Evaluation of Direct Machine Translation System For Punjabi To Hindi

Gurpreet Singh Josan[1], Gurpreet Singh Lehal[2]

[1] Yadawindra College of Engg., Punjabi University Guru Kashi Campus, Talwandi Sabo.
josangurpreet@rediffmail.com
[2] Deptartment of Computer Science, Punjabi University Patiala. gslehal@yahoo.com

Abstract

The Direct MT system is based upon exploitation of syntactic similarities between more or less related natural languages. Both Punjabi and Hindi languages have originated from Sanskrit which is one of the oldest language. In terms of speakers, Hindi is third most widely spoken language and Punjabi is twelfth most widely spoken language. Punjabi language is mostly used in the Northern India and in some areas of Pakistan as well as in UK, Canada and USA. Hindi is the national language of India and is spoken and used by the people all over the country. Hindi and Punjabi are closely related languages with lots of similarities in syntax and vocabulary. In the present study, a Punjabi to Hindi machine translation system using Direct MT approach has been developed and its output is evaluated by already prescribed methods in order to get the suitability of the system. It was observed that a fairly high accuracy Punjabi to Hindi Machine Translation System has been developed by direct word-for-word translation. The major inaccuracies in the direct translation are due to poor word choice for ambiguous words and some corrections regarding post positions in Hindi.

1. Introduction

Evaluation is without doubt a major aspect of language engineering, including Machine Translation (MT). It plays an important role for system developers (to tell if their system is improving), for system integrators (to determine the appropriate approach) and for consumers (to identify which system will best meet a specific set of needs). Beyond this, evaluation plays a critical role in guiding and focusing research [Bharati et al. 2004; Hajic et al. 2000; Marrafa et al. 2001]. Despite the fact that history of Machine translation is quite old, the number of really successful systems is not very impressive. The main reason why the field of MT has not met the expectations of sci-fi literature as well as of scientific community is the complexity of the task itself. The general opinion is that it is easier to create an MT system for a pair of related languages [Hajic et al. 2000]. We argue that for really close languages it is possible to obtain better translation quality by means of simpler methods. This study presents the evaluation and analysis of direct machine translation system for closely related languages namely Punjabi to Hindi. Although it is true that no consensus exists regarding the best way to evaluate software, there is a general agreement about some of the factors which must be taken into account while deciding what form an evaluation should take.

In our approach, we include the subjective as well as quantitative parameters. Only sentence level translation is considered i.e. a single sentence however long it may

be, is treated as a single unit. Subjective test include Intelligibility Test and Adequacy Test which determines the fitness of an MT system with respect to comprehensibility of translation. Quantitative test include word error rate and sentence error rate. These tests are diagnostic tests to identify limitations, errors and deficiencies of the system. Subjective evaluation is typically performed by potential users and/or purchasers of systems (individuals, companies, or agencies) and diagnostic evaluation is the concern mainly of researchers and developers. In the next sections, we throw some light on Punjabi and Hindi languages. Then we discuss the evaluation results in the remaining section.

2. About Languages

Both Punjabi and Hindi languages have originated from Sanskrit which is one of the oldest language. In terms of speakers, Hindi is third most widely spoken language and Punjabi is twelfth most widely spoken language. Punjabi language is mostly used in the Northern India and in some areas of Pakistan as well as in UK, Canada and USA. Hindi is the national language of India and is spoken and used by the people all over the country.

The script of Punjabi is Gurmukhi. Gurmukhi alphabet was devised during the 16th century by Guru Nanak, the first Sikh guru, and popularised by Guru Angad, the second Sikh guru. It was modelled on the Landa alphabet. The name Gurmukhi means "from the mouth of the Guru".

Notable Features of Gurmukhi

- This is a syllabic alphabet in which all consonants have an inherent vowel. Diacritics, which can appear above, below, before or after the consonant they belong to, are used to change the inherent vowel.
- When they appear at the beginning of a syllable, vowels are written as independent letters.
- When certain consonants occur together, special conjunct symbols are used which combine the essential parts of each letter.
- Punjabi is a tonal language with three tones. These are indicated in writing using the voiced aspirates consonants (gh, dh, bh, etc) and the intervocal h.

Gurmukhi script

Vowels and Vowel diacritics (Laga Matra)

ਅ	ਆ	ਇ	ਈ	ਉ	ਊ	ਏ	ਐ	ਓ	ਔ
Mukta	Kanna	Sihari	Bihari	Onkar	Dunlankar	Lavan	Dulavan	Hora	Kanuara
a	ā	i	ī	u	ū	e	ai	o	au
[ə]	[ɑ]	[ɪ]	[i]	[ʊ]	[u]	[e]	[æ]	[o]	[ɔ]

ਕ	ਕਾ	ਕਿ	ਕੀ	ਕੁ	ਕੂ	ਕੇ	ਕੈ	ਕੋ	ਕੌ
ka	kā	ki	kī	ku	kū	ke	kai	ko	kau

Consonants (Vianjans)

ੳ Ura a, ā, ai, au	ਅ Aira i, ī, e	ੲ Iri u, ū, o	ਸ Sussa sa [sə]	ਹ Haha ha [hə]
ਕ Kakka ka [kə]	ਖ Khukha kha [kʰə]	ਗ Gugga ga [gə]	ਘ Ghugga gha [gə]	ਙ Unga ṅa [ŋə]
ਚ Cucā ca [ʧə]	ਛ Chucha cha [ʧʰə]	ਜ Jujja ja [dʒə]	ਝ Jhujja jha [dʒə]	ਞ Yanza ña [ɲə]
ਟ Tainka ṭa [ʈə]	ਠ Thutha ṭha [ʈʰə]	ਡ Duda ḍa [ɖə]	ਢ Dhuda ḍha [ɖə]	ਣ Nahṇa ṇa [ɳə]
ਤ Tutta ta [tə]	ਥ Thutha tha [tʰə]	ਦ Duda da [də]	ਧ Dhuda dha [də]	ਨ Nunna na [nə]
ਪ Puppa pa [pə]	ਫ Phupha pha [pʰə]	ਬ Bubba ba [bə]	ਭ Bhubba bha [bə]	ਮ Mumma ma [mə]
ਯ Yaiyya ya [jə]	ਰ Rara ra [rə]	ਲ Lulla la [lə]	ਵ Wawa wa [wə]	ੜ Rahṛa ṛa [ɽə]

ਸ਼ ša [ʃə]	ਜ਼ za [zə]	ਫ਼ fa [fə]	ਖ਼ xa [χə]	ਗ਼ ɣa [ɣə]	ਲ਼ la [lə]

Other symbols

ਟਿੱਪੀ (ṭippī) – indicates nasalization.
Used with a, i and u, and also with ū when in final position.

ਬਿੰਦੀ (biṃdī) – indicates nasalization. Used with all other vowels.

ਅੱਦਕ (addak) – doubles the consonant before which it appears.

The Nāgarī (lit. 'of the city') or Devanāgarī ('divine Nagari') alphabet descended from the Brahmi script sometime around the 11th century AD. It was originally developed to write Sanksrit but was later adapted to write many other languages.

Notable Features of Devanagri

- Some scholars use the term alphasyllabary to describe Devanāgarī, while others call it an abugida.

- Consonant letters carry an inherent vowel which can be altered or muted by means of diacritics or *matra*.
- Vowels can be written as independent letters, or by using a variety of diacritical marks which are written above, below, before or after the consonant they belong to. This feature is common to most of the alphabets of South and South East Asia.
- When consonants occur together in clusters, special conjunct letters are used.
- The order of the letters is based on articulatory phonetics.

Devanāgarī alphabet
Primary vowels

	Short		Long		Diphthongs	
	Initial	Diacritic	Initial	Diacritic	Initial	Diacritic
Unrounded low central	अ a	प pa	आ ā	पा pā		
Unrounded high front	इ i	पि pi	ई ī	पी pī		
Rounded high back	उ u	पु pu	ऊ ū	पू pū		
Syllabic variant	ऋ ṛ	पृ pṛ	ॠ ṝ	पॄ pṝ		

Secondary vowels

	Short		Long		Diphthongs	
Unrounded front			ए e	पे pe	ऐ ai	पै pai
Rounded back			ओ o	पो po	औ au	पौ pau

Other symbols

अं aṅ *anusvāra* - nasalises vowel
अँ aṃ *anunāsika/candrabindu* - nasalises vowel

अः aḥ *visarga* - adds voiceless breath after vowel
प् p *virāma* - mutes vowel

Consonants

Occlusives

	Voiceless plosives		Voiced plosives		Nasals	
	unaspirated	aspirated	unaspirated	aspirated		
Velar	क ka	ख kha	ग ga	घ gha	ङ ṅa	
Palatal	च ca	छ cha	ज ja	झ jha	ञ ña	
Retroflex	ट ṭa	ठ ṭha	ड ḍa	ढ ḍha	ण ṇa	
Dental	त ta	थ tha	द da	ध dha	न na	
Labial	प pa	फ pha	ब ba	भ bha	म ma	

Sonorants and fricatives

	Palatal	Retroflex	Dental	Labial
Sonorants	य ya	र ra	ल la	व va
Sibilants	श śa	ष ṣa	स sa	

Other letters

ह ha   ळ ḷa

A selection of conjunct consonants

क्ष kṣa ज्ञ jña त्त tta त्र tra प्य pya त्क tka ट्क ṭka ह्य hya त्व ttva

Except minor differences, most of the characters are same in both the scripts. In Gurmukhi there is no equivalent character for characters "ृ, श्र, ऋ in Devnagri. Some characters of Devnagri are for double sounds like त्र, ज्ञ, क्ष but no such characters are available in Gurmukhi.

