CS626: Speech, NLP and the Web

Machine Translation

10th November 2020

Content we are going to cover today

- What is Machine Translation (MT)?
- Why is MT difficult?
- Different paradigms of Machine Translation
    - Rule based Machine Translation
    - Statistical Machine Translation
    - Example based Machine Translation
    - Neural Machine Translation
- Details of Neural Machine Translation
- Evaluation of Machine Translation systems

What is Machine Translation?

- Automatic conversion of text from one language to another
  - Preserve the meaning
  - Fluent output text

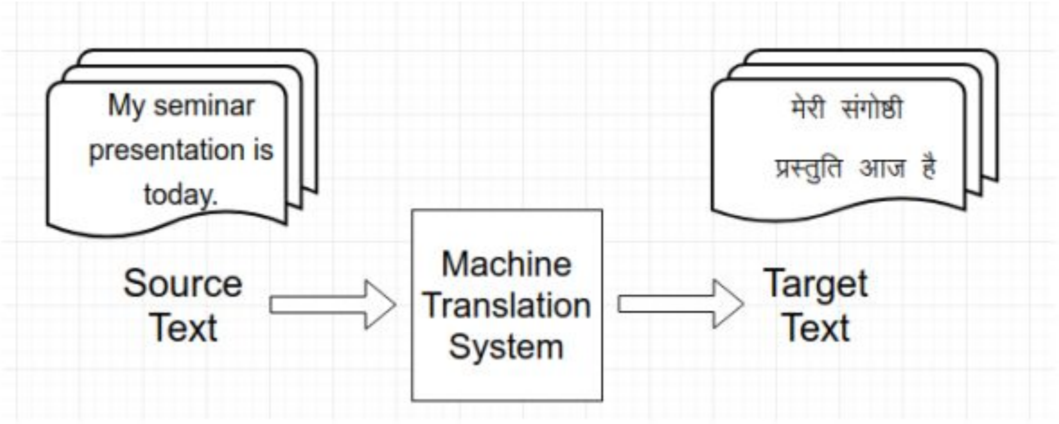


Figure: Machine Translation

History of MT

- 1954: First public demo of MT by IBM
  - Georgetown IBM experiment
- 1956: First MT conference
- 1972: Logos MT system
  - Translating military manuals into Vietnamese
  - Rule based approach
- 1993: Statistical MT
  - IBM models
- 2013: Neural Machine Translation

Why MT is hard?

Why MT is hard?

Language Divergence

Language divergence

- Languages express meaning in divergent ways
- **Syntactic divergence**
  - Arises because of the difference in structure
- **Lexical semantic divergence**
  - Arises because of semantic properties of languages

Different kinds of syntactic divergence

- Constituent order divergence (Word order)

English: He is waiting for him.
Hindi: वह उसके लिए इंतजार कर रहा है।

Subject	He	वह
Verb	waiting	इंतजार कर रहा है
Object	him	उसके

- Adjunction divergence

English: Delhi, the capital of India, has many historical buildings.
Hindi: भारत की राजधानी दिल्ली में बहुत सी एतिहासिक इमारतें हैं

- Null subject divergence

English: I am going.
Hindi: जा रहा हूँ।।

Different kinds of lexical semantic divergence

- Conflational divergence

English: He stabbed him.
Hindi: उसने उसे छुरे से मारा

- Categorial divergence (Lexical category change)

English: They are competing.
Hindi: वे प्रतिस्पर्धा कर रहे हैं

- Head-swapping divergence (Promotion or demotion of logical modifier)

