Interactive English to Urdu Machine Translation using Example-Based Approach

Maryam Zafar Department of Computer Science Military College of Signals, NUST Rawalpindi, Pakistan <u>maryamzafar@yahoo.com</u>

Abstract—This work is first attempt towards English to Urdu Machine Translation (MT) using example based approach. We have developed an interactive MT system to facilitate the user to customize the translation to his needs, thereby improving the performance of the translation. Our MT system supports idioms, homographs, and some other features in addition to the ability of the bilingual corpus to evolve. In the end, we compare the features of the MT system developed by the center for research in Urdu Language Processing (CRULP) with those of our MT system.

Keywords-Machine Translation; Bilingual Corpus

I. INTRODUCTION

Knowledge is the key to progress and English is one language that preserves a tremendous amount of knowledge. There is a huge literature of Sciences and Engineering available in English. Rightly so, English is believed to be an international language. It becomes more important to be able to understand English.

Urdu is a language that is spoken in all over Pakistan and many parts of India and some other South Asian countries [1]. This makes Urdu a very important language.

Only 5% to 10% people of Pakistan are familiar with English [2]. To be able to move with the world, people of Pakistan must get the latest knowledge but the current standing in terms of literacy and the understanding of English finds a gulf that needs to be abridged. This suggests that there is a need of the interface between the two languages, i.e. English and Urdu.

Solution comes in the form of Machine Translation (MT). MT refers to the use of computing resources to facilitate the translation of the content available in one language to its equivalent content in another language [3].

There are three main techniques for MT known as Rule based Machine translation (RBMT), Statistical Machine Translation (SMT), and Example Based Machine Translation (EBMT). RBMT depends upon linguistic rules to carry out translation from one language to another. It performs morphological, syntactic and semantic analysis on the language and then transfers it into target language [4]. SMT is a corpus based approach that uses statistical models to carry out translation [5]. Asif Masood

Department of Computer Science Military College of Signals, NUST Rawalpindi, Pakistan amasood@mcs.edu.pk

Third approach for MT is Example Based Machine Translation (EBMT). It was introduced by Nagao [6] in 1984. Using EBMT, translation is carried out by decomposing English sentence into fragments, then finding corresponding translation for those fragments in Urdu and then recombining translated fragment into Urdu sentence.

English and Urdu are structurally different languages [1]. For structurally different languages EBMT is a better choice than others [7]. This suggests us to use Example based approach for our MT system.

A quality translation cannot be developed now unless the user gives a feedback to the MT system as discussed in [8], and [9]. An interactive system that takes the feedback of the user to improve the translation quality is developed in our MT system.

The scheme of this paper is as follows. We present related work in section 2, our proposed method in section 3, discuss bilingual corpus in section 4, compare the work that we have done with that done by CRULP in section 5, after which conclusion and future work is provided in section 6.

II. RELATED WORK

EBMT has been attempted for many languages, including English, French, Spanish, German, Japanese, Chinese, Turkish, Arabic, Indian, and even Sign language, as for instance in [10] – [16].

An MT system from English to Urdu is recently developed by CRULP that is available online [17]. This MT system uses rule based technique, i.e. this MT system uses syntactic parsing trees, the parts of speech tagger, and the grammatical rules for the translation purpose.

Work has been done on a web-based interactive MT system [18]. The system is a chat-style MT. The idea is to provide a broad coverage machine translation using the user's response to improve the results inspired by [8], and [9].

To the best of our knowledge, no work is available in literature for English to Urdu MT using example based approach. Also, interactive features are not available for an MT system from English to Urdu. This has motivated us to build an interactive EBMT system for the translation from English to Urdu. Also, we aim to overcome the discrepancies that are present in the CRULP MT system. The proposed method is discussed in the next section.

III. PROPOSED METHOD

The proposed method utilizes the phrase based set of examples accompanied by interactivity with the user. The methodology is divided into the following main phases.