Because of same origin, both languages have very similar structure and grammar. The difference is only in words and in pronunciation e.g. in Punjabi the word for boy is ਮੁੰਡਾ and in Hindi it is लड़का. The inflection forms of both these words in Punjabi and Hindi are also similar. There are examples where words are also same but pronunciation is different e.g. ਘਰ and घर. Although Sanskrit is *Sanyogatmik* (Synthetic), Its decendent Hindi is *Viyogatmic* (Analytic)[Singh 1991]. It means we need to add some words, known as prepositions, to convey the relation in Hindi. E.g. In Sanskrit we say बालतः and in Hindi the same meaning is conveyed by adding postposition as बाल का. In comparison to Hindi, Punjabi is also Viyogatmic, but it is not completely Viyogatmik. There are many examples which show that Punjabi is Sanyogatmik also. As for example, The meaning conveyed by two words in Hindi (घर से) is conveyed by one word in Punjabi i.e. ਘਰੋਂ Although it can be written in Punjabi as ਘਰ ਤੋਂ which is Viyogatmic nature, but in general ਘਰੋਂ is more popular. Thus Punjabi is not purely Viyogatmic. It still inherits some properties of its mother language [Singh 1991].

Structurally both Punjabi and Hindi are same. In both languages sentence is comprised of Subject and Predicate. In both languages, the basic elements are *Kaaraka.* Both have eight numbers of *Kaaraka* which by combining with each other create a sentence. The general sequence for transitive Sentence is *Karta, Karam , Kria* e.g. ਰਾਮ ਨੇ ਰੋਟੀ ਖਾਧੀ। and for intransitive sentence is *karta, kriya* e.g. ਰਾਮ ਰੋਇਆ। In both languages the relation between kaarka's are shown by prepositions. The available prepositions in Punjabi are ਦਾ, ਦੇ, ਦੀ, ਦੀਆਂ, ਨੇ, ਨੂੰ etc. and in hindi are का, के, की, को, ने etc. Total eight part-of-speech are recognized in both Punjabi and Hindi. Beside this, both have same types of Nouns, Genders, Number, Person tense and Cases [Singh 1991; Singh and Singh 1986]. Sentence structure is as shown in figure 2.1 and 2.2 on next page.
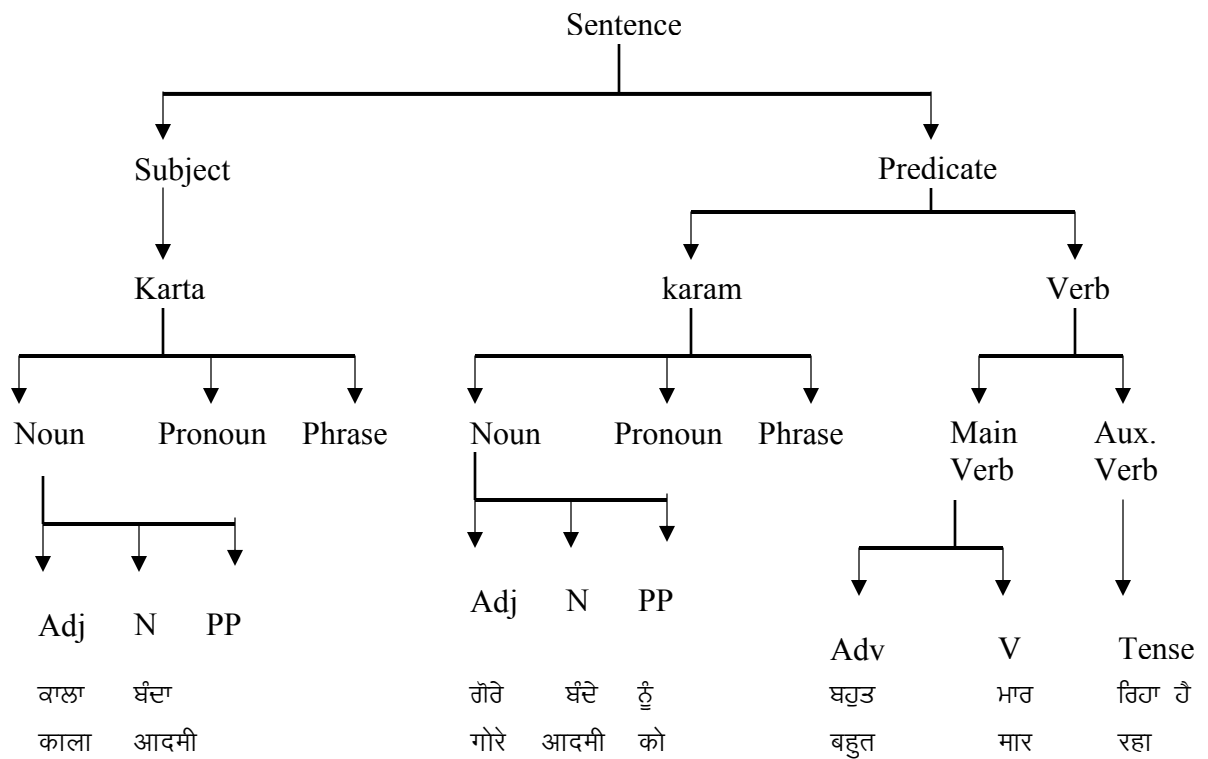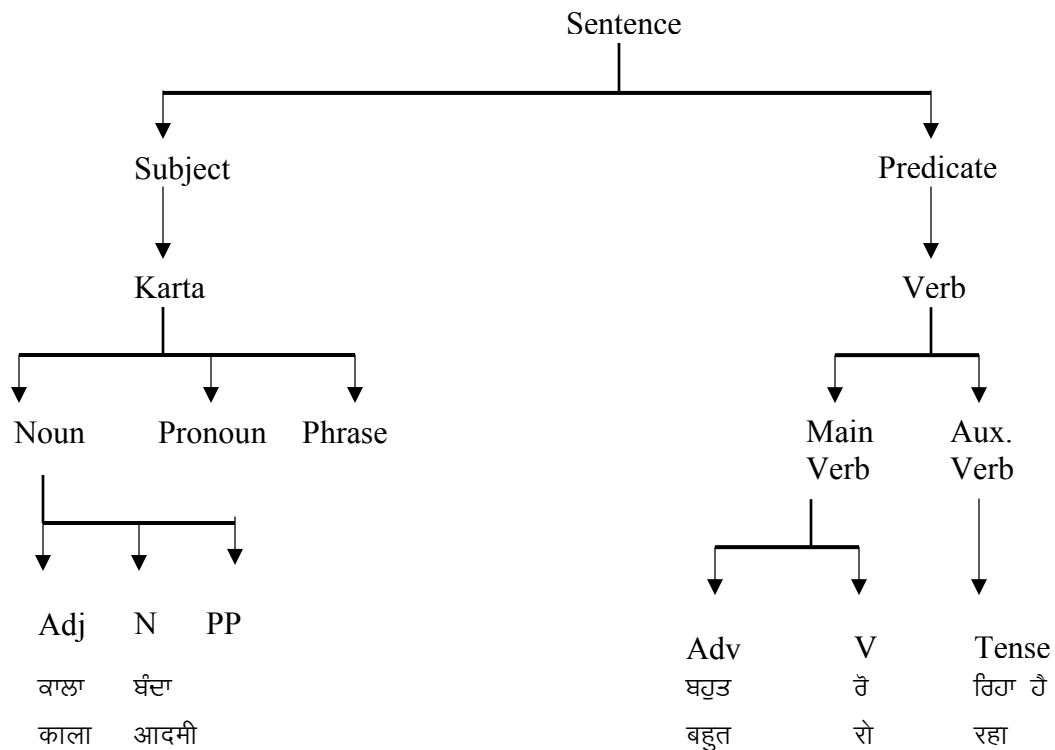
Sentence
├─ Subject
│ └─ Karta
│   ├─ Noun
│   │ ├─ Adj — ਕਾਲਾ / काला
│   │ ├─ N — ਬੰਦਾ / आदमी
│   │ └─ PP
│   ├─ Pronoun
│   └─ Phrase
└─ Predicate
   ├─ karam
   │ ├─ Noun
   │ │ ├─ Adj — ਗੋਰੇ / गोरे
   │ │ ├─ N — ਬੰਦੇ / आदमी
   │ │ └─ PP — ਨੂੰ / को
   │ ├─ Pronoun
   │ └─ Phrase
   └─ Verb
      ├─ Main Verb
      │ ├─ Adv — ਬਹੁਤ / बहुत
      │ └─ V — ਮਾਰ / मार
      └─ Aux. Verb
         └─ Tense — ਰਿਹਾ ਹੈ / रहा

Fig. 2.1 Tree view for Transitive sentences

Sentence
├─ Subject
│ └─ Karta
│   ├─ Noun
│   │ ├─ Adj — ਕਾਲਾ / काला
│   │ ├─ N — ਬੰਦਾ / आदमी
│   │ └─ PP
│   ├─ Pronoun
│   └─ Phrase
└─ Predicate
   └─ Verb
      ├─ Main Verb
      │ ├─ Adv — ਬਹੁਤ / बहुत
      │ └─ V — ਰੋ / रो
      └─ Aux. Verb
         └─ Tense — ਰਿਹਾ ਹੈ / रहा

Fig. 2.2 Tree view for Intransitive sentences

From the direct translation point of view the important differences between Punjabi and Hindi languages are:

Sometime the gender of word is changed in the translated language e.g.

