Anuvadaksh: An Evaluation

ANUVADAKSH : AN EVALUATION

Aphale Neha, Doctor R, Shinde Babita
Centre for Development of Advanced Computing, Pune
{nehaa, rdoctor, babitag}@cdac.in

Abstract:

In order to ensure seamless transfer of Information across languages, TDIL DeitY has initiated various research areas viz. development of machine translation, cross lingual information access system, optical character recognition. This paper presents an overview of the Procedures as well as Methodologies deployed for evaluating one of the many MT Systems: Anuvadaksh. CDAC GIST has been entrusted with the task of developing an evaluation strategy and evaluating the output of the Translations. The paper is divided into 4 parts.
To set things in perspective, Part One of the paper presents a broad overview of the various machine translation systems, all the more so, since the evaluation methodology is common for all the machine translation systems. An overview of Anuvadaksh system along with other existing MT systems is presented.
The Objectives underlying evaluation are given in Part two.
Part three deals with the development and fine-tuning of the machine translation evaluation strategy and the details of evaluation procedure followed.
As a logical corollary, Part four gives a brief analysis of Anuvadaksh system. The next part provides steps for improving the overall translation quality.
It needs to be noted that, this paper deals only with machine translation evaluation strategy & evaluation procedure as pertinent to Anuvadaksh and hence the output of the system in terms of accuracy is not presented.

1. Introduction:

To build machine translation systems different institutes & organizations have been working for several years to overcome the language barriers and have generated a large repository of linguistic tools & resources, pertinent to machine translation. The Department of Information Technology (DeitY), Ministry of Communications and Information Technology, Government of India has taken the initiative and provided them a common platform in consortium mode, through which these machine translation systems are made available to public mass usage through TDIL-DC portal. (http://translation.tdil-dc.in/). Anuvadaksh MT systems have completed One Phase of development, testing and evaluation (31/12/2007 to 30/6/2010), and are now in their second phase.

India is a multilingual country. There is a great demand for translation of documents from one language to another. This will ensure larger flow of information across different languages.
Anuvadaksh: an English to Indian language machine translation system is part of the machine translation consortia the other three being, AnglaMT, ILMT & Sanskrit to Hindi Translation. Anuvadaksh aims to develop and deploy a machine translation system from English to chosen Indian languages in Tourism and Health domains and supports the following languages.

1. English →Hindi
2. English →Marathi
3. English →Bangla
4. English →Urdu
5. English →Tamil
6. English →Odia
7. English →Gujarati (newly added in phase-II)

8. English →Bodo (newly added in phase-II)

Anuvadaksh is developed by EILMT consortium. C-DAC Pune is a Consortium Leader in association with 12 institutes - IIT- Hyderabad, C-DAC Mumbai, IIT- Bombay, Jadavpur - University, IIIT- Allahabad, Utkal University - Bhubaneswar, Amrita- university , Banasthali Vidyapeeth, North Maharastra University, Jalgaon, Dharamsinh Desai University, Nadiad and North Eastern Hill University, Shillong. This is a multi engine; multi output machine translation system which has TAG based MT (TAG), Example Based MT (EBMT) & Statistical Machine Translation (SMT) systems. Whereas the TAG system supports all the 8 indian languages, SMT supports three languages Hindi, Marathi & Bangla and EBMT supports only Hindi.

2. Objectives of Evaluation:

The aim of the evaluation is to evolve a strategy for evaluating output of the different Translation Engines and provide a methodology at the national level for machine translation evaluation, which will focus on linguistic analysis of machine translation system as well as to provide end to end system performance in quantitative measure. There exist international standards for evaluation like BLEU, NIST & Meteor. But these do not cater to Indian language complexities. However in close consultation with consortia leaders it was felt that in the initial stage the tried and tested approach of manual testing be carried out to ensure that the translation system meets the main requirements: usability & native user expectations. Our objective has been therefore to develop in consultation with the Consortia an evaluation procedure which evaluates output of the translation system in terms of comprehensibility, fluency and linguistic integrity also, which can handle the morphological complexities of Indian languages.