English: The play is on.
Hindi: खेल चल रहा है

The Vauquois Triangle

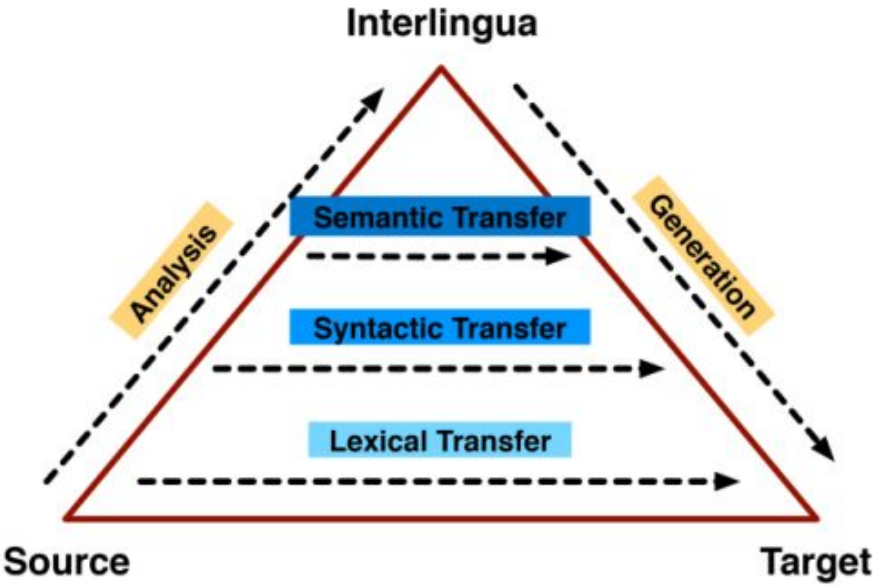


Image source: http://www.cs.umd.edu/class/fall2017/cmsc723/slides/slides15.pdf

Different paradigms of Machine Translation

- Rule based Machine Translation
- Statistical Machine Translation
- Example based Machine Translation
- Neural Machine Translation

Rule based Machine Translation

- Linguists create rules
- Three types
  - Direct
    - Map input to output with basic rules
  - Transfer based
    - Direct + Morphological and Syntactic analysis
    - The level of transfer is dependent on the language pairs
  - Interlingua based
    - Use an abstract meaning
    - Interlingua: Represent meaning of text unambiguously
    - It works at the highest level of transfer
- Performance of system highly dependent on experts who are creating rules

Statistical Machine Translation

- Learning from parallel corpora
- Three important things
  - Word translation
  - Word alignment
  - Word fertility management
- Problem to solve for SMT

$$\hat{e} = \arg\max_{e} \left(P(e|f)\right) = \arg\max_{e} \left(P(e).P(f|e)\right)$$

e is target language sentence, f is source language sentence, P(e) is language model in target language and P(f|e) is translation model.

Example based Machine Translation

- Majorly based on textual similarity
- Process
  - Analysis
    - Phrasal fragments of the input sentence
  - Transfer
    - Finding the aligned phrases from the database of examples
  - Generation
    - Recombination (Stitch together the aligned phrases)

Example based Machine Translation: Example

- He buys a book on Machine Translation.
- Phrasal fragments: He buys, a book, on, Machine Translation
- Aligned phrases: Identifies the aligned phrases from the database

He buys: वह खरीदता है

a book: एक पुस्तक

on: पर

machine translation: मशीन अनुवाद

- Recombination: Recombine those phrases to construct a sentence (Adjusting morphology, reordering)

वह मशीन अनुवाद पर एक पुस्तक खरीदता है।

Phrase based Statistical Machine Translation

- Why?
  - Translation of phrases is more intuitive
- Process involved
  - Two-way alignment
  - Symmetrization
  - Expansion of aligned words to phrases (Phrase table construction)

Phrase based SMT: English to Hindi alignment

	वह	आज	शाम	को	केक	बनाने	की	योजना	बना	रहा	है
He	✓										
is											✓
planning								✓			
to											
make						✓					
a											
cake											
in											
the											
evening			✓								

Phrase based SMT: Hindi to English alignment

	He	is	planning	to	make	a	cake	in	the	evening
वह	✓									
आज										
शाम										✓
को										
केक							✓			
बनाने					✓					
की										
योजना			✓							
बना										
रहा										
है		✓								