ਸਾਧਨਾਂ ਦੀ ਦੁਰਵਰਤੋਂ   साधनों का दुरुपयोग
ਪੂਰੀ ਵਾਹ ਲਗਾਈ   पुरा जोर लगाया

Some words can be used in both senses i.e. famine and masculine. e.g.

ਦੂਸਰੇ ਪਾਸੇ   दूसरी तरफ
ਦੂਸਰੇ ਲੋਕ   दूसरे लोग

Poorly understood grammar of some constructions e.g.

ਇਹ ਨਹੀਂ ਸੀ ਹੋ ਸਕਦਾ। is a valid sentence in Punjabi but
यह नही था हो सकता। is not a valid sentence in Hindi

For some phrase structure, a preposition is inserted but in some cases the rule is not followed as shown in following example..

ਵਰਤਣ ਲਈ   **उपयोग के लिए**
ਇਸ ਲਈ   **इस लिए**

In the following sections, we discuss the evaluation and results of direct translation system for the language pair of Punjabi and Hindi.

3. History

The first attempt to verify the hypothesis that related languages are easier to translate started in mid 80s at Charles University in Prague [FEMTI; Hajic et al. 2000]. The project was called RUSLAN and aimed at the translation of documentation in the domain of operating systems for mainframe computers. From that date to till date so many examples are there in history which support the argument that with close languages, the quality of MT system, with simple techniques, is better. To name a few one are CESILKO (a system for translating Czech and Slovak), MT system for translating Turkish To Crimean Tatar etc. We are also trying to strengthen the same concept by experimenting with a direct translation system for Punjabi to Hindi. These languages are very closely related and have many features in common.

4. System description

To start with, a direct translation system is created in which words from source language are chosen, their equivalents in target language are found out from the lexicon and are replaced to get target language. For the words with multiple meanings, the most frequently used meaning is selected. Translations are obtained from this system and made available to the evaluators. For Intelligibility test, the evaluators do not have any clue about the source language. They judge each sentence on the basis of its comprehensibility. The target user is a layman who is interested only in the comprehensibility of translations. The scoring is done based on the degree of intelligibility and comprehensibility. For Adequacy test, evaluators are provided with source text along with translated text. The evaluators give score to each translation

according to the scoring scheme (as discussed in next section). On the basis of these scores results are generated using simple statistical techniques.

Error analysis is done with the help of error list which is prepared in advance in consultation with the linguists. Word error rate and sentence error rate are found out. This analysis helps to improve the performance of an MT system

The rest of paper discusses evaluation methodology, result and conclusion for future work.

5. Evaluation Techniques and Methodology:

Based on the previous approaches, following evaluation methods and techniques are applied.

5.1 Selection set of sentences: Input sentences are chosen randomly from newspapers, articles, reviews, court's orders, stories, office letters and people's day to day conversations. All possible constructs including simple as well as complex ones are incorporated in the set. The sentence set also contains all types of sentences such as declarative, interrogative, imperative and exclamatory. The size of input is shown in table 5.1.

Table 5.1 Size of Input for Direct MT System

	Stories	Essays	Court Orders	Office letters	People's conversation
Total Articles	10	10	10	10	10
Total Sentences	360	2340	340	166	284
Total Words	1564	10068	1874	876	872

5.2 The Tests:
Following tests are selected to check the creditability as well as quantitative analysis of the system [FEMTI; Marrafa et al. 2001; Tomas et al. 2003; Wagner].

5.2.1 Subjective tests:
- **Intelligibility Test:** It is a subjective test which is used to check how intelligible a system is? Intelligibility is effected by grammatical errors, miss-translations, and un-translated words.
- **Accuracy Test:** A highly intelligible output sentence need not be a correct translation of the source sentence. It is important to check whether the meaning of the source language sentence is preserved in the translation. This property is called accuracy.

Any variation between the comprehensibility rating and the fidelity rating is due to additional distortion of the information, which can arise from:

- Loss of information (silence) - example: word not translated
- Interference (noise) - example: word added by the system
- Distortion from a combination of loss and interference - example: word badly translated

5.2.2 Error test:
To check the Error rate of the Direct Translation System, some quantitative metrics are also evaluated. These include:

- **Word Error Rate:** It is defined as %age of words which are to be inserted, deleted, or replaced in the translation in order to obtain the sentence of reference.
- **Sentence Error Rate:** Which is %age of sentences, whose translations have not matched in an exact manner with those of reference

Error analysis is done against pre classified error list. All the errors in translated text were identified and their frequencies were noted. Main categories of errors are:

- Wrongly translated word or expression. e.g. मैं is translated into मैं but in some places it must be मैंने.
- Addition or removal of words. e.g. ਜਾਣ ਨਾਲ should be translated into जाने के साथ but it is translated as जाने साथ.
- Un-translated words. e.g. Non Sense words like धाणी in पानी.धाणी
- Wrong choice of words. e.g. ambiguous words i.e. ਰੱਖ in Punjabi can be understood as रखना (to place) or रख (Reserved area for forest animals)

Errors were just counted and not weighted.

5.3 Scoring Procedure for subjective tests:

The evaluators are provided with four point scale for Intelligibility test and Accuracy test. The scoring scheme is given below

5.3.1 For Intelligibility Test
A Four point scale is made in which highest point is assigned to those sentences that look perfectly alike the target language and lowest point is assigned to the sentence which is un-understandable. The scale looks like:

3.	The sentence is perfectly clear and intelligible. It is grammatical and reads like ordinary text.	**e.g.** जाड़ों के दिन थे। **or** मेरा लड़का छह साल का है।
2.	The sentence is generally clear and intelligible. Despite some inaccuracies, one can understand immediately what it means.	e.g. बरसात बरसता है।
1.	The general idea is intelligible only after considerable study. The sentence contains grammatical errors &/or poor word choice.	e.g. बदलां में लुके चंन द्रिश ते बार वारी चानण का लिशकारा मार कर बदलोटीआं की हनेरे में से उसी को समोहत करते बदलोटीआं के हनेरे में लोप जांदे हैं । (Some word are translated, other are left)
0.	The sentence is unintelligible. Studying the meaning of the sentence is hopeless.	e.g. परंतू वीहवीं सदी दे मूंह झाखरे समें काविमझी भाषा दा युग बीत चुका सी ।

	Even allowing for context, one feels that guessing would be too unreliable.	(i.e whole sentence is transliterated character by character)

Table 5.2 Score Sheet for Intelligibility Test

5.3.2 For Accuracy Test

A Four point scale is made in which highest point is assigned to those sentences that look perfectly alike the target language and lowest point is assigned to the sentence which is un-understandable and unacceptable as described by Van Slype. The scale looks like:

0	Completely Unfaithful. Doesn't make sense.	शायद पत्थर करना भी इसी प्रभाव आधीन होता रहा हो।
1	Barely faithful: less than 50 % of the original information passes in the translation.	अगर पुत्र माथे लग गया न तो समझ भई देश न उचित शहर तो छोड़ ही जाओगे।
2	Fairly faithful: more than 50 % of the original information passes in the translation.	हम तन के कुष्ठों की बात नही करनी।
3	Completely Faithful	जैसे ही वह ठीक हुई उस से अपराधी का नाम पता पुछा गया।

Table 5.3 Score Sheet for Accuracy Test

6. Experiments

The survey is done by the 20 peoples of different professions who know the target language (Hindi) very well. Each person evaluated one translation on one criterion, so that each translation is rated for intelligibility by 10 persons and for accuracy by another 10 persons. Average ratings for the sentences of the individual translations were then summed up (separately according to intelligibility and accuracy) to get the average score. Percentage of accurate sentences and intelligible sentences is also calculated separately by counting down the number of accurate sentences.

7. Results

7.1 Subjective test analysis

Initially Null hypothesis is assumed i.e. the system's performance is NULL. We assume that system is dumb and does not produce any valuable output. By the Intelligibility analysis and Accuracy analysis, we prove this wrong.

Accuracy is measured with the help of a 4 point scale. The figure 2.48 denotes the average score of a sentence in accuracy test. From the accuracy analysis total number of accurate sentence are calculated and then their %age is found out which is come out to be 76.28%.

Overall score for accuracy of the translated text comes out to be 2.48. The accuracy %age for the system is found out to be 76.28%. This is comparable with other similar systems as shown in table 7.1.

MT SYSTEM	Accuracy
RUSLAN	40% correct 40% with minor errors. 20% with major error.
CESILKO (Czech-to-Slovak)	90%
Czech-to-Polish	71.4%
Czech-to-Lithuanian	69%
Our System	76.28%

Table 7.1 Comparative analysis of %age accuracy

Further investigation reveals that from the remaining 23.72%,
- 85.93% sentences achieve a match between 50 to 99% and
- Remaining 11.05% of sentences were marked with less than 50% match against the correct sentences.
- Only 3.02 % sentences are those which are found unfaithful.

A match of lower than 50% does not mean that the sentences are not usable. After some post editing, they can fit properly in the translated text. Percentage accuracy of individual articles is shown in table 7.2.