- Sentence Fragmentation
- Searching in Corpus
- N-ary Product based Retrieval
- Ordering of Translated Text

A. Sentence Fragmentation

Fragmentation of sentences into phrases is important to improve the scope of the sentences that the translator can handle. This result can be obtained alternatively by keeping sentences in corpus and by gaining a broad coverage by fragmentation and combination to get new sentences using the genetic algorithm at run time [19]. The problem of fragmenting a sentence into simpler sentences and phrases is handled using the concept of idioms, the connecting words and the cutter points that are explained as follows.

1) Idioms

An idiom is a phrase whose meaning is not determined word for word, but is instead related to the common use of the native speakers, for details consult any standard text in English grammar, for instance [20]. A list of idioms is developed to handle idiomatic phrases. The idioms are picked from [20].

2) Connecting Words

Connecting words are the words that separate two or more sentences that are present in a large sentence e.g. and, because etc available at [21].

3) Cutter Points

Cutter points are the words that separate different meaningful phrases in a sentence. These words are auxiliary verbs, possessive pronouns, and prepositions.

FRAGMENTATION ALGORITHM

Input: A sentence in English Output: A set of English phrases Algorithm body: For each idiom, in the list of Idioms Find idiom in the input sentence If idiom is found Tag it as an idiom If ends For loop ends Find connecting words For each connecting word Separate the fragments of sentences into new sentences

For loop ends

Find cutter points

For each cutter point Separate the fragments of sentences into phrases For loop ends

EXAMPLE1:

I cannot back out of my promise.

This sentence contains an idiom 'back out of'; therefore this is first tagged as an idiom, and the word 'of' in the idiom is not considered as a cutter point. Thus, the sentence is fragmented into the following phrases.

- I
- Cannot back out of
 - My promise

EXAMPLE2:

I love to play football because it thrills.

This sentence is first decomposed into two sentences namely,

- I love to play football
- Because it thrills

The first sub-sentence is further divided into phrases using cutter points as follows.

- I love
- To play football

Its fragmented phrases are:

- I love
- To play football
- Because it thrills

B. Source Text

The source text in English is input and is fragmented into phrases using the fragmentation algorithm.

C. Searching in Corpus

Searching comprises of the idea of finding whether or not an input phrase is available in the bilingual corpus, discussed in section IV. If it doesn't find the exact match, it attempts to find the close match. The measure of closeness is done using the threshold at two levels; one for the exact match, the other for the close match.

This is done in two ways using Levenshtein Algorithm, and semantic distance algorithm. The user is given the option to choose from the both approaches.

LEVENSHTEIN ALGORITHM

This is a character based algorithm that compares an input string with the target string [22]. We have used the concept of Levenshtein algorithm for the word based comparison of two phrases or two sentences as follows.

 TABLE I.
 DESCRIPTION OF USING LEVENSHTEIN ALGORITHM

Maryam Zafar et al / International Journal on Computer Science and Engineering Vol.1(3), 2009, 275-282

		Is	Eating
	0	1	2
Was	1	1	2
Eating	2	2	1

1 in the bottom right corner indicates that one operation is needed to convert the source text 'is eating' into target text 'was eating'.

The 0 Levenshtein distance means the exact match. A positive finite Levenshtein distance means that it requires some finite operations to get the target text from the input text. A threshold 0 < T < n/2 means a closer match between two strings. We provide interactivity for the string having closer match.

Semantic distance measures how close the two strings are semantically. Semantic distance algorithm is an open source project [23]. This algorithm returns score, a coefficient measure of semantic closeness between two strings. A score of 1 means exact match, a score 0.9 < s < 1 means a synonymous word, where closeness is towards 1. We provide interactivity for 0.8 < s < 0.98. These thresholds have been set after testing on some examples. They can be adjusted in some cases as suitable.