3. Machine Translation Evaluation: An Overview

Once the objective is set in place, Evaluation in conformity with the Objective needs to be defined. Evaluation is necessarily a two pronged process. Creating a Strategy: A strategy for evaluation needs to be defined in terms of what is being targeted. Implementing the Strategy: Once this strategy is formulated and finalized, a method for implementing the strategy needs to be developed. Each of these will be treated in what follows:

3.1 Development of a Strategy:

The Evaluation strategy evolved has undergone mutations in the various stages of its development, as and when successive evaluations have taken place and feedback from the consortia team has been received. In what follows a diachronic development of this strategy is presented. In short the strategy has moved from Linguistic Testing to Sprachgefühl i.e. native speaker's acceptance of the output in terms of comprehensibility and fluency.

3.1.1 Phase 1:Linguistic testing[1]

Initially the strategy proposed was that of Linguistic Testing through a test bed of patterns based on Spelling, Morphology, Syntax, Semantics, lexicology and norms. Stress was also laid on whether the systems produced text which was in compliance with the Akshara theory. A short synopsis of the major parameters of this approach is given below: .

- **Orthography:**
-

1. Misspellings: Spelling correction: e.g. strenght for strength, dias for dais
2. Spelling variants: American vs. British: e.g. Organise vs. Organize

1 All examples are provided in roman for intelligibility

3. Abbreviations and Truncations: e.g. it's, 'tis I'd ain't
4. Acronyms: e.g. IT: Income Tax or Information Technology
5. Misused terms: e.g. loose for lose

- ***Morphology:***

1. Moving from a low level morphological language to a high level one e.g. ENGLISH to TAMIL will test the translation systems ability to lemmatize data as well as the POS Tagging and parsing "skills" of the MT system.
2. Morphological variants: E.g. Urdu: joining the helper verb or keeping the helper verb distinct kiyaagayaa written together instead of separately.

- ***Morphotactics:***

1. Suffix ordering : EN->MAR From this very table: tebl+aa+var+caa+c in Marathi EN->GUJ From the boy's side (with stress): chokr+aa+vaaLaa+o+maaN+thi+ya En->MAL maram+il -> marattil (geminated sandhi)
2. Sandhi: e.g. Vowel harmony in Bangla verbs e.g. khaa vs. khe

- ***Syntax:***

1. Handling of word order: SVO->SOV.

2. Use of correct negation: mat and nahiN in Urdu.

3. Handling dummy constructs: e.g. DO in English "I DO not understand", "DOES he know?" vs.. He DID the work.

4. Anaphora and cataphora interpretation from Source to Target Language.

    a. E.g. It's well known that Rajasthan has forts. The dog came. Its tail was wagging.

5. IF constructs (conditionals) If I were..., If I win..., If he went..., If he had gone....

6. POS typologies and their translations: Pronouns, Adjectives (Non-qualifiers), Adverbs

7. Negation: He has NOT come, he does NOT know

- ***Semantics:***

1. Here stress will be basically on Semantic competence. Some test cases are provided below.

2. Semantic Ambiguity:
    a. E.g. American head seeks arms Identification and explanation: The homograph "head" can be interpreted as a noun meaning either chief or the anatomical head of a body.
3. Lexical Ambiguity:
    a. E.g. Teacher strikes idle kids.
    b. Ambiguity type: Lexical (part of speech or category ambiguity). Identification and explanation: "strikes" can occur as either a verb meaning to hit or a noun meaning a refusal to work.
4. Structural Ambiguity:
    a. E.g. stolen painting found by tree Ambiguity type: Structural.