Phrase based SMT: Phrase generation

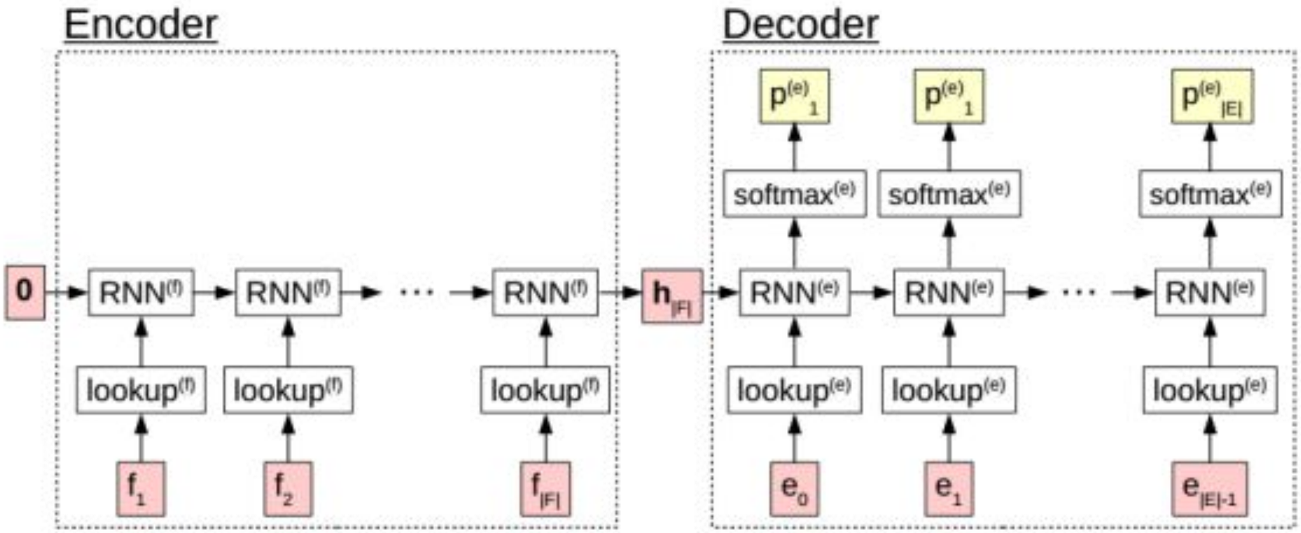
	वह	आज	शाम	को	केक	बनाने	की	योजना	बना	रहा	है
He	✓										
is											✓
planning								✓			
to											
make						✓					
a											
cake					✓						
in											
the											
evening			✓								

- Principle of coverage: Every word must be in a phrase
- Principle of non-vacuousness: No empty phrases
- Principle of consistency:The aligned phrases must be consistent in the sense all words of phrase in source languages

Neural Machine Translation

- Use of Neural network to predict the translation of a sentence
- Based on word sequence labeling
- Encoder-Decoder approach
  - Encoder encode the source sentence
  - Decoder generate the target sentence

NMT: Encoder-Decoder paradigm



- f_i = Source sentence words
- e_i = Target sentence words

Image source: http://www.phontron.com/class/mtandseq2seq2017/mt-spring2017.chapter7.pdf