 TABLE II.
 TRADE-OFF BETWEEN USING LEVENSHTEIN ALGORITHM AND SEMANTIC DISTANCE ALGORITHM

	Processing	Coverage
Levenshtein	Fast	Low
Semantic	Slow	High



Figure 1. Screenshot for interactive searching of phrases

We calculate the number of phrases for which a match is found divided by the total number of phrases in the input sentence, if the ratio is greater than or equal to 0.75, the program attempts to find the meanings of all the words in the remaining phrases from the dictionary. If the meanings of all the words in a phrase are found, the phrase is added to the corpus with its Urdu equivalent, otherwise, the phrase is returned to the user with whatever portion of the phrase is found in the dictionary accompanied by the word whose meanings are not found.

SEARCHING ALGORITHM Input: English Phrase **Output: OK or SENTENCE IS NOT SUPPORTED** Algorithm body: For each phrase Compare the phrase with the set of phrases in the corpus If exact match found Return OK Else Show options similar to the phrase taken from Corpus If user selects an option Return OK Else Return NOT FOUND If ends If ends For loop ends If total number of OK responses divided by total phrases >= 0.75For each word of the phrase Find meaning of the word from the dictionary If all found Add the phrase to the corpus Return OK Else Return the found words' meanings or words If ends For loop ends Else **Return SENTENCE IS NOT SUPPORTED** If ends The tool returns a string "SENTENCE IS NOT SUPPORTED" if the ratio of successful matches < 0.75.

D. N-ary Product Based Retrieval

This phase comprises of the steps used to retrieve the translation of the source text. There is a possibility of many translations available for an input sentence. We gather the possibilities and we use the idea of n-ary product to list all the possible sentences. The idea is discussed as follows.

RETRIEVAL ALGORITHM

Input: Set of all English phrases of the input sentence Output: Set of all possible Urdu translations Algorithm body: Set i=1 //i is a counter of the number of phrases For each English phrase e_i Find the phrase e_i in the corpus

🔜 Translator

Make a set U_i Find all the translations of e_i and add them

to U_i

Add 1 to i For loop ends Find the n-ary product of all U_i 's

The problem is that the number of phrases for a particular input sentence is not known before program's execution and it becomes available only at runtime. Also, the number of translation equivalents in Urdu for a particular phrase gets available only at execution time. This poses a great problem in implementation.

We resolve it by considering a set E of English phrases e_1, e_2, \ldots, e_n and corresponding to each phrase e_i of English, a set of its Urdu equivalents is considered as follows.

 $E = \text{Set of English phrases} = \{e_1, e_2, \dots, e_n\}$ $U_1 = \text{Set of Urdu translations of } e_1 = \{u_{11}, u_{12}, \dots, u_{1n_1}\}$ $U_2 = \text{Set of Urdu translations of } e_2 = \{u_{21}, u_{22}, \dots, u_{2n_2}\}$ \vdots $U_4 = \text{Set of Urdu translations of } e_1 = \{u_{21}, u_{22}, \dots, u_{2n_2}\}$

$$U_n$$
 = Set of Urdu translations of $e_n = \{\mathbf{u}_{n1}, \mathbf{u}_{n2}, \dots, \mathbf{U}_{nn_n}\}$

U = n-ary product of all the sets of Urdu translations of allEnglish phrases = $U_1 \times U_2 \times ... \times U_n$ = { $(u_{11}, u_{21}, ..., u_{n1}), ..., (u_{1n_1}, u_{2n_2}, ..., u_{nn_n})$ }

The total number of Urdu translations is given by $n_1 * n_2 * ... * n_n$, where n_i is the cardinality of U_i .

EXAMPLE 3:

A student, 'aik talib ilm'

Has been studying, 'parrh raha hai', 'parrh rahi hai',

For two hours, 'do ghante se'.