5. A tree found a stolen painting. A person found a stolen painting near a tree. (* Indicates an unacceptable sentence)

- ***Lexicology:***

1. Borrowing: A word taken directly from another language e.g. Strawberry, Hadran Collider in Target languages Loan vs. Source translation: Installments vs. kishton
2. Calque: SL foreign word/phrase is translated and incorporated into TL free verse
3. Transliteration & Named Entity Recognition: Taj Mahal, Eiffel Tower Translation of "function words" such as for: Rajasthan is famous for its castles.
4. Translation of Phrasal verbs: go in for, go for, go into,
5. Divalence of POS: yellow (Adj., Noun, Verb)

- ***Norms:***

1. Spelling Norms: Compliance with spelling norms of the respective target languages:
   i. e.g. Urdu: Imlaanaamaa Bangla: Bangla Akademi
2. Storage Norms: Compliance with Unicode Based on the above parameters the following 7-POINT Russian Grading scale (slightly modified) which deals with translation as a process of visibility of text was proposed. Usability and Transmission of Information is the prime criteria on which this grading scale was based on.

7-POINT Grading scale

1.	**Opacity**	The rendering is absolutely useless for any purpose. Such a rendering shall be deemed as of **UNACCEPTABLE** quality.
2.	**Semi-Opaque**	Some parts are comprehensible, but on the whole the picture still remains difficult to get and the text evades the target user. Such a rendering shall be deemed as **POOR** quality.
3.	**Part Visibility**	The user can get a grasp of the over-all intention of the text, but on the whole, the user has to work hard to get at the meaning of the text and large fragments are practically opaque and incomprehensible. Such a rendering shall be deemed as **LOW** quality.
4.	**Half Visibility**	The rendered text is quite comprehensible to the target reader and can be used by him/her as can be used as a rough draft for improvement. Such a rendering shall be deemed as **DRAFT** quality.
5.	**Near Visibility**	Text is clear enough and all pertinent information can be drawn from it. However, the text is hard to read due to language errors and require further filtration. Such a rendering shall be deemed as of **ACCEPTABLE** quality.
6.	**Near-total visibility**	The rendering has stylistic errors and also some difficult grammatical, syntactic, lexical issues are not clarified. However, it transmits the information needed to the target user. Such a rendering shall be deemed as of **SATISFACTORY** quality.
7.	**Total visibility**	The rendering passes muster though not stylistically perfect. Such a rendering shall be deemed as of **HIGH** quality.

However, as the Anuvadaksh machine translation system is a research project it was felt that these parameters could be applied at a later stage. Moreover it was observed that the more the number of scales, the more will be errors in human judgment and more training is required for human evaluators. Hence this strategy based on linguistic testing was replaced by one based on the native speaker's evaluation.

3.1.2 Phase 2: Comprehensibility

A new approach was formulated with stress on Sprachgefühl i.e. focus on usability and the native speaker's expectations and the translation quality was provided in terms of comprehensibility of output. With this in view, the following 4 point grading scale was finalized.

Grade 0	No output provided by the engine concerned.
Grade 1	The translated output is not comprehensible.
Grade 2	Comprehensible after accessing the source text.
Grade 3	Comprehensible with difficulty.
Grade 4	Acceptable since the text is comprehensible.

For calculating the ratings, the following formula was deployed:

$$\text{Final Evaluation Mean For Each Sentence (Within 0 - 4)} = \frac{\sum_{T=1}^{T=3} \text{Grade Given By Tester T}}{3}$$

$$\text{Total Average} = \sum_{S=1}^{S=100} \text{Final Evaluation Mean Of Sentence S}$$

$$\text{Accuracy Of Engine (In \%)} = \frac{\text{Total Average}}{\text{Number Of Sentences}} * 25$$

3.1.3 Phase 3: Present Evaluation Strategy: Fluency And Comprehensibility

Based on the learning experience from evaluation and suggestions from consortia the grading scale was changed. And this because of the following reason: There was confusion in the above grading scale, where "Grade 2" -"Comprehensible after accessing the source text" was useful only in Open testing and had no significance in Blind testing. This made a difference in results of MT systems. So in consultation with all the MT consortia leaders it was decided to change the grading scale for testing of MT systems when Phase-II was deployed.