	Stories	Essays	Court Orders	Office letters	People's conversation
%age Accuracy	68.89	80.08	78.18	77.17	66.13

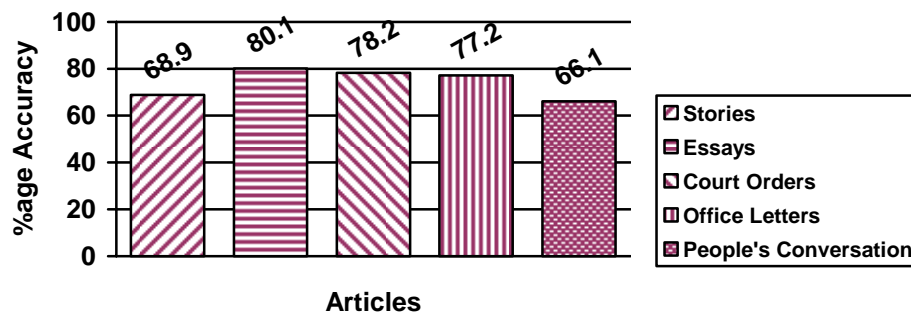**Table 7.2 Percentage Accuracy of different articles**

Figure 7.1 Percentage Accuracy for Different Articles

The main reason that accuracy is less in case of stories and People's conversation is that the language is not standardized. People generally use slang which causes the failure of the translation software as the slang available in one language is not present in other language. Also un-standardized language cause more ambiguities.

The results of Intelligibility test are as follow:

43.17 % of the sentences got the score of 3 i.e. they are perfectly clear and intelligible.
47.64 % of the sentences got the score of 2 i.e. they are generally clear and intelligible.
7.99 % of the sentences got the score of 1 i.e. they are hard to understand.
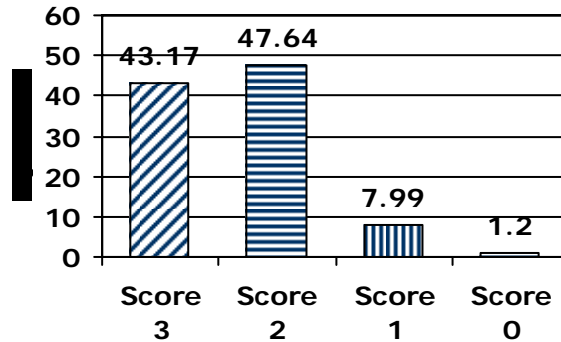**1.2 %** of the sentences cannot be understood at all.

Fig. 7.2 Charts for Intelligibility test.

So we can say that about 90.81% sentences are intelligible. These sentences are those which has score 2 or above. Thus, we can say that the direct approach can translate Punjabi text to Hindi text with a tolerably good accuracy. The sentences which cannot be recognized at all are mostly Idioms and phrases. The sentences, which got the score of 1, includes Idioms and Phrases and sentence containing some non sense words which has no meaning in both languages like पाळी in पाळी-पाळी and ਸ਼ੋਟੀ in ਰੋਟੀ-ਸ਼ੋਟੀ etc.

Percentage intelligibility of individual articles is shown in table 7.2. Again those sentences are selected which has score 2 or above.

	Stories	Essays	Court Orders	Office letters	People's conversation
%age Intelligibility	89.79	93.89	91.79	91.37	87.22

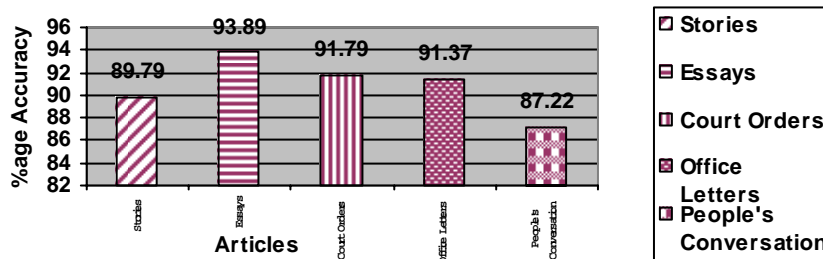**Table 7.3 Percentage Intelligibility of different articles**

Table 7.3 Intelligibility for different articles

7.2 Error Analysis

Error analysis is done for diagnostic evaluation. All errors in the translated text were identified and their frequencies were noted. The types of errors looked for along with their %age in the translation is listed in table 7.3. Word Error rate is found out to be 6.54% which is comparably lower than that of the general systems, where it ranges from 9.5 to 12.

Wrongly translated word or expression.	44.68%
Addition or removal of words.	31.15%
Un-translated words.	9.15%
Wrong choice of words.	15.01%

Table 7.3 Error Analysis
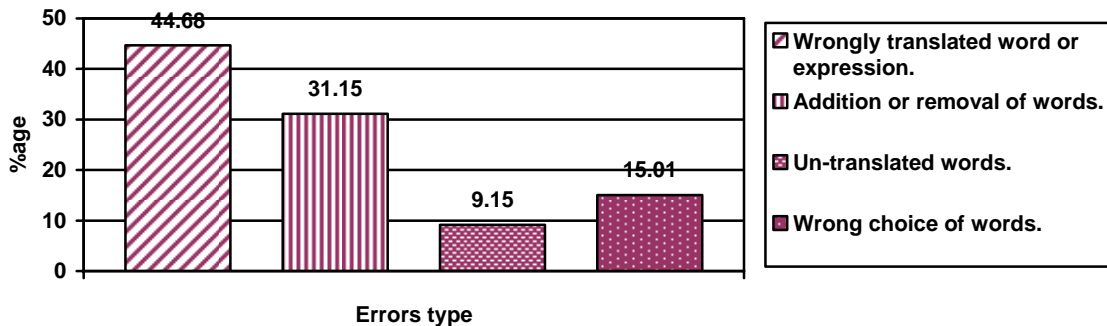
Figure 7.3 Different Types of Errors

As shown in table 7.3, out of 6.54% wrong words, majority of the errors are due to wrongly translated words. The figures show that for improvements in results, bilingual dictionary must be prepared carefully. After enhancing the dictionary, good results are expected. The errors related to addition or removal of words also requires developer's attention. Only 15.01% errors are related to wrong word choice i.e. a problem of word sense disambiguation. Direct approach must be hybrid with some other algorithms for word sense disambiguation so that results with higher accuracy can be produced. Word error rate for individual articles is as shown in following table.

	Stories	Essays	Court Orders	Office Letters	People's Conversation
WER (%age)	6.96	5.03	5.22	5.16	7.13

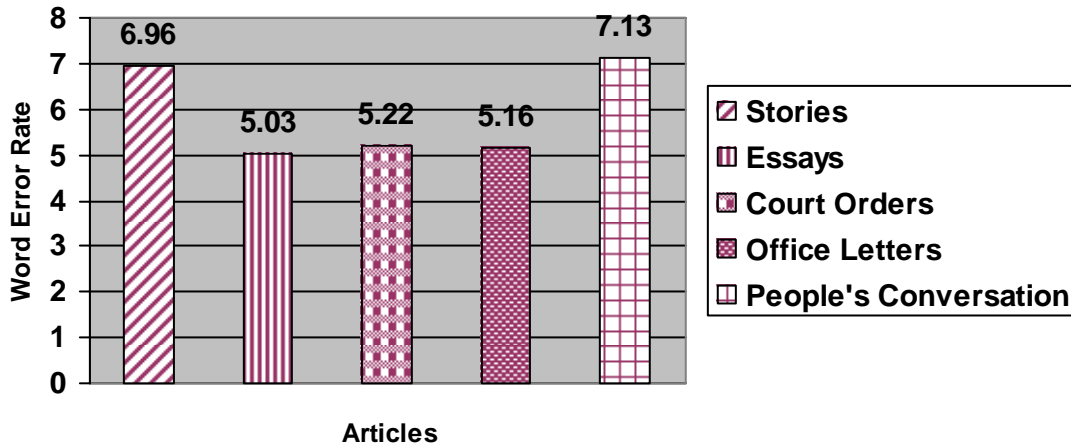**Table 7.4 Word Error Rate of different articles**

Figure 6.4 Word Error Rate for Different Articles

It is found that in stories, word error rate is more i.e. 6.96%. This is due to the fact that language in stories is not standardized. Similar is the case with people's conversation where the error rate is found out to be 7.13% because they use Idioms and phrases more frequently. In articles with standardized language like essays, the word error rate is found out to be 5.03%. For court orders and office letters, the error rate is 5.22% and 5.16% respectively. Again most of the errors are due to post positions adjustments. This suggests that more attention is required in those cases where language is not standardized. In other words for getting higher accuracy, the input should be standardized. The detailed error analysis is shown in following table.