E= {e1=A student, e2=has been playing, e3=for two hours}

 $U_1 = \{u_{11} = \text{aik talib e ilm}\}$

 $U_2 = \{u_{21} = \text{parrh raha hai}, u_{22} = \text{parrh rahi hai}\}$

 $U_3 = \{u_{31} = \text{do ghante se}\}$

 $U = U_1 * U_2 * U_3 = \{(u_{11}, u_{21}, u_{31}), (u_{11}, u_{22}, u_{31})\} = \{(aik talib e ilm, parrh raha hai, do ghante se), (aik talib e ilm, parrh rahi hai, do ghante se)\}$

I am writing to tell	you about myself	
1	Translate	
	-	Search by Levenstein Search by Semantic Distance
Urdu Sentence O	uput	•••••••••••••••••••••••••••••••••••••••

- 🗆 ×

Figure 2. Screenshot for interactive retrieval phase

We have two triplets (3-tuples) in this case, i.e. we $n_1*n_2*n_3 = 1*2*1 = 2$ triplets (3-tuples).

For details of n-ary product, consult any standard text in Discrete Mathematics; see for instance [24].

E. Ordering of Translated Phrases

This one aspect is usually ignored in the research articles. However, the problem is important in the translation from English to Urdu and other distant language pairs.

The ordering algorithm is as under:

1) Ordering Rule 1

If there are $e_1, e_2, ..., e_n$ are the phrases of English whose Urdu equivalents are $u_1, u_2, ..., u_n$, then the translation of the sentence $[e_1 e_2 ... e_n]$ is given by $[u_1 u_n u_{n-1} ... u_3 u_2]$.

EXAMPLE 4:

She will work hard to pass the exam.

- She
- Will work hard
- To pass the exam

'Woh imtihan mein kamiabi ke liye mehnat kare gi'

- Woh
- Mehnat kare gi
- Imtihan mein kamiabi ke liye

By applying the rule above we get

Woh imtihan me kamiabi ke liye mehnat kare gi

2) Ordering Rule 2

For complex sentences, it is observed that the pattern repeats when some English sentences are combined to form a larger sentence using connecting words.

EXAMPLE 5:

My aunt is a doctor and my uncle works in a factory that makes television sets.

In this sentence, 'and', and 'that' are connecting words that split the sentence into three sub-sentences as under.

- My aunt is a doctor
- And my uncle works in a factory
- That makes television sets

These sentences are decomposed into the following way

- My aunt, is a doctor
- And my uncle works, in a factory
- That makes television sets

Using the rule stated above the translation is given as

Meri aunti doctor hain, aur mere uncle kaam karte hain factory mein, jo television set banati hai.

For more complex sentences in the sense of ordering of their translations, we provide an interactive tree structure that a user can utilize to adjust the phrases to get a correct order of Urdu translations.



Figure 3. Screenshot for interactive ordering of translated text

IV. BILINGUAL CORPUS

Bilingual corpus is one of the most important parts of this tool, since the tool relies mainly on the corpus and a set of rules to combine the phrases in the corpus.

For the construction of the corpus, we take the textbooks of the secondary school [20], [25], and [26], pick random sentences, and then we fragment them based on the idea of connecting words, and the cutter points, details of which are discussed in section III A. Corpus is built using a phrase-based set of examples, rather than sentence based examples to incorporate more sentences and scenarios. A list of idioms is separately developed to supplement the bilingual corpus, as discussed in III A 1). Also, the homographs, are handled inspired by [27] as in example 6, these are the words that have same spelling but different meanings. Gender differences are also incorporated for the phrases of neutral gender as in example 7, and the singularity differences are handled for the phrases of neutral singularity as in example 8.

EXAMPLE 6:

- He gets an apple, 'use aik saib mila',
- He gets an idea, 'use aik tarkeeb soojhi',
- He gets to home, 'woh ghar ko pahuncha'.

In these sentences, 'gets' is the homograph that gives different meanings in different contexts.

EXAMPLE 7:

A player is playing, 'aik khilarri khel raha hai', 'aik khilarri khel rahi hai'. In this example, 'player' is the neutral gender word.

EXAMPLE 8:

You are eating; 'aap kha rahe hain', 'tum kha rahe ho' etc. In this example, 'you' is the neutral singular word.