Also it was decided that, NO OUTPUT by system and buffer clearance issue will be graded as "-1". As per previous grading scale this was graded as "0" which largely affected the performance of the system. Performance of the systems will be given on two parameters, (a) comprehensibility and (b) fluency.

This led to the following rating system:

Grade 0	Nonsense (if the sentence doesn't make any sense at all - it is like someone speaking to out in al language you don't know)
Grade 1	Some parts make sense but is not comprehensible over all (e.g., listening to a language which has lots of borrowed words from your language - you understand those words but nothing more)
Grade 2	Comprehensible but has quite a few errors (e.g., someone who can speak your language but would make lots of errors. However, you can make sense out of what is being said)
Grade 3	Comprehensible, occasional errors (e.g., someone speaking Hindi getting all its genders wrong)
Grade 4	Perfect (e.g., someone who knows the language)

The results are thus calculated on two parameters, (a) comprehensibility and (b) fluency. Both (a) and (b) are be calculated by considering the average of the score given by all the evaluators for every sentence.

(a) Comprehensibility is calculated by taking out the percentage of the number of sentences getting an average score bet 2-4 out of the total number of sentences in the set. Specifically, let S_i be the number of sentences with a grade of i (i=0, 1, 2, 3, 4). Then comprehensibility C = sum_over_{i=2,3,4}(S_i)/ sum_over_all_i(S_i).

(b) For fluency, the average scores will be measured against [4 * total number of sentences in the set]. Specifically, let S_i be the number of sentences with a grade of i (i=0, 1, 2, 3, 4).
Then fluency Q = sum_over_all_i(i *S_i)/[4 * sum_over_all_i(S_i)].

3.2 Development of an Evaluation Procedure:

Once these testing parameters were set in place, the following step-by-step procedure was adopted. The main features of the evaluation procedure are laid down point wise for the purpose of clarity.

3.2.1 Sampling of Test Data:

Test data for both tourism & health domain was provided by consortia. Out of the provided test data, randomly selected sentences were used for evaluation. For the initial round of testing 100 sentences were used and later number of test sentences were increased to 1000 for extensive testing. All these sentences were randomly selected from the test data provided by consortia. Phase-II evaluation of Anuvadaksh system is carried out on the 100 sentences for tourism & health domain. All these sentences were categorized grammatically structure wise by consortia.

3.2.2 Identification of Evaluators:

Once the data was finalized, the choice of evaluators was determined. This was conditioned by the strategy. As user expectations is a main criteria of machine translation evaluation, instead of linguists 3 to 5 native users were used for evaluation. Since all evaluation of a translation is subjective; to bring in a modicum of objectivity, 3 -5 evaluators were used. Evaluators were selected on the basis of their proficiency in the source and target language. To ensure a broad sampling spectrum, they were chosen from different professions and age-groups.

Thus for English →Marathi, the following sampling of professionals was chosen:

English →Marathi					
Evaluator ID	Eval 1	Eval 2	Eval 3	Eval 4	Eval 5
Qualification	B.Ed.	B.E.	M.A.	M.A.	B.A.
Profession	Teacher	I.T.Professional	Freelancer	student	Freelancer

3.2.3 Training To Evaluators:

Once the evaluators were determined, a workshop for formal training of the evaluators was conducted to make them aware about the grading scale, along with the sample examples of each grading scale. This was followed by a Hands-on Training session to make them familiar with the procedure and also to resolve all doubts. The evaluation was a double-bind test: Open vs. Blind testing. A double-bind testing was proposed i.e. Blind vs. Open testing. Initially the evaluators were supposed to evaluate the output alone in terms of the grading scale and in the second half, propose a grading after seeing both source and target. The grading may change in Open Testing for following two reasons

a. Some translations deemed as incomprehensible will become “comprehensible with great difficulty” i.e. after accessing the source text

b. However it may also be that the translation deemed correct will be proved incorrect, since a wrong choice of word is involved.