NMT: Attention based Encoder-Decoder paradigm

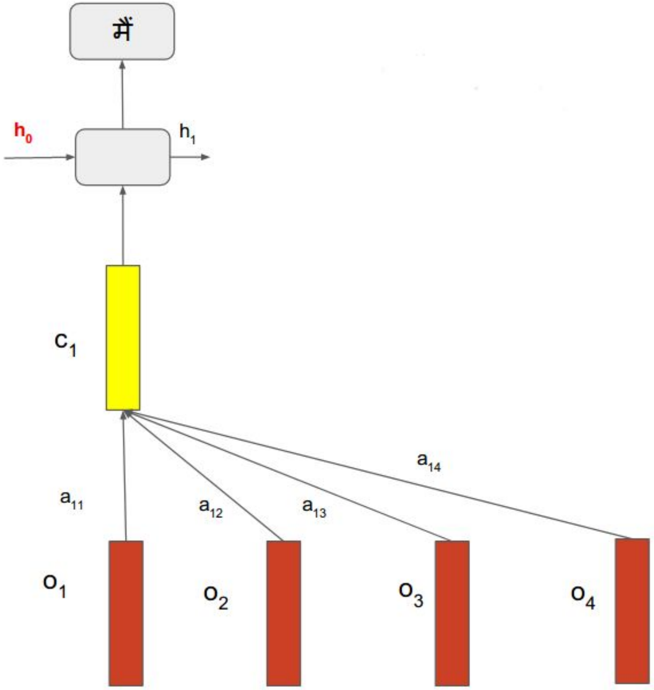


Image source- http://www.iitp.ac.in/~shad.pcs15/data/nmt-rudra.pdf

Different types of attention mechanism

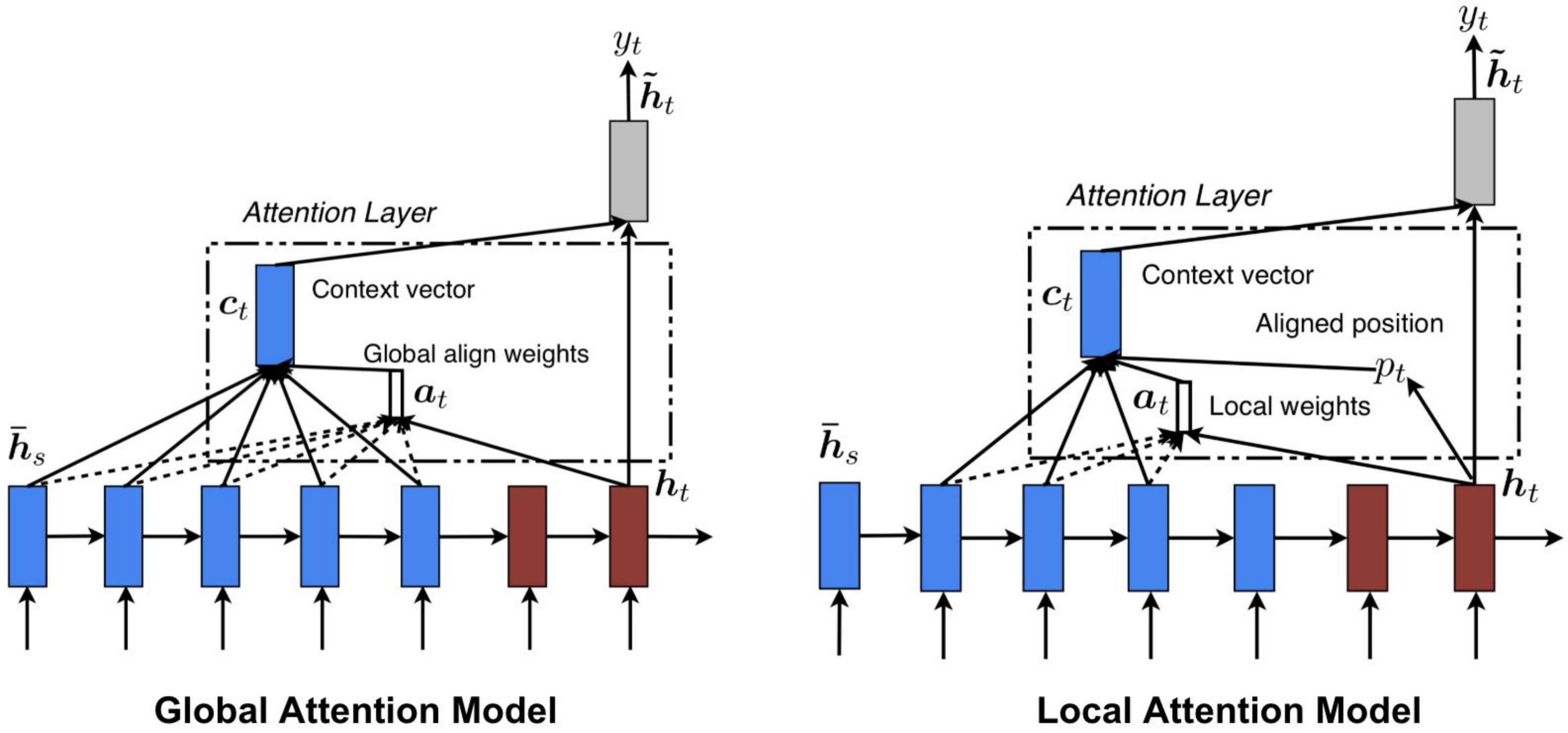
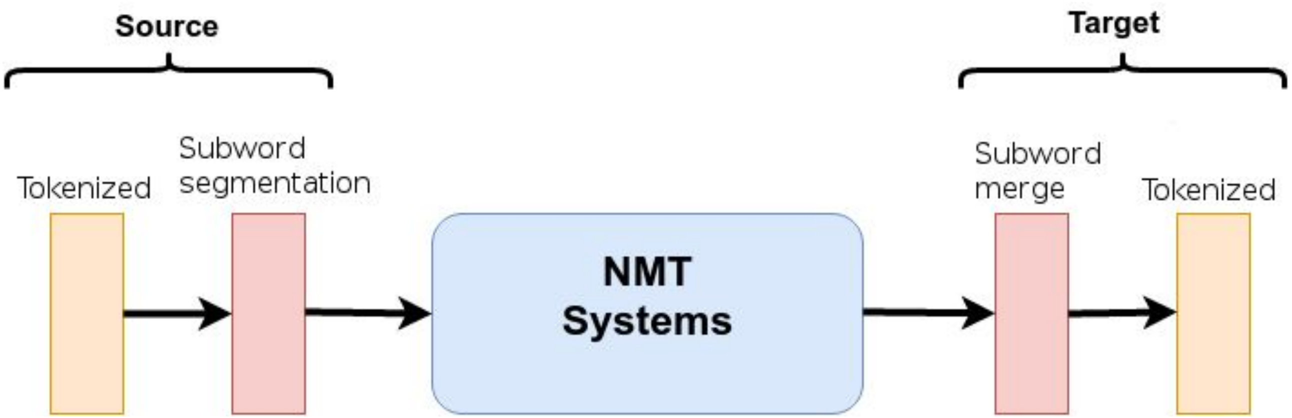