Since the homographs are included in the corpus, and also the gender differences are recognized, therefore, the corpus size is not homogeneous, i.e. the column size is variable because one English phrase may have many Urdu equivalents as in example 9.

EXAMPLE 9:

Enjoys cricket; 'cricket pasand karta hai', 'cricket pasand karti hai'.

In this example, enjoys cricket is stored in one row, and the corresponding equivalents are placed in the same row in different columns.

There is another interesting feature of the tool associated with the corpus construction. There is a separate module that ensures that no phrase is duplicated in the corpus, although one phrase can have more than one translation equivalents as has been discussed above. However, in case a phrase occurs more than once in the construction phase, its frequency is added thereby improving its probability in the sense of Bayesian probability theory [28]. This idea is useful when there are scarce computing resources at our disposal as in the case of embedded devices.

A. Dictionary

A dictionary can be used in an EBMT approach to support translation process as in [29] The idea is to let the program learn the new words from the dictionary whenever the program fails to translate a sentence by a small fraction of the unrecognized phrases to the total number of phrases in the sentence. We build a specific dictionary whose words are taken from [30]. This helps in evolving the bilingual corpus at runtime.

V. FEATURES ANALYSIS OF OUR MT SYSTEM COMPARED WITH CRULP MT SYSTEM

CRULP system is a good attempt towards English to Urdu MT and it facilitates the correct translations in many situations. However, there are some discrepancies that we have attempted to overcome in our system.

Examples illustrating idioms:

EXAMPLE 10:

Input:

The conspiracy was brought to light by policeman

CRULP MT system response:

'saazish police ke afsar ke paas roshni ki taraf layi gayi'

Our MT system response:

'saazish police afsar se manzar e aam par aayi'

EXAMPLE 11:

Input:

He has come of age today

CRULP MT system response:

'woh aaj umer ka aaya hai'

Our MT system response:

'woh aaj baaligh hua hai'

In these examples, the idioms 'brought to light', and 'come of age' have been considered for which CRULP MT system attempts to translate word for word, whereas, our MT system attempts the idiomatic phrase correctly.

Examples illustrating homographs:

EXAMPLE 12:

Input:

- He gets an apple
- He gets an idea

CRULP MT system response:

'use saib milta hai'

'use khayal milta hai'

Our MT system response (as options):

Translation options for first sentence:

'use mila ailk saib'

'use soojha aik saib'

Translation options for second sentence:

'use soojha aik khayal'

'use mila aik khayal'

EXAMPLE 13:

Input:

- He works in a bank
- He is waiting near the bank of a river

CRULP MT system response:

'woh bank mein kaam karta hai'

'daria ke bank ke qareeb he is waiting'

Our MT system response (as options):

Translation options for first sentence:

'Woh kaam karta hai bank mein'

'Woh kaam karta hai kinare mein'

Translation options for second sentence:

'Woh dariya ke kinare ke qareeb intizaar kar raha hai'

'Woh daria ke bank ke qareeb intizaar kar raha hai'

In example 12, the word 'get' is taken in two senses, i.e. to come into possession of, and to perceive. In example 13, the word 'bank' has two meanings i.e. bank as a financial institution, and bank as the slope beside a body of water. CRULP MT system doesn't support the multiple uses of the word bank and get in these examples, but our MT system provides such support.

Examples illustrating the gender and the words taken in singular and plural sense together:

EXAMPLE 14:

Input:

They are playing in the garden CRULP MT system response: 'woh baagh mein khel rahe hain' Our MT system response: 'woh baagh mein khel rahe hain' 'who baagh mein khel rahi hain'

EXAMPLE 15:

Input: It is his work CRULP MT system response: 'yeh uss ka kaam hai' Our MT system response: 'yeh uss ka kaam hai' 'yeh unn ka kaam hai' VII.