E.g. Translation for: I saw the man may be rendered in output as: मैं को आदमी काटा which may be judged as partly comprehensible when in fact it has distorted the meaning of the English and the rating will be changed to Incomprehensible.

3.2.4 Evaluation by Evaluators:

Each evaluator was provided two sheets. One sheet for Blind Testing and one for Open testing. Each evaluator first undertook blind testing and subsequently Open Testing. Along with the testing sheets, an instruction sheet was provided to each evaluator for their guidance and to avoid the confusion in grading the sentences. In case of Blind Testing, the evaluator did not have access to Source sentence. And in Open Testing s/he was provided machine translated output along with the Source sentence. To ensure objectivity, if multiple outputs were obtained from the System, the top five outputs were considered for evaluation, from which the evaluator selected the better machine translated output for each input sentence. The evaluators were given enough time to grade the sentences. Comprehensibility & fluency are calculated on the basis of grading given by evaluators. The grading provided by the evaluators was furnished to the Consortia for their feedback and for bettering the system.

4. Analysis of Anuvadaksh:

Various rounds of evaluation have been carried out with test data provided by consortia for all the language pairs. The quality of output has improved over each round of testing. Our major observations are provided in what follows:

4.1 General Observations:

The Overall performance of Tourism domain is better than health domain. This is possibly because the initial phase was for Tourism and the Health domain was added on later within a comparative perspective, TAG & SMT has better translation quality than EBMT engine. Performance of system in terms of quality of translation is increased from Phase-I, for English to Hindi (for both TAG & SMT engine), English to Bangla (TAG) and English to Urdu (TAG).But the quality of translation is a bit less for English to Marathi (TAG), English to Odia & English to Tamil.

4.2 Syntactic Issues:

The following observations were noted for syntactic evaluation of the data: Irrespective of language, it was found that in Tourism domain, the system performs well for the following sentence structures "Simple, co-ordinate, copula, pp initial, participle, relative clause, that clause"; However, improvement for structure type "Conditional, wh-clause, adverb initial, apposition, discourse connector" needs to be implemented.

In Health domain it is observed that system perform better for the following sentence structure types “Simple, Relative clause, participial/gerund, compound, pp initial, adverb initial, copula, that clause, infinitive constructs, appositional, wh-clause" and needs improvement for the structure type "Discourse connector, imperative, complex, multiple verb, participle constructs, conditional clause".

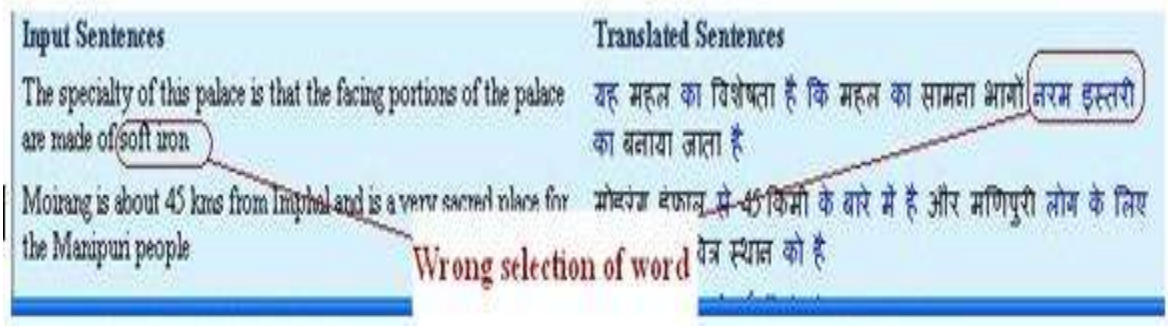
4.3 Error Diagnostics

On the basis of evaluation, we have analyzed the following most prominent errors. Quality of the translation can be considerably increased if these certain issues are addressed.