Articles / Type of error(%age)	Stories	Essays	Court Orders	Office Letters	People's Conversation
Wrong Translation	44.53	43.8	44.93	45.01	44.66
Addition or removal of words	19.94	36.11	41.42	41.91	16.37
Wrong word choice (Ambiguities)	17.4	12.88	13.55	12.98	18.4
Un translated words	18.13	7.21	0.1	0.1	20.57

Table 7.5 Error Analysis of Different Articles

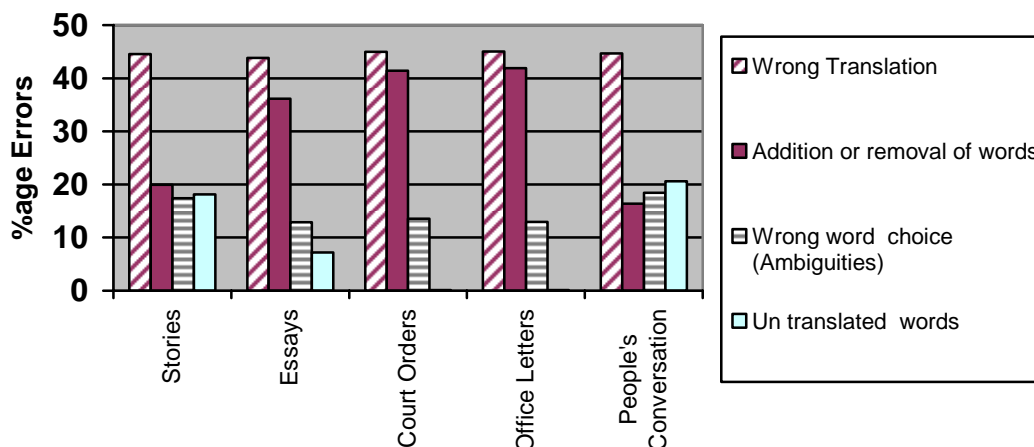## Comparative view of different types of errors

Figure 7.5 Error Analysis of Different Articles

Similarly the Sentence Error rate is found out to be 61.59%. The Sentence error rate for individual articles is as follow:

	Stories	Essays	Court Orders	Office Letters	People's Conversation
SER (%age)	67.3	58.67	57.3	59.13	66.03

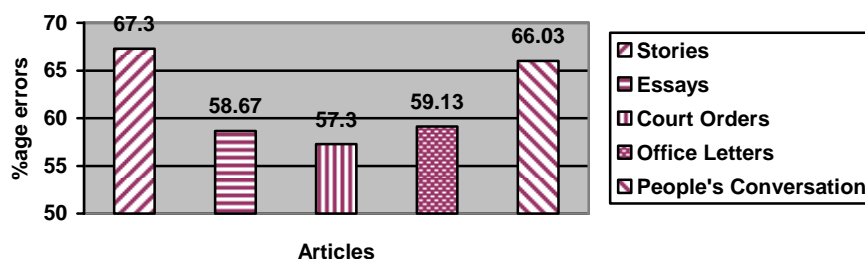**Table 7.6 Sentence Error Rate of Different Articles**

Figure 7.6 Sentence Error Rate of Different Articles

As discussed earlier, the SER of un-standardized matter i.e. stories and people's conversation is higher than the standardized matter. It strengthens the fact that better input gives the better output. If some pre editing of the text is performed then better results may be expected.

8. Multiword Units: Idioms and Adages

Languages are full of idioms and adages such as ਸਿਰ ਤਲੀ ਤੇ ਧਰਨਾ which do not obey the principal of compositionality. The idioms are group of words that have an established meaning that is not apparent from looking at the individual words. An adage

is a short, but memorable saying, which holds some important fact of experience that is considered true by many people, or it has gained some credibility through its long use. The problem with idioms and adages, in an MT context, is that it is not usually possible to translate them using the normal rules. There are exceptions, for example ਚੋਰ ਦੀ ਦਾੜ੍ਹੀ ਵਿਚ ਤਿਣਕਾ (meaning `Guilty person always afraid and doubt that every person is talking about him') can be translated literally into Hindi as "चोर की दाढ़ी में तिनका", which has the same meaning. But, for the most part, the use of normal rules in order to translate idioms will result in nonsense. Instead, one has to treat idioms as single units in translation.

One problem with sentences which contain idioms is that they are typically ambiguous; in the sense that either a literal or idiomatic interpretation is generally possible (i.e. the phrase ਉੱਲੂ ਬੋਲਣੇ can really be about the sound of owl). The real problem with idioms is that they are not generally fixed in their form, and that the variation of forms is not limited to variations in inflection (as it is with ordinary words). Thus, there is a serious problem in recognizing idioms.

Researches are going on to deal with the problems of multiword units in different languages. Many algorithms are proposed but none of them is fully accurate. A full proof system for detecting Idioms and phrases is yet to be developed. We have advantage of language similarity in our case which eliminates most of the problems. Many Idioms and proverbs in Punjabi have similar structure in Hindi. Moreover, ambiguities are also preserved in target language. So, word to word translation can give better results in case of similar languages. To demonstrate this idea, an accuracy test for the most commonly used 200 Idioms and phrases was performed whose results are as shown in table 8.1

	Idioms	Adages
%age Accuracy	72.3%	52.45%

Table8.1 Accuracy of Idioms and Phrases

Score	%age of Idiom	%age of Adages
3	72.37	52.45
2	18.16	24.63
1	7.13	17.89
0	2.34	5.03

Table 8.2 Score of Different Idioms and Phrases

The data shows that about 72.37 % Idioms are common and can be translated word to word with out distorting the meaning. No other system is known that is claiming as much accuracy as the Punjabi-Hindi translation system. This further pushes the idea of having a successful MT system between Punjabi and Hindi. The accuracy in the case of adage is lower because in adages words are absorbed from the local languages which are not standardized and some time no word in the target language is available for the corresponding word in source language. e.g. consider the proverb in Punjabi ਹੱਥ ਪੁਰਾਣੇ ਘੋਸੜੇ ਬਸੰਤੇ ਹੋਰੀ ਆਏ। No word corresponding to ਘੋਸੜੇ & ਬਸੰਤੇ ਹੋਰੀ is found in Hindi language making it difficult for translation.
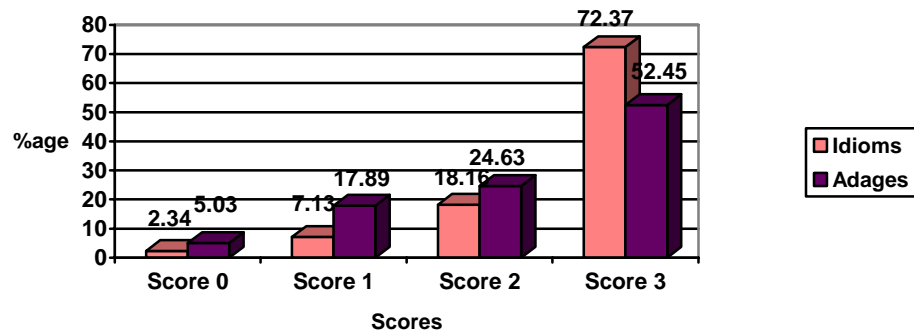
Fig. 8.1 Graph for %age Accuracy

9. Conclusion

The accuracy of the translation achieved by our system justifies the hypothesis that word-for-word translation might also be a solution for language pair of Punjabi and Hindi. The major inaccuracies in the direct translation are due to poor word choice for ambiguous words and some corrections regarding post positions. The lack of information in glossaries and dictionaries sometimes causes an unnecessary translation error. Here it should be worth mentioning that the lexicon used for the system must be exhaustive one, which must contain each word along with every inflection it can have. This will give rise to a huge dictionary. Information retrieval from such dictionary is also a point of consideration.

We can conclude that this study encourages the idea of direct translation by revealing the key problem areas and it shows the next area where concentration is required in order to achieve the objective of creating an MT system for Punjabi to Hindi.

Appendix - Sample Output of Some Articles

Following is some sample input and corresponding output. Coding scheme is as follow:

Bold Text	Completely Unfaithful. Doesn't make sense.
Underline Text	Barely faithful: less than 50 % of the original information passes in the translation.
Italic Text	Fairly faithful: more than 50 % of the original information passes in the translation.
Regular Text	Completely Faithful

ਵਿਆਹ ਵਿਚ ਵੱਟਾ ਕਰਨ ਤੋਂ ਭਾਵ ਆਪਣੀ ਲੜਕੀ ਦਾ ਵਿਆਹ ਉਸੇ ਘਰ ਵਿੱਚ ਕਰਨਾ ਹੁੰਦਾ ਹੈ ਜਿਸ ਘਰ ਵਿੱਚ ਮੁੰਡੇ ਦਾ ਵਿਆਹ ਕੀਤਾ ਹੁੰਦਾ ਹੈ।	विवाह में पत्थर करने से भाव अपनी लड़की का विवाह उसी घर में करना होता है जिस घर में लड़के का विवाह किया होता है
ਸ਼ਾਇਦ ਵੱਟਾ ਕਰਨਾ ਵੀ ਇਸੇ ਪ੍ਰਭਾਵ ਅਧੀਨ ਹੁੰਦਾ ਰਿਹਾ ਹੋਵੇ।	शायद पत्थर करना भी इसी प्रभाव आधीन होता रहा हो
ਆਮ ਤੌਰ ਤੇ ਅਜਿਹੇ ਵੱਟੇ ਸੱਟੇ ਨੀਮ ਪਹਾੜੀ ਇਲਾਕੇ ਜਾਂ ਪਛੜੇ ਇਲਾਕੇ ਵਿਚ ਹੁੰਦੇ ਸਨ।	आम तौर पर ऐसे पत्थर सटे अर्ध पहाड़ी ईलाके या पछड़े ईलाके में होते थे
ਇਹੋ ਜਿਹਿਆਂ ਬਾਰੇ ਲੋਕ ਹਮੇਸ਼ਾਂ ਇਉਂ ਕਹਿੰਦੇ	इस जैसों बारे लोग हमेशा ऐसे कहते सुनते हैं :