In example 14, the word 'they' is taken in the masculine sense as well as feminine sense. In example 15, the word 'his' is taken in two senses in Urdu, i.e. when we want to give regard some third person, and when no such regard is needed. In these examples, the CRULP MT system does not support the words having the two gender senses, and it also does not discriminate between the word senses for the regard, which is usual in Urdu.

	CRULP MT system	Proposed system	
Idioms	Not supported	Supported	
Homographs	Not supported	Supported	
Gender	Not supported	Give options	
Processing time	High	Low	
Up gradation	Not supported	Upgrades its corpus	

TABLE 3: COMPARISON OF CRULP MT SYTEM WITH OUR MT SYSTEM

Our MT system fails to provide a correct ordering of the translated text when there are connecting words that also lie in the category of cutter points, e.g. 'and', 'or' etc, or when a conjunction is simply a separator in the list of some nouns or verbs. This also doesn't handle the ordering of the translated text correctly when there are phrases like 'Government of Pakistan', in which the word 'of' is not to be considered as a cutter point. Such ordering problems are ignored in the automated translations, as for instance discussed in [23]. However, we provide support to handle such issues as discussed in [ordering of translated text].

VI. CONCLUSION AND FUTURE WORK

We have developed an interactive MT system, thereby providing support for idioms, homographs, gender, the words having plural and singular senses together, ability of the corpus to evolve for a broader coverage, and also the support for the problem of ordering of the translated text. Ordering is a problem that is hard for computer but easier for humans. The interactive system brings ease to the users thereby complying with the basic goal of the research in this direction, i.e. facilitating the user in the translation process and improving the efficiency of the task.

Our MT system works particularly well for the situations that it has been trained to handle. This suggests us that an MT system for domain specific needs can be built using the ideas discussed here. Also, the size of our corpus suggests that it can be used in embedded devices with low memory resources. The phrase based set of examples are used in this MT system, this gives it a broader coverage.

Work is in progress to extend the bilingual corpus, and to improve the algorithm for the ordering of the translated text. The extended corpus and the improved ordering algorithm improve the scope of our MT system, enhance the quality of the translation, and optimize the efficiency of the translation process.