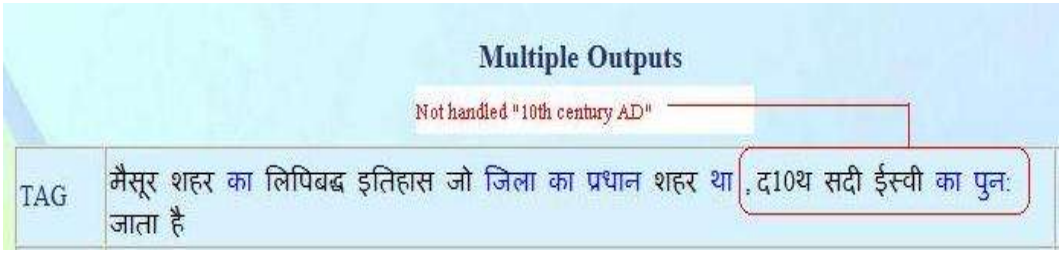
1. Ranking Module - As Anuvadaksh is multi engine translation system, if three engines (TAG, SMT & EBMT) are integrated along with the ranking module, better output can be provided to the end user.

2. WSD module - Improvement in WSD module will help to avoid problem of wrong selection of words in target language, where a word has multiple meanings.

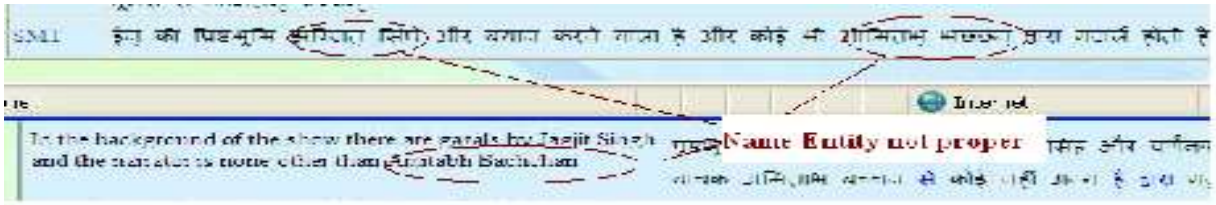
3. ***Preprocessing Module*** -Cases of Acronyms and short abbreviations need to be addressed in the preprocessing module –Thus 12th is glossed as 12थ, instead of 12वी.,



4. ***NER module*** – This is a complex area and all MAT systems encounter difficulties in correctly identifying Named Entities. Certain named entities are not identified and transliterated correctly. The quality of transliteration is quite good but if improved will provide a boost to rating.



5. Sometimes Indian language text is not rendered properly



EBMT	वह आवास् लगातार् वह् पथ् युक्त् है का प्राचीन् तच्ट् कीचद्ः झोपद्ःइयों और् वह् पथ् का जीवत्व् यद्यपि सत्कार् है सम्पूर्णतया जातीय्

5. Conclusion

The Road Ahead Evaluation in terms of both procedure and strategy and its implementation has been a huge learning experience. Fine tuning the Procedure will better the evaluation approach. From the feedback we have received from the evaluators, it was felt that within the present system more clarity is needed for differentiating Grade 2 and Grade 3. While at present test data is proposed by the consortia for testing he output, in subsequent testing grounds, the test data will be provided by the testing team itself. Since the engines are available for testing on the TDIL-DC website, a large sampling of sentences has been provided by users and its is suggested that this valuable user data be used for bettering the quality of translation and also for testing purposes. At present no module- wise analysis of the output is being tested. Reports for such and evaluation will enable the developers to pinpoint with accuracy which modules are efficient and which needs to be fine-tuned. Machine Translation is a major challenge world-wide and Anuvadaksh's contribution towards bridging the language divide between English and Indian Languages cannot be understated. Within the perspective of the evaluation carried out, we are confident that the MT System will provide near-native translation output.