ਸੁਣਦੇ ਹਨ: ਦੇਖੀ ਕਿਤੇ ਸਾਲਾ ਮੱਥੇ ਲੱਗ ਜੇ	देखना कहीं साला माथे लग अगर
ਜੇ ਪੁੱਤ ਮੱਥੇ ਲੱਗ ਗਿਆ ਨਾ ਤਾਂ ਸਮਝ ਬਈ ਦੇਸ਼ ਨਾ ਸਹੀ ਸ਼ਹਿਰ ਤਾਂ ਛੱਡ ਈ ਜਾਏਂਗਾ	अगर पुत्र माथे लग गया न तो समझ भई देश न उचित शहर तो छोड़ ही जाओगे
ਇਸ ਵਹਿਸ਼ੀਆਨਾ ਕੁਕਰਮ ਦੇ ਨਤੀਜੇ ਵਜੋਂ ਮਾਇਆ ਬੇਹੋਸ਼ ਹੋ ਗਈ ਅਤੇ ਉਸ ਨੂੰ ਹਸਪਤਾਲ ਲੈ ਜਾਇਆ ਗਿਆ।	इस अभय कुकर्म के परिणाम वश माया बेहोश हो गई और उस को अस्पताल ले व्यर्थ गया
ਵੇ ਭਾਈ! ਇਕੇਰਾਂ ਦੇਖ ਲੈ, ਪਿੱਛੇ ਨੂੰ ਘੜੀ ਮੁੜੀ ਤਕਦਾ ਹੈਂ।	वे भाई ! एक बार देख ले , पीछे को घड़ी मुड़ी देखते हो
ਅਸੀਂ ਇਕ ਮਿੱਤਰ ਨੂੰ ਦੱਸਿਆ।	हम एक मित्र को बताया
ਮਿੱਤਰ ਠਹਾਕਾ ਮਾਰ ਕੇ ਹੱਸ ਪਿਆ ਤੇ ਕਹਿਣ ਲੱਗਾ, ਨਹੀਂ ਇਹ ਗਿਣਤੀ ਏਦੂੰ ਵੀ ਵੱਧ ਐ।	मित्र ठहाका मार कर हँस पड़ा पर कहने लगा, नही यह गिनती इस से भी ज्यादा हैं
ਆਪਾਂ ਤਨ ਦੇ ਕੋੜ੍ਹੀਆਂ ਦੀ ਗੱਲ ਨਹੀਂ ਕਰਨੀ।	हम तन के कुष्ठों की बात नही करनी
ਕੋੜ੍ਹੀਆਂ ਦੀਆਂ ਹੋਰ ਬਹੁਤ ਕਿਸਮਾਂ ਨੇ	कुष्ठों की और बहुत किस्में ने
ਅਸੀਂ ਪੁੱਛਿਆ ਤਾਂ ਮਿੱਤਰ ਅੱਖਾਂ ਮਟਕਾ ਕੇ ਬੋਲਿਆ, ਕੁਝ ਲੋਕ ਸ਼ਕਲ ਦੇ ਕੋੜ੍ਹੀ ਹੁੰਦੇ ਐ	हम पुछा तो मित्र आंखें मटका कर बोला, कुछ लोग शक्ल के कुष्ठी होते हैं
ਉਨ੍ਹਾਂ ਦੀ ਸ਼ਕਲ ਦੇ ਹਮੇਸ਼ਾਂ ਈ ਬਾਰਾਂ ਬਜੇ ਰਹਿੰਦੇ ਐ	उन्हे की शक्ल के हमेशा ही बारह बजे रहते हैं
ਪੰਚਤੰਤਰ ਦੀ ਇਕ ਕਥਾ ਵੀ ਹੈਗੀ ਇਹੋ ਜਿਹਿਆਂ ਵਾਸਤੇ	पंचतंत्र की एक कथा भी है इस जैसों लिए
ਆਪਾਂ ਆਪਣਾ ਰਾਜਾ ਹੋਰ ਚੁਣੀਏ	हम अपना राजा और चुनें
ਸਾਰਿਆਂ ਖੂਬ ਸੋਚ ਵਿਚਾਰ ਕੀਤੀ	सभी खूब सोच विचार की
ਉਹ ਉਨ੍ਹਾਂ ਦੇ ਬੱਚਿਆਂ ਦੀ ਚੰਗੀ ਰਾਖੀ ਕਰੂ	वह उन्हे के बच्चों की अच्छी रक्षा करेगा
ਪੰਛੀਆਂ ਨੇ ਕਾਂ ਨੂੰ ਆਪਣਾ ਫੈਸਲਾ ਦੱਸਿਆ	पक्षियों ने कौआ को अपना फैसला बताया
ਸੱਚਮੁੱਚ ਉਸ ਦੇ ਜੀਵਨ ਵੇਰਵੇ ਇਨਸਾਨੀ ਜੀਵਨ ਦੀਆਂ ਸੰਭਾਵਨਾਵਾਂ ਦੀ ਬੜੀ ਰੌਸ਼ਨ ਝਾਕੀ ਪ੍ਰਸਤੁਤ ਕਰਦੇ ਹਨ।	सचमुच उस के जीवन व्योरे इन्सानी जीवन की सम्भावनाएं की बहुत रौशन झांकी प्रस्तुत करते हैं
ਮਾਇਆ ਅਜੇ ਮਸਾਂ ਸੱਤ ਕੁ ਵਰ੍ਹੇ ਦੀ ਮਾਸੂਮ ਬਾਲੜੀ ਹੀ ਸੀ ਕਿ ਉਸ ਦੀ ਮਾਂ ਦੇ ਇਕ ਪ੍ਰੇਮੀ ਨੇ ਉਸ ਨਾਲ ਬਲਾਤਕਾਰ ਦਾ ਘਿਨੌਣਾ ਕੁਕਰਮ ਕੀਤਾ।	माया अभी मुश्किल से सात सा वर्ष की मासूम अनजान लड़की ही थी कि उस की माता के एक प्रेमी ने उस के साथ बलात्कार का घिनौना कुकर्म किया
ਰਿਹਾਈ ਤੋਂ ਬਾਅਦ ਅਜੇ ਉਹ ਆਪਣੇ ਘਰ ਨੂੰ ਵਾਪਸ ਹੀ ਜਾ ਰਿਹਾ ਸੀ ਕਿ ਲੋਕਾਂ ਦੀ ਭੜਕੀ ਹੋਈ ਭੀੜ ਨੇ ਉਸ ਤੇ ਹਮਲਾ ਕਰ ਕੇ ਉਸ ਨੂੰ ਕੁੱਟ ਕੁੱਟ ਕੇ ਮਾਰ ਦਿੱਤਾ।	रिहाई से बाद अभी वह अपने घर को वापिस ही जाओ रहा था कि लोगों की भड़की हुई भीड़ ने उस पर हमला कर कर उस को पीट पीट कर मार दिया
ਮਾਇਆ ਐਂਜਲਿਯੂ ਨੂੰ ਮਹਿਸੂਸ ਹੋਇਆ ਕਿ ਜੇ ਉਹ ਅਪਰਾਧੀ ਬਾਰੇ ਨਾ ਬੋਲਦੀ ਤਾਂ ਵਿਚਾਰੇ ਦੀ ਜ਼ਿੰਦਗੀ ਖਤਮ ਨਹੀਂ ਸੀ ਹੋਣੀ।	माया अैंजलियू को महसूस हुआ कि अगर वह अपराधी बारे न बोलती तो बेचारे की ज़िन्दगी समाप्त नही थी होनी
ਉਸ ਨੂੰ ਆਪਣੇ ਤੇ ਗਿਲਾ ਹੋਇਆ ਅਤੇ ਪਸ਼ਚਾਤਪ ਦੀ ਖਾਤਰ ਉਸ ਨੇ ਪੰਜ ਵਰ੍ਹਿਆਂ ਵਾਸਤੇ ਬਿਲਕੁਲ ਵੀ ਨਾ ਬੋਲਣ ਦੀ ਸਹੁੰ ਖਾ ਲਈ ਜੋ ਉਸ ਨੇ ਪੂਰੀ ਤਰ੍ਹਾਂ ਨਿਭਾਈ।	उस को अपने पर गिला हुआ और पश्चात्ताप की खातिर उस ने पाँच सालों लिए बिल्कुल भी न बोलने की कसम खा के लिए जो उस ने पूर्ण तरह निभाई
ਮੈਂ ਤਾਂ ਪਰਸੋਂ ਡਿਊਟੀ ਦੇਣ ਉਪਰੰਤ ਨੌਂ ਵਜੇ ਹੀ ਕਰ ਸਕਾਂਗਾ।	मैं तो परसों डिउटी देने के पश्चात नौ बजे ही कर सकूंगा