Image source: Luong, T., Pham, H. and Manning, C.D., 2015, September. Effective Approaches to Attention-based Neural Machine Translation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 1412-1421).

Subword NMT

- Compound words, words with morphological variation (need for morphological segmentation), named entities are very common
- We can utilise this phenomena, if we look into subword level.



Inference using beam search

- In greedy search, at each time step, one best hypothesis is considered, in beam search at each step b best hypothesis is considered

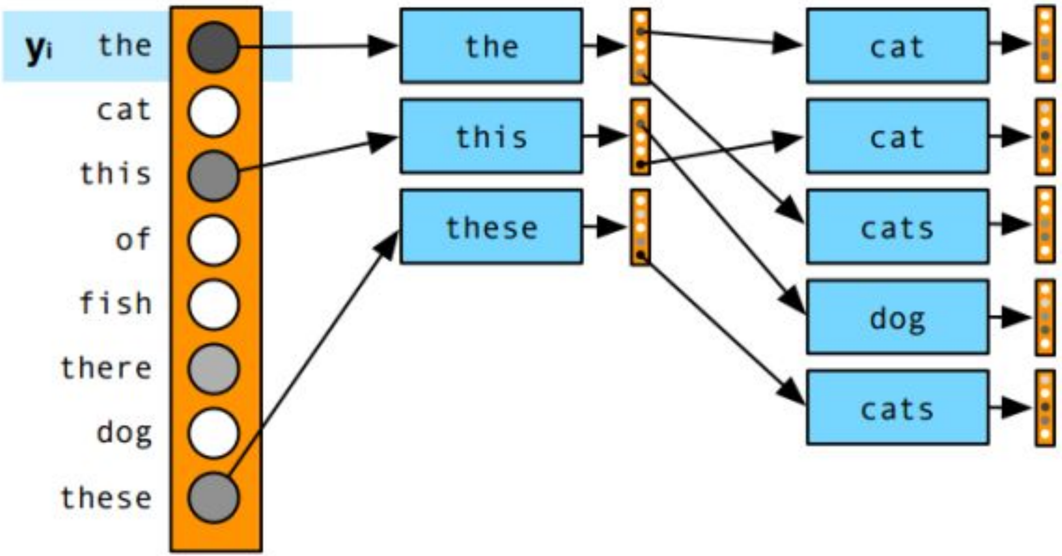


Image source: Philipp Koehn. Neural machine translation. CoRR, abs/1709.07809, 2017.

Back-Translation

- NMT needs large number of parallel sentences to train a model
  - Costly and time consuming task
- Can we utilize monolingual data?
- Back-Translation
  - What we need?
    - MT system (L2->L1) and L2 monolingual data
  - From monolingual data in L2 (target language), produce synthetic translation in language L1.
  - Train model for L1→L2