6. Acknowledgement

We would like to express our special gratitude and thanks to TDIL-Deity officials for supporting and guiding this activity. We would also like to thank all the consortia leaders of machine translation systems Dr. R.M.K. Sinha, Dr. Rajiv Sangal, Dr. hemant Darbari and MR. V.N. Shukla for their valuable and constructive suggestions during the evolution of machine translation evaluation strategy, In addition, we would also like to thank chief investigator of the Testing, benchmarking & Evaluation project, Mr. Mahesh Kulkarni for his continuous support and linguistic Resource Cell at C-DAC GIST for getting the evaluators on board.

References:

1. Amengual, J.C., J.M. Benedí, F. Casacuberta, A. Castao, A. Castellanos, D. Llorens, A. Marzal, F.Prat, E. Vidal and J.M. Vilar; 1997 Using Categories in the EUTRANS System, Spoken Language Translation: Proceedings of a Workshop Sponsored by the Association for Computational Linguistics and the by the European Network in Language and Speech (ELSNET), Madrid, pp. 44-53.

2. Arnold, Doug, Louisa Sadler and R. Lee Humphreys;1993, Evaluation: An Assessment, Machine Translation 8, 1 24.

3. Bharati, Akshar, Chaitanya, Vineet, Kulkarni, Amba P., Sangal, Rajeev Anusaaraka: Machine Translation in stages . Vivek, A Quarterly in Artificial Intelligence, Vol. 10, No. 3 July 1997,NCST, India, pp. 22-25.http://arxiv.org/pdf/cs.CL/0306130

4. Blekhman Michael S. 2005 MACHINE TRANSLATION: PROFESSIONAL EXPERIENCE An introduction to an introduction.

5. Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, and Nizar Habash. 2002. DUSTer: A Method for Unraveling Cross-Language Divergences for Statistical Word Level Alignment. In Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002,Tiburon, CA.

6. Bonnie J. Dorr. 1993. Machine Translation: A View from the Lexicon. MIT Press, Cambridge, MA.

7. Carter, David, Manny Rayner, Robert Eklund, Catriona MacDermid and Mats Wirén ;2000, "Evaluation", in Manny Rayner, David Carter, Pierrette Bouillon, Vassilis Digalakis and Mats Wirén (eds) The Spoken Language Translator, Cambridge: Cambridge University Press, pp. 297–312.

8. Chung Hye Han, Benoit Lavoie, Martha Palmer, Owen Rambow, Richard Kittredge, Tanya Korelsky, Nari Kim, and Myunghee Kim. 2000. Hadling Structural Divergences and Recovering Dropped Arguments in a Korean/English Machine Translation System. In Proceedings of the Fourth Conference of the Asoociation for Machine Translation in the Americas,AMTA-2000, Cuernavaca, Mexico.

9. Dash, Niladri Sekhar, Chaudhuri, Bidyut Baran 2000. Why do we need to develop corpora in Indian languages? . A paper presented at SCALLA 2001 conference, Bangalore.http://www.elda.fr/proj/scalla/SCALLA2001/SCALLA2001Dash.pdf

10. Dave, Shachi, Parikh, Jignashu and Bhattacharyya, Pushpak Interlingua Based English Hindi Machine Translation and Language Divergence, Journal of Machine Translation, Volume 17, September, 2002.

11. Deepa Gupta and Niladri Chatterjee. 2003a. Divergence in English to Hindi Translation: Some Studies. International Journal of Translation, Bahri Publications, New Delhi. (In print).

12. Durgesh Rao. 2001. Human Aided Machine Translation from English to Hindi: The MaTra Project at NCST.In Proceedings Symposium on Translation Support Systems, STRANS-2001, I.I.T. Kanpur.

13. Gates, Donna, Alon Lavie, Lori Levin, Marsal Gavaldà,Monika Woszczyna and Puminhg Zhan; 1997, End-to-End Evaluation in JANUS: A Speech-to-Speech Translation System", in E. Maier, M. Mast and S. Luperfoy (eds) Dialogue Processing in Spoken Language Systems, Berlin: Springer, pp.195-206.