REFERENCES

- [1] Tafseer Ahmed Sadaf Alvi, English to Urdu translation system, University of Karachi, Pakistan, 2002.
- [2] Sarmad HUSSAIN "Urdu Localization Project: Lexicon, MT and TTS (ULP)" Center for Research in Urdu Language Processing, National University of Computer and Emerging Sciences.
- [3] W. John Hutchins and Harold L. Somers, "An Introduction to Machine Translation". A CADEMIC PRESS. (London),1992.
- [4] Cyfieithu Peirianyddol a'r Gymraeg: y Ffordd Ymlaen, "Machine translation and welsh: the way forward", A Report for The Welsh Language Board, July 2004.
- [5] <u>http://en.wikipedia.org/wiki/Statistical_machine_translation</u>
- [6] M. Nagao, "A framework of a mechanical translation between Japanese and English by analogy principle," in Proceedings of the international NATO symposium on Artificial and human intelligence, pp. 173– 180,1984.
- [7] Chris Brockett, Takako Aikawa, Anthony Aue, Arul Menezes, Chris Quirk, & Hisami Suzuki, "English-Japanese example-based machine translation using abstract linguistic representation", Coling-2002 workshop "Machine translation in Asia", Taipei, Taiwan, 1 September 2002.
- [8] Mark Seligman. 1997. Six issues in speech trans- lation. In Steven Kra.uwer et al., editors, Spo- ken Language Translation Workshop, pages 83--89, Madrid, July.
- [9] Robert Frederking, Alexander Rudnicky, Er Rudnicky, Christopher Hogan, "Interactive speech translation in the DIPLOMAT project", In Proceedings of the Spoken Language Translation Kluwer Academic Publishers Hingham, MA, USA, pp 27-42, 2000.
- [10] Yang, Muyun; Jiang, Hongfei; Tiejun, Zhao; Li, Sheng; Liu, Daxin, "Domain Sensitive Chinese-English Example Based Machine Translation", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08, Volume 2, Page(s):453 - 456 18- 20 Oct 2008.
- [11] Morrissey S. and A. Way. "An Example-Based Approach to Translating Sign Language", In Proceedings of the Second Workshop on Example-Based Machine Translation, Phuket, Thailand, pp.109—116, 2005.
- [12] Alp, N. D.; Turhan, C, "English to Turkish Example-Based Machine Translation with Synchronous SSTC", Fifth International Conference on Information Technology: New Generations, Page(s):674 – 679, 7-9 April 2008.
- [13] Ralf D. Brown, "Example-Based Machine Translation in the Pangloss System", In Proceedings of the 16th International Conference on Computational Linguistics (COLING-96), p. 169-174. Copenhagen, Denmark, August 5-9, 1996.
- [14] Nicolas Stroppa, Declan Groves, Andy Way and Kepa Sarasola. "Example-Based Machine Translation of the Basque Language".In Proceedings of the 7th biennial conference of the Association for Machine Translation in the Americas, pages 232-241, Cambridge, Massachusetts, August 2006.
- [15] Vamshi Ambati & Rohini U: "A hybrid approach to example based machine translation for Indian languages". ICON-2007: 5th International Conference on Natural Language Processing, IIIT Hyderabad, India, pp 4-6, January 2007.
- [16] Kfir Bar, Y.Choueka, & N.Dershowitz: "An Arabic to English examplebased translation system". ICTIS 2007: Information and Communication Technologies International Symposium. Workshop on Arabic natural language processing, pp. 355-359, 3-5 April 2007.
- [17] Center of Research for Urdu Language Processing (CRULP), http://www.crulp.org (Last accesss 10/07/2009).
- [18] Christopher Hogan, and Robert Frederking, "WebDIPLOMAT: A Web-Based Interactive Machine Translation System", Proceedings of the 18th conference on Computational linguistics - Volume 2. PP 1041 – 1045, 2000.
- [19] Hiroshi Echizen-ya, Kenji Araki, Yoshio Momouchi, Koji Tochinai, " Machine translation method using inductive learning with genetic algorithms" Proceedings of the 16th conference on Computational linguistics, August 05-09, 1996, Copenhagen, Denmark.

Maryam Zafar et al / International Journal on Computer Science and Engineering Vol.1(3), 2009, 275-282

- [20] Afzal Anwar mufti, "English Grammer and Composition", Qaumi Kutub Khana, Lahore, Pakistan.
- ", Qaumi Kutub [26] John Jackman, and Wendy Wren, "Book 2", Sunrise/ Nelson publishers. [27] Michael McCarthy, Felicity O'Dell, English Vocabulary in Us
- [21] <u>http://daphne.palomar.edu/mlloyd-</u> jones/handouts/CONNECTING%20WORDS.pdf
- [22] http://en.wikipedia.org/wiki/Levenshtein_distance
- [23] Troy Simpson, Thanh Dao; WordNet-based semantic similarity measurement, <u>http://www.codeproject.com/KB/string/semanticsimilaritywordnet.aspx</u>
- [24] James L Hein, "Discrete Mathematics", Jones and Bartlett books in mathematics.
- [25] John Jackman , and Wendy Wren, "Book 5", Sunrise/ Nelson publishers.
- [27] Michael McCarthy, Felicity O'Dell, English Vocabulary in Use Advanced, Cambridge University Press.
- [28] Stuart Russell, and Peter Norvig, "Artificial intelligence: A modern approach" Second Edition. Prentice Hall series in Artificial Intelligence.
- [29] Sadao Kurohashi, Toshiaki Nakazawa, Kauffmann Alexis, Daisuke Kawahara: "Example-based Machine Translation Pursuing Fully Structural NLP", In Proceedings of International Workshop on Spoken Language Translation (IWSLT'05), pp.207-212, Pittsburgh, Oct 2005.
- [30] Apni Urdu, http://www.apniurdu.com (Last accesss25/06/2009).