0.55
అవళ్ళు ఆశీర్వాద దదింద సురిసల్ప ట్టళ్ళు	0	అవళ్ళు బహళష్టు ఆశీర్వాద దగళ్ళన్న పడెదళ్ళు	0.70

<b>Original Sentence:</b> He stretched the truth			
<b>Expected Translation:</b> అతడు అబద్ధమ ఆడాడు / उसने झूठ बोला / ಅವನು ಸುಳ್ಳನ್ನು ಹೇಳಿದನು			
<b>Paraphrased Sentence (by T5-I) :</b> He used falsehood			

<b>Google Translate's Translation Of</b>			
Original Sentence	BLEU Score	Paraphrased Sentence	BLEU Score
సత్యాన్ని సాగదీశాడు	0	అతను అసత్యాన్ని ఉపయోగించాడు	0.76
उसने सच फैलाया	0.76	उन्होंने झूठ का इस्तेमाल किया	0.66
ಅವರು ಸತ್ಯವನ್ನು ವಿಸ್ತರಿಸಿದರು	0	ಅವರು ಸುಳ್ಳನ್ನು ಬಳಸಿದರು	0.76

<b>Original Sentence:</b> I buy peace of mind by being silent			
<b>Expected Translation:</b> నేను మౌనంగా ఉండి మనశ్శాంతిని పొందుతాను / मुझे चुप रहकर सुकून मिलता है / నాను మౌనదింద నేమ్మదియన్ను గಳిసిదే.			
<b>Paraphrased Sentence (by T5-I) :</b> My peace of mind comes by being silent			
<b>Google Translate's Translation Of</b>			
Original Sentence	BLEU Score	Paraphrased Sentence	BLEU Score
నేను మౌనంగా ఉండడం ద్వారా మనశ్శాంతిని కొనుక్కుంటాను	0.79	మౌనంగా ఉండడం వల్ల నా మనశ్శాంతి కలుగుతుంది	0.0
मैं चुप रहकर मन की शांति खरीदता हूँ	0.43	मेरे चुप रहने से आती है मन की शांति	0.68
నాను మౌనవా గిరువుదర మొలక	0.64	మౌనదింద నన్న మనస్సిగే	0.69

<sup>14</sup> BLEU (BiLingual Evaluation Understudy) is a metric for automatically evaluating machine-translated text. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a

set of high quality reference translations. A value of 0 means that the machine-translated output has no overlap with the reference translation (low quality) while a value of 1 means there is perfect overlap with the reference translations (high quality).



ಮನಸ್ಸಿನ ಶಾಂತಿಯ ನ್ನು ಖರೀದಿಸು ತ್ತೇನೆ		ಶಾಂತಿ ಸಿಗುತ್ತದೆ	
--	--	--------------------	--

<b>Original Sentence:</b> An idea sprouted in his mind			
<b>Expected Translation:</b> ಅತನಿಗೆ ಒಕ ಆಲೊಂನ ವಚ್ಚಿಂಧಿ / उसको मन में एक विचार आया / ಅವನಲ್ಲಿ ಒಂದು ಉಪಾಯವು ಚಿಗುರೊಡೆಯಿತು.			
<b>Paraphrased Sentence (by T5-I) :</b> An idea came to his mind			
<b>Google Translate's Translation Of</b>			
Original Sentence	BLEU Score	Paraphrased Sentence	BLEU Score
ಅತನಿ ಮದಿಲೊ ಒಕ ಆಲೊಂನ ಮೊಲಕೆ ತ್ತೊಂಧಿ	0.56	ಅತನಿ ಮದಿಲೊ ಒಕ ಆಲೊಂನ ವಚ್ಚಿಂಧಿ	0.56
उसके मन में एक विचार कौथा	0.50	उसके दिमाग में एक विचार आया	0.48
ಅವನ ಮನಸ್ಸಿನ ಲ್ಲಿ ಒಂದು ಕಲ್ಪನೆ ಚಿಗುರೊಡೆ ಯಿತು	0.79	ಅವನ ಮನಸ್ಸಿಗೆ ಒಂದು ಉಪಾಯ ಹೊಳೆಯಿ ತು	0.67

It can be noticed that even where the individual BLEU score did not show improvement with the paraphrasing by T5-I, the translation of the paraphrased sentence is much more meaningful than the one without. Taking a corpus score on the 12 Google Translate translations above the BLEU score for translations improved from 0.39 to 0.6 with paraphrasing by T5-I.