14. Hideo Watanabe, Sadao Kurohashi, and Eiji Aramaki. 2000. Finding Structural Correspondences from Bilingual Parsed Corpus for Corpus-based Translation. In Proceeding of COLING-2000, Saarbrucken, Germany.

15. Hutchins, W. John, Somers, Harold L.; 1992. An Introduction to Machine Translation. Academic Press, London,.

16. Jain, A.N., A.E. McNair, A. Waibel, H. Saito, A.G. Hauptmann and J. Tebelskis ;1993; "Connectionist and Symbolic Processing in Speech-to-Speech Translation: The JANUS System", in Sergei Nirenburg (ed.) Progress in Machine Translation, Amsterdam: IOS Press and Tokyo: Ohmsha, pp.153-160.

17. Kay, Martin, Jean Mark Gawron and Peter Norvig;1994; Verbmobil: A Translation System for Faceto-Face Dialog, Stanford, CA: CSLI.

18. Krauwer, Steven; 2000, "Introduction: Special Issue on Spoken Language Translation", Machine Translation 15, 1–2.

19. Levin, Lori, Alon Lavie, Monika Woszczyna, Donna Gates, Marsal Gavaldà, Detlef Koll and Alex Waibel ; 2000, "The JANUS-III Translation System: Speechto-Speech Translation in Multiple Domains", Machine Translation 15, 3–25.

20. Matsumoto, Yuji, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka and Masayuki Asahara ;2000; Morphological Analysis System ChaSen version 2.2.1, Technical Report, NAIST, Nara, Japan; available at http://chasen.aist-nara.ac.jp/.

21. Muriel Vasconcellos, (ed.) MT Evaluation: Basis for Future Directions. Proceedings of a workshop sponsored by the National Science Foundation, San Diego, California

22. Murthy, B. K., Deshpande, W. R. 1998. Language technology in India: past, present and future .http://www.cicc.or.jp/english/hyoujyunka/mlit3/7-12.html

23. Nida E. 1969, The Theory and Practice of Translation -Brill,

24. Nizar Habash and Bonnie J. Dorr. 2002. Handling Translation Divergences: Combining Statistical and Symbolic Techniques in Generation-Heavy Machine Translation. In Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas, AMTA-2002,Tiburon, CA.

25. Nizar Habash. 2002. Generation-Heavy Hybrid Machine Translation. In Proceedings of the International Natural Language Generation Conference (INLG'02), New York.

26. R. M. K. Sinha et. al. 2002. An English to Hindi Machine Aided Translation System based on ANGLABHARTI Technology "ANGLA HINDI", I.I.T. Kanpur, http://anglahindi.iitk.ac.in/.

27. R.D. Brown. 1996. Example-Based Machine Translation in the Pangloss System. In Proceeding of COLING-96: Copenhagen, pp. 169-174.

28. Sangal Rajeev et al. 2003. Machine Translation System:Shakti", http://gdit.iiit.net/~mt/shakti/.

29. Rao, Durgesh 2001. Machine Translation in India: A Brief Survey . SCALLA 2001 conference,Bangalore.http://www.elda.fr/proj/scalla/SCALLA2001/SCALLA2001Rao.pdf

30. Seligman, Mark ; 2000, "Nine Issues in Speech Translation", Machine Translation 15, 149–185.

31. Tillmann, Christoph, Stephan Vogel, Hermann Ney and Hassan Sawaf (2000) "Statistical Translation of Text and Speech: First Results with the RWTH System", Machine Translation 15, 43–74.

32. Vinay, J. P., and J. Darbelnet. 1977. Stylistique comparée du français et de l'anglais. Beauchemin.

33. Vinay, J. P., and J. Darbelnet. 1970. Cahier d'exercices (Work Book) 1. Stylistique comparée du français et de l'anglais. Beauchemin.

✡✡✡