ਨਾਲ ਦੀ ਸਵਾਰੀ ਨੂੰ ਵਾਰ ਵਾਰ ਤੰਗ ਕਰਦਾ ਹੈ।	साथ की स्वारी को बार बार तंग करता हो
ਇੱਕ ਸਬਜ਼ੀ ਵੇਚਣ ਵਾਲਾ ਆਪਣੇ ਆਲੇ ਦੁਆਲੇ ਰੱਖੀਆਂ ਸਬਜ਼ੀਆਂ ਦੇ ਭਾਅ ਗਾਹਕਾਂ ਨੂੰ ਦੱਸਦਾ ਹੈ।	एक सबज़ी बेचने वाला अपने आले चारों और रखी सबज़ीयों के भाव ग्राहकों को बताता है
ਤੁਸੀਂ ਪਾਣੀ ਪੀਣ ਲਈ ਕਿਸ ਤੋਂ ਪੁੱਛ ਕੇ ਗਏ ਸੀ ਤੇ ਕੀਹਨੂੰ ਦੱਸ ਕੇ ਗਏ ਸੀ?	आप पानी पीने के लिए किस से पूछ कर गए थे पर किसे बता कर गए थे ?
ਇਹ ਇਕ ਰਿਵਾਜ਼ ਨਹੀ ਸੀ ਬਲਕਿ ਮੁਸਲਮਾਨੀ ਦੇ ਵੇਲੇ ਤੋਂ ਇਕ ਪਿਰਤ ਜਿਹੀ ਹੀ ਚਲੀ ਆ ਰਹੀ ਸੀ।	यह एक रिवाज़ नही था बल्कि मुसलमानी के समय से एक प्रथा जैसी ही चली आ रही थी
ਹਿੰਦੂ ਸਮਾਜ ਵਿਚ ਪਹਿਲਾਂ ਪਹਿਲ ਲੜਕੀ ਦਾ ਰਿਸ਼ਤਾ ਕਰਨ ਸਮੇ ਦੋਹਾਂ ਪਾਸਿਆਂ ਦੇ ਤਿੰਨ ਤਿੰਨ ਗੋਤ ਵਿਚਾਰ ਜਾਂਦੇ ਸਨ।	हिन्दु समाज में पहले पहिल लड़की का रिश्ता करने समय दोनों तरफ के तीन तीन गोत्र विचार जाते थे
ਭਾਵ ਦੋਨਾਂ ਤਰਫ ਤੋਂ ਬਾਪ ਦਾ ਗੋਤ, ਮਾਂ ਦਾ ਗੋਤ ਅਤੇ ਦਾਦੀ ਦਾ ਗੋਤ ਵਿਚਾਰੇ ਜਾਂਦੇ ਸਨ।	भाव दोनो तरफ से पिता का गोत्र , माता का गोत्र और दादी का गोत्र बेचारे जाते थे
ਰਜਿਸਟਰਾਰ ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਪਟਿਆਲਾ ਅਤੇ ਡਾਇਰੈਕਟਰ ਯਾਦਵਿੰਦਰਾ ਕਾਲਜ ਆਫ ਇੰਜਿਨੀਅਰਿੰਗ ਤਲਵੰਡੀ ਸਾਬੋ ਵੱਲੋਂ ਉਪਰੇਟ ਸਟੇਟ ਬੈਂਕ ਆਫ ਪਟਿਆਲਾ ਪੰਜਾਬੀ ਯੂਨੀਵਰਸਿਟੀ ਪਟਿਆਲਾ ਵਿਖੇ ਖਾਤਾ ਨੰ: 50068 ਵਿੱਚ ਇਸ ਕਾਲਜ ਦੇ ਵਿਦਿਆਰਥੀਆਂ ਨੂੰ ਸਕਾਲਰਸ਼ਿਪ ਦੇਣ ਹਿਤ ਦਾਨੀ ਪੁਰਸ਼/ਸੰਸਥਾਵਾਂ ਵੱਲੋਂ ਪ੍ਰਪਤ ਰਾਸ਼ੀ ਜਮਾ ਕਰਵੀ ਗਈ ਹੈ।	रजिसटरार पंजाबी विश्व विद्धाल्य पटियाला और डायरैक्टर यादविंदरा कालेज आफ ईंजीनियरिंग तलवंडी साबो द्वारा उुपरेट सटेट बैंक आफ पटियाला पंजाबी विश्व विद्धाल्य पटियाला में खोले खाता नं:50068 में इस कालेज के विद्यार्थीयों को सकालरशिप देने हित दानी पुरूष/संसथांएं तरफ से प्राप्त राशी जमां करवाई गई है
ਬੈਂਕ ਖਾਤੇ ਵਿਚ ਰਾਸ਼ੀ ਟਰਾਂਸਫਰ ਕਰਨ ਸਬੰਧੀ ਵਾਈਸ ਚਾਂਸਲਰ ਸਾਹਬ ਦੇ ਆਦੇਸ਼ ਪਹਿਲਾਂ ਹੀ ਲਏ ਸਨ ਜੋ ਕਿ ਫਾਈਲ ਵਿਚ ਹਨ ਜੀ।	बैंक खाते में राशी टरांसफर करने सम्बंध में वाईस चांसलर साहब के आदेश पहले ही लिए थे जोकि फाईल में हैं जी
ਕੋੜੀਆਂ ਦੀ ਬਸਤੀ	कुष्ठों की बसती
ਦੁਨੀਆਂ ਦੇ ਇਕ ਤਿਹਾਈ ਕੋੜੀ ਸਾਡੇ ਦੇਸ਼ 'ਚ ਰਹਿੰਦੇ ਐ।	दुनीयां के एक तिहाई कुष्ठी हमारे देश में रहते हैं
ਆਪਾਂ ਗਾਂਧੀ ਦੇ ਚੇਲੇ ਆਂ	हम गांधी के चेले हैं !
ਤਨ ਦੇ ਕੋੜੀਆਂ ਤੋਂ ਇਲਾਵਾ ਵੀ ਹੋਰ ਕੋੜੀ ਹੁੰਦੇ ਐ?	तन के कुष्ठों से अतिरिक्त भी और कुष्ठी होते हैं !
ਚਿਹਰਾ ਇਉਂ ਹੁੰਦੈ ਜਿਵੇ ਹੜਤਾਲ ਹੋਈ ਹੋਵੇ	चेहरा ऐसे होता है जैसे हड़ताल हुई हो
ਉਸ ਦੇ ਦਰਸ਼ਨ ਕਰਨੇ ਵੀ ਅਗਲਾ ਪਾਪ ਸਮਝਦੈ	उस के दर्शन करने भी अगला पाप समझता है
ਇਕ ਵਾਰ ਪੰਛੀਆਂ ਨੇ ਇਕੱਠ ਕੀਤਾ ਅਤੇ ਮਤਾ ਪਕਾਇਆ ਬਈ ਗਰੁੜ ਤਾਂ ਵਿਸ਼ਨੂੰ ਦੇ ਈ ਵੜਿਆ ਰਹਿੰਦੈ	एक बार पक्षियों ने इकट्ठ किया और फैसला पकाया भई गरूड़ तो विष्णु के ही घूसा रहता है
ਕਾਂ ਉਸ ਸਭਾ ਵਿੱਚੋਂ ਗੈਰਹਾਜ਼ਰ ਸੀ	कौआ उस सभा में से गैरहाजिर था
ਅਖੀਰ ਫੈਸਲਾ ਇਹ ਕੀਤਾ ਕਿ ਉੱਲੂ ਰਾਤ ਨੂੰ ਜਾਗਦਾ ਰਹਿੰਦੈ	अंत फैसला यह किया कि उल्लु रात को जागता रहता है
ਪੰਛੀਆਂ ਨੇ ਆਮ ਸਹਿਮਤੀ ਨਾਲ ਉੱਲੂ ਨੂੰ ਰਾਜਾ ਚੁਣ ਲਿਆ	पक्षियों ने आम सहमती के साथ उल्लु को राजा चुन लिया