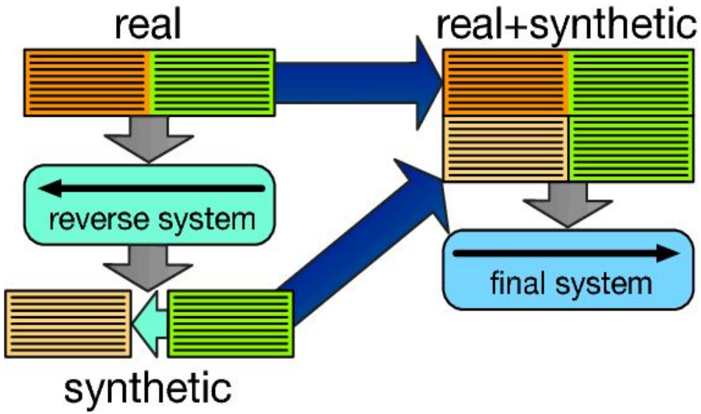


Image source: Hoang, V.C.D., Koehn, P., Haffari, G. and Cohn, T., 2018, July. Iterative Back-Translation for Neural Machine Translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation* (pp. 18-24).

MT Evaluation

- Manual evaluation
- Quality of sentence depends on two factors
  - Adequacy
    - How faithful the meaning of a sentence is transferred
  - Fluency
    - Acceptability of the native speaker

More fluent: मुझे भूख लग रही है।
Less fluent: मैं भूखा महसूस कर रहा हूँ।

- Automatic evaluation measures
  - Word/phrase matching based
  - Edit distance based
  - Ranking based

BLEU score

- Bilingual Evaluation Understudy
- Word/Phrase matching based

$$BLEU = BP.exp(\sum_{n=1}^{N}(w_n.log(p_n))$$

- BP is brevity penalty, to penalize based on the length of the generated sentence.

$$BP = \begin{cases} 1 & c > r \\ e^{(1-r/c)} & c <= r \end{cases}$$

c = the length of the candidate translation, r = the effective reference corpus length, p_n is modified n-gram precision, w_n is weight(uniform in BLEU)

BLEU score: Example

English: He is a painter.

Hindi (Candidate): वह एक चित्रकार चित्रकार चित्रकार है।

Hindi (Reference): वह एक चित्रकार है।

- Example:
  - 1-gram precision is 1.
  - Modified 1-gram precision is 4/6.
- The ratio of the number of phrases of length n present in candidate translation that are also present in reference translation and total number of phrases of length n in candidate translation.
- In modified n-gram precision maximum count from reference translation is considered for that particular n-gram.

TER

- Translation Edit Rate
- Edit operations
  - Insertion
  - Deletion
  - Substitution
  - Shift

$$TER = \frac{Number\ of\ edits}{Average\ number\ of\ reference\ words}$$

- Example:
  Reference: The TAs decided to give two homeworks this week.
  Candidate: This week the TAs decided to give two homeworks.
  - **This week** is shifted in the last. (cost is 2 units)

References

- Pushpak Bhattacharyya. Machine translation. CRC Press, 2015.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics, pages 311-318. Association for Computational Linguistics, 2002.
- Philipp Koehn. Neural machine translation. CoRR, abs/1709.07809, 2017.
- Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. In Proceedings of association for machine translation in the Americas, volume 200, 2006.

References

- Weiyue Wang, Jan-Thorsten Peter, Hendrik Rosendahl, and Hermann Ney. Character: Translation edit rate on character level. In Proceedings of the First Conference on Machine Translation: Volume Shared Task Papers, volume 2, pages 505-510, 2016.
- R Ananthakrishnan, Pushpak Bhattacharyya, M Sasikumar, and Ritesh M Shah. Some issues in automatic evaluation of english-hindi mt: more blues for bleu. ICON, 2007.
- Gregor Leusch, Nicola Ueffing, and Hermann Ney. Cder: Efficient mt evaluation using block movements. In 11th Conference of the European Chapter of the Association for Computational Linguistics, 2006.

- Gregor Leusch, Nicola Ueffing, and Hermann Ney. Cder: Efficient mt evaluation using block movements. In 11th Conference of the European Chapter of the Association for Computational Linguistics, 2006.

References

- Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. In Proceedings of association for machine translation in the Americas, volume 200, 2006.
- Kevin Duh. Ranking vs. regression in machine translation evaluation. In Proceedings of the Third Workshop on Statistical Machine Translation, pages 191-194. Association for Computational Linguistics, 2008.
- Alon Lavie and Abhaya Agarwal. Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments. In Proceedings of the Second Workshop on Statistical Machine Translation, pages 228-231. Association for Computational Linguistics, 2007.

Thank You