<sup>15</sup> We leveraged existing code of CNN2d and T5 models from Google Colab's (Colaboratory is cloud based ML resource) pytorch libraries.

## 4 Conclusion

Where the CNN2d correctly identified the existence of Indication, it performed very well in identifying the sub-categories of Indication except in the case of the two Super-imponent Indications. We believe this because of the lack of equal number of examples across categories in our training dataset. The improvement in BLEU score achieved for translations with paraphrasing by T5-I is significant and encouraging. The difference in the translation with and without T5-I paraphrasing was very evident in more complicated literary usages of sentences (and not just metaphors).

For the purpose of the pilot we trained black-box<sup>15</sup> implementations of CNN2d and T5-base in with a small dataset. We believe fine-tuning of the model architecture and a limited increase in the dataset can improve accuracy of the models for paraphrasing and translating sentences in Literary works with Indicated meaning to a higher level of accuracy. Where word embeddings are available the trained models should also work across languages..

There are other use cases as well, where understanding the real intent of a sentence depends on understanding of the Indicated meaning, including dialogue systems, sentiment analysis and emotion analysis.

## 5 Acknowledgements

The authors would like to acknowledge the efforts of Dr Sai Susarla in reviewing the methodology and providing critical inputs, Sriram Subramanian in helping with creation of dataset and providing review comments for the paper, and Manjula Pilaka for providing guidance in the programming that was needed to train and test the machine learning models.

## 6 References

Almahasees, Zakaryia Mustafa, Machine Translation Quality of Khalil Gibran's the Prophet (November 9, 2017). AWEJ for translation & Literary Studies Volume, 1 Number 4, October 2017

Anandavardhanacharya, Dhvanyāloka, Editor, Acharya Jagannath Pathak, Chowkambha Vidyabhavan, Varanasi, 2009

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In Proceedings of the International Conference on Learning Representations (ICLR), May.

Besacier, Laurent and Lane Schwartz. 2015. Automated translation of a literary work: a pilot study. In Proceedings of the Fourth Workshop on Computational Linguistics for Literature, pages 114–122.

Acharya Mammata, Kāvyaṣākāśa, , Editor Acharya Satyanarayanasastri Khanduri, Chawkambha Krishnadas Academy, Varanasi, 2016

James Hadley, Maja Popović, Haithem Afli, Andy Way (Editors), Proceedings of Qualities of Literary Machine Translation, Machine Translation Summit XVII, August 2019, Dublin, Ireland

Jerry Lui, N. O. (2020). Metaphor Detection Using Contextual Word Embeddings From Transformers. Proceedings of the Second Workshop of Figurative Language Processing, Association of Computational Linguistics

M. R. Kale, kumārasambhavam of kālidāsa – English Translation, Second Edition, The Standard Publishing Co. Bombay, 1917

Mahamahopadhyaya Dr. Sir Ganganatha Jha kāvyapākāśa, of Mammata with English Translation, Chapters I to X, Bharatiya Vidya Prakasan, 1966

Matusov, E. (2019). The Challenges Of Using Neural Machine Translation for Literature. The Qualities of Literary Machine Translation. Dublin.

P. V. Kane, History of Sanskrit Poetics, Motilal Banarasiidass Publishers Pvt. Ltd. Delhi, 2005

Rabinovich, Ella, Shachar Mirkin, Raj Nath Patel, Lucia Specia, and Shuly Wintner. 2016. Personalized machine translation: Preserving

original author traits. arXiv preprint arXiv:1610.05461.

Sennrich, Rico, Barry Haddow, and Alexandra Birch. 2016a. Improving neural machine translation models with monolingual data. pages 86–96, August

Rob Voigt, D. J. (2012). Towards a Literary Machine Translation: The Role of Referential Cohesion. Workshop on Computational Linguistics for Literature (pp. 18-25). Montreal, Canada: Association for Computational Linguistics.

Rui Mao, C. L. (2018). Word Embedding and WordNet Based Metaphor Identification and Interpretation. Proceedings of 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: Association for Computational Linguistics.

Sir Rabindranath Tagore, Gitanjali, English Translation, 1912

Taja Kuzman, Špela Vintar, Mihael Arčan, Machine, Neural Machine Translation of Literary Texts from English to Slovene-Translation Summit XVII, August 2019, Dublin, Ireland

Vamanacarya Ramabhata Jhalakikara , bālabodhini Samskrit commentary on kāvyapākāśa, Second Edition, Government Central Book Depot 1901