ਇੰਨੇ ਨੂੰ ਕਾਂ ਵੀ ਆ ਪਹੁੰਚਿਆ	इतने को कौआ भी आ पहुंचा
ਪਿਛਲੇ ਦਿਨੀ ਅੰਗਰੇਜ਼ੀ ਟ੍ਰਿਬਿਊਨ ਵਿਚ ਇਕ ਐਫਰੋ – ਅਮਰੀਕਨ ਲੇਖਕਾ ਮਾਇਆ ਐਂਜਲਿਯੂ ਬਾਰੇ ਕੁਝ ਸਮੱਗਰੀ ਪ੍ਰਕਾਸ਼ਿਤ ਹੋਈ ਸੀ।	पिछले दिनों अंग्रेज़ी ट्रिब्यून में एक अैफरो अमरीकन लेखिका माया अैंजलियू बारे कुछ सामग्री प्रकाशित हुई थी
ਜਿਵੇਂ ਹੀ ਉਹ ਠੀਕ ਹੋਈ ਉਸ ਤੋਂ ਅਪਰਾਧੀ ਦਾ ਨਾਮ ਪਤਾ ਪੁੱਛਿਆ ਗਿਆ।	जैसे ही वह ठीक हुई उस से अपराधी का नाम पता पुछा गया
ਉਸ ਨੇ ਸਭ ਕੁਝ ਦੱਸ ਦਿੱਤਾ।	उस ने सभी कुछ बता दिया
ਅਪਰਾਧੀ ਤੇ ਮੁਕੱਦਮਾ ਤਾਂ ਚੱਲਿਆ ਪਰ ਉਸ ਨੂੰ ਜ਼ਮਾਨਤ ਤੇ ਰਿਹਾਅ ਵੀ ਕਰ ਦਿੱਤਾ ਗਿਆ।	अपराधी पर मुकदमा तो चला लेकिन उस को ज़मानत पर मुक्त भी कर दिया गया
ਉਸ ਤੋਂ ਬਾਅਦ ਵੀ ਉਹ ਬੋਲਣ ਕ੍ਰਿਆ ਬਾਰੇ ਬਹੁਤ ਸੁਚੇਤ ਰਹੀ।	उस से बाद भी वह बोलने क्रिया बारे बहुत सुचेत रही
ਉਹ ਜਾਣਦੀ ਸੀ ਕਿ ਬੋਲ ਬਾਣੀ ਕਿਵੇਂ ਲੋਕਾਂ ਅਤੇ ਕੌਮਾਂ ਨੂੰ ਇਕ ਦੂਜੇ ਦੇ ਲਹੂ ਦੇ ਤਿਹਾਏ ਬਣਾ ਦਿੰਦੀ ਹੈ।	वह जानती थी कि बोल वाणी कैसे लोगों और कौमों को एक दूसरे के रक्त के प्यासे बना देती है
ਬੋਲਣਾ ਸੱਚਮੁੱਚ ਹੀ ਇਕ ਹਿੰਸਾਤਮਕ ਕ੍ਰਿਆ ਹੈ।	बोलना सचमुच ही एक हिंसातमक क्रिया है
ਜੇ ਤੁਸੀ ਅਹਿੰਸਾਵਾਦੀ ਰਹਿਣਾ ਹੈ ਤਾਂ ਕੁਝ ਨਾ ਬੋਲੋ।	अगर आप अहिंसावादी रहना है तो कुछ न बोलो
ਇਹ ਕੰਮ ਨਾ ਮੈਂ ਕੱਲ੍ਹ ਕਰ ਸਕਿਆ ਹਾਂ, ਨਾ ਅੱਜ ਕਰ ਸਕਾਂਗਾ, ਭਲਕੇ ਵੀ ਵਿਹਲਾ ਨਹੀਂ।	यह काम न मैं कल कर सका हुँ , न आज कर सकूंगा , कल भी फुरसत में नही
ਤੁਸੀ ਚਾਹੁੰਦੇ ਹੋ ਇਹ ਕੰਮ ਤੁਰੰਤ ਹੋ ਜਾਵੇ।	आप चाहते हो यह काम तुरन्त हो जाए
ਰਾਤੀ ਜਾਂ ਐਤਵਾਰ ਨੂੰ ਮੈਂ ਵਿਹਲਾ ਨਹੀਂ ਹੋਵਾਂਗਾ।	रात को या रविवार को मैं फुरसत में नही हूँगा
ਤੈਥੋਂ ਸਿੱਧੇ ਹੋ ਕੇ ਨਹੀਂ ਬੈਠ ਹੁੰਦਾ, ਖ਼ਬਰਦਾਰ ਜੇ ਪਿੱਛੇ ਦੁਬਾਰਾ ਤੱਕਿਆ।	तेरे से सीधे हो कर नही बैठ होता , खबरदार अगर पीछे दुबारा देखा
ਕਈ ਵਾਰ ਤਾਂ ਉਹ ਹਰ ਸਬਜ਼ੀ ਦਾ ਨਾਂ ਲੈ ਕੇ ਭਾਅ ਦਸਦਾ ਹੈ, ਪਰ ਕਈ ਵਾਰੀ ਕਹਿੰਦਾ ਹੈ, ਲਓ ਜੀ, ਇਹ ਦੋ ਰੁਪਏ ਕਿਲੋ, ਅਹੁ ਸਵਾ ਰੁਪਈਆ ਕਿਲੋ।	कई बार तो वह हर सबजी का नाम ले कर भाव बताता है , लेकिन कई बार कहता है , लो जी , यह दो रूपए किलो , वह सवा रूपया किलो
ਨਹੀਂ ਜੀ, ਉਹ ਤੁਹਾਡੇ ਖਰੀਦਣ ਦੀ ਨਹੀਂ, ਬਾਸੀ ਹੈ।	नही जी , वह आपके खरीदने की नही , बासी है
ਮੁਸਲਮਾਨ ਸਮਾਜ ਵਿੱਚ ਇਕ ਦੁੱਖ ਦਾ ਰਿਸ਼ਤਾ ਭਾਵ ਭੈਣ ਭਰਾ ਛੱਡ ਕੇ ਹੋਰ ਸਾਰੇ ਰਿਸ਼ਤੇਦਾਰਾਂ ਵਿਚ ਵਿਆਹ ਹੋ ਜਾਂਦਾ ਹੈ।	मुसलमान समाज में एक दुःख का रिश्ता भाव बहन भाई छोड़ कर और सभी रिश्तेदारों में विवाह हो जाता है
ਬਾਅਦ ਵਿਚ ਦੋ ਗੋਤ ਅਤੇ ਹੁਣ ਆ ਕੇ ਇਕੋ ਗੋਤ ਵਿਚਾਰਿਆ ਜਾਂਦਾ ਹੈ।	बाद में दो गोत्र और अब आ कर एक गोत्र विचारा जाता है
ਹਿੰਦੂ ਅਤੇ ਸਿੱਖ ਸਮਾਜ ਵਿੱਚ ਉਸੇ ਗੋਤ ਵਿਚ ਲੜਕੀ ਦਾ ਵਿਆਹ ਕਰਨਾ ਠੀਕ ਨਹੀਂ ਸਮਝਿਆ ਜਾਂਦਾ।	हिन्दु और सिक्ख समाज में उसी गोत्र में लड़की का विवाह करना ठीक नही समझा जाता
ਗੋਤਾਂ ਦੀ ਪਰਖ ਵਿਚ ਸ਼ਾਇਦ ਇਕ ਵਿਗਿਆਨਕ ਪੱਖ ਕੰਮ ਕਰਦਾ ਹੈ।	गोत्रों की परख में शायद एक वैज्ञानिक पक्ष काम करता है
ਇਸ ਖਾਤੇ ਵਿੱਚੋਂ ਲੱਖ ਰੁਪਏ ਤਲਵੰਡੀ ਸਾਬੋ ਦੇ ਖਾਤਾ ਨੰ: 50110 ਵਿਚ ਟਰਾਂਸਫਰ ਕਰਵਾਣ ਲਈ ਸਟੇਟ ਬੈਂਕ ਆਫ ਪਟਿਆਲਾ ਪੰਜਾਬੀ	इस खाते में से 20 लाख रूपए तलवंडी साबो के खाता नं: 50110 में ट्रांसफर करवाने के लिए सटेट बैंक आफ पटियाला पंजाबी विश्व विद्दालय के नाम

ਯੂਨੀਵਰਸਿਟੀ ਦੇ ਨਾਮ ਪੱਤਰ ਭੇਜਿਆ ਜਾ ਰਿਹਾ ਹੈ।	पत्र भेजा जा रहा है
ਕਿਰਪਾ ਕਰਕੇ ਇਸ ਤੇ ਹਸਤਾਖਰ ਕਰਨਾ ਤਾਂ ਜੋ ਬੈਂਕ ਨੂੰ ਇਹ ਰਾਸ਼ੀ ਟਰਾਂਸਫਰ ਕਰਨ ਲਈ ਪੱਤਰ ਭੇਜਿਆ ਜਾ ਸਕੇ।	कृप्या करके इस पर हसताक्षर करना तो जो बैंक को यह राशी ट्रांसफर करने के लिए पत्र भेजा जा सके

Reference:

Bharati A, Moona R., Singh S., Sangal R., Sharma D.S., "MTeval: An evaluation methodology for Machine Translation Systems", Proceedings of the SIMPLE-Symposium on Indian Morphology, Phonology and Language Engineering 2004, IIT Kharagpur, INDIA

FEMTI - a Framework for the Evaluation of Machine Translation in ISLE
http://www.isi.edu/natural-language/mteval/

Hajic J, Hric J, Kubon V., "Machine Translation of Very Close Languages", proceedings of the 6th Applied Natural Language Processing Conference, April 29--May 4, 2000, Seattle, Washington, USA. pp 7-12

Marrafa, Palmira and Ribeiro A., "Quantitative Evaluation of Machine Translation Systems: Sentence level", Proceedings of the MT Summit VIII Fourth ISLE workshop 2001, spain, pp. 39-43.

Singh H. "*Saadian Bhashawan*", ed. 1, Punjabi Academy, Delhi. 1991.

Singh H. Singh L. "College Punjabi Viakaran". Punjab State University text Book Board, Chandigarh. 1986.

Tomas J, Mas J. A., Casacuberta F., "A Quantitative Method for Machine Translation Evaluation", presented in workshop of 11th Conference of the European Chapter of the Association for Computational Linguistics April 12-17, 2003 Agro Hotel, Budapest, Hungary.

Wagner S, "Small Scale Evaluation Method", from website http://www.ifi.unizh.ch/CL/swagner/SmallScale.rtf.

[1]. Bharati A, Moona R., Singh S., Sangal R., Sharma D.S., "MTeval: An evaluation methodology for Machine Translation Systems", Proceedings of the SIMPLE-Symposium on Indian Morphology, Phonology and Language Engineering 2004, IIT Kharagpur, INDIA
[2]. HAJIC J, HRIC J, KUBON V., "Machine Translation of Very Close Languages", proceedings of the 6th Applied Natural Language Processing Conference, April 29--May 4, 2000, Seattle, Washington, USA. pp 7-12

[3]. Marrafa, Palmira and Ribeiro A., "Quantitative Evaluation of Machine Translation Systems: Sentence level", Proceedings of the MT Summit VIII Fourth ISLE workshop 2001, spain, pp. 39-43.

[4] Singh H. "*Saadian Bhashawan*", ed. 1, Punjabi Academy, Delhi. 1991.

[5] Singh H. Singh L. "College Punjabi Viakaran". Punjab State University text Book Board, Chandigarh. 1986.

[6]. Slype V., "Critical Methods for Evaluating the Quality of Machine Translation (Final Report)", prepared for the commission of the European committees, Brussels.

[7]. Tomas J, Mas J. A., Casacuberta F., "A Quantitative Method for Machine Translation Evaluation", presented in workshop of 11th Conference of the European Chapter of the Association for Computational Linguistics April 12-17, 2003 Agro Hotel, Budapest, Hungary.

[8]. Wagner S, "Small Scale Evaluation Method", from website
http://www.ifi.unizh.ch/CL/swagner/SmallScale.rtf.

[9]. FEMTI - a Framework for the Evaluation of Machine Translation in ISLE
http://www.isi.edu/natural-language/mteval/

--

--