Multi-Engine Machine Translation Guided by Explicit Word Matching

Author：Shyamsundar Jayaraman and Alon Lavie

Professor：陳嘉平

Reporter：陳逸昌

Outline

- Introduction
- The MEMT Algorithm
- Experimental Setup
- Results
- Conclusion and Future Work

Introduction(1/2)

- It would be beneficial to have an effective framework for combining different systems into an MT system.
- It attempt to combine the best of outputs to produce output than any individual systems.

Introduction(2/2)

- Some MEMT approach uses standard Levenshtein edit distance but they doesn't accurately capture phrase movement like：
  - In the street, the children cried.
  - The children cried in the street
    1. Deleting the words "The children cried" in the second sentence.
    2. Inserting "The children cried" at the end of the sentence.

The MEMT Algorithm

- The word alignment matcher
- Basic hypothesis generation
- Part-of-speech based matching
- Matching window
- Alignment horizon
- Scoring of hypotheses
- A complete example

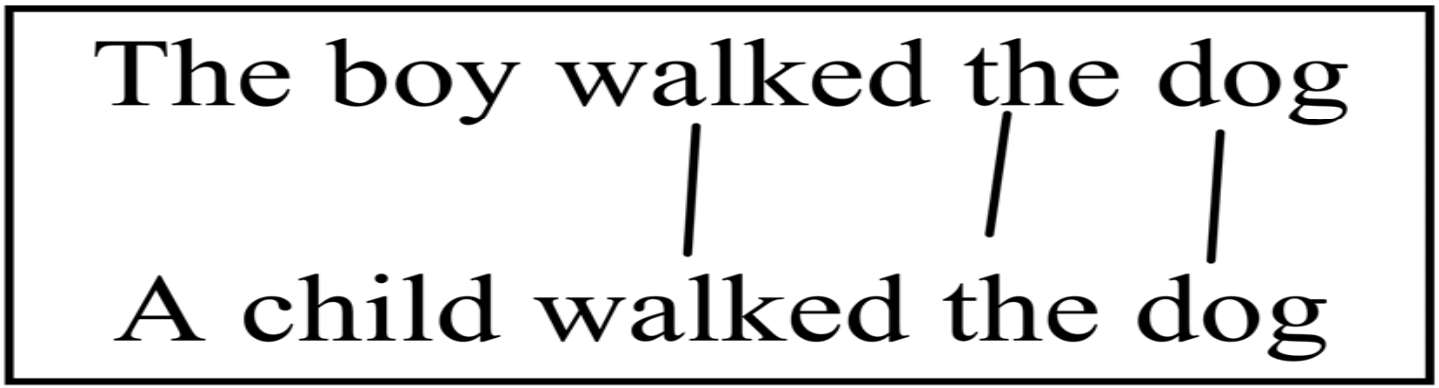
The word alignment matcher(1/3)

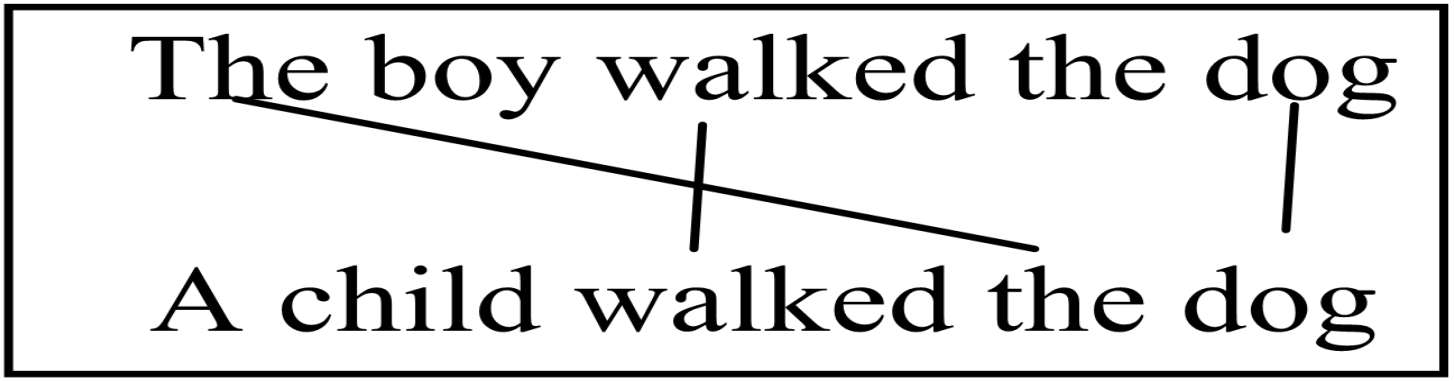
- Identical words, ignoring case, that appear in both input sentences are potential matches.
- This alignment is defined as the alignment that has the smallest number of “crossing edges”.

The word alignment matcher(2/3)

- The boy walked the dog
- A child walked the dog
  - The “walked” and “dog” occur only once in each sentence, their alignments are fixed.
  - The word “the” appears twice in the first sentence and only once in the second sentence.

The word alignment matcher(3/3)





Basic hypothesis generation(1/4)

- The generation algorithm is an iterative process and produces these translation hypotheses incrementally.
- The existing set of partial hypotheses is extended by incorporating an “unused” word from one of the original translations.
- A data structure keeps track of the accounted for “used” words that are associated with any partial hypothesis.

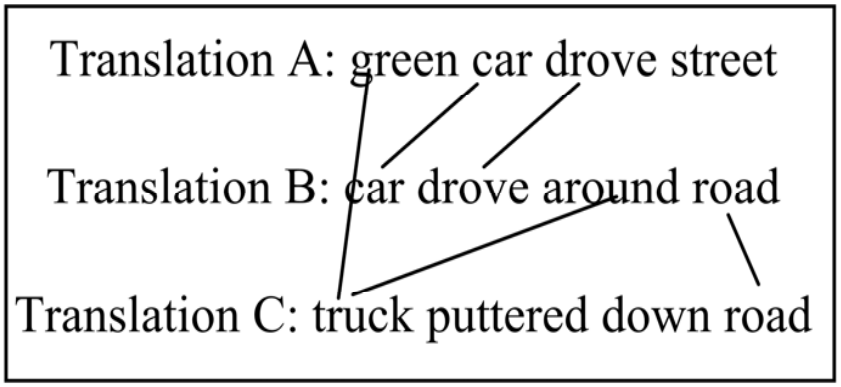
Basic hypothesis generation(2/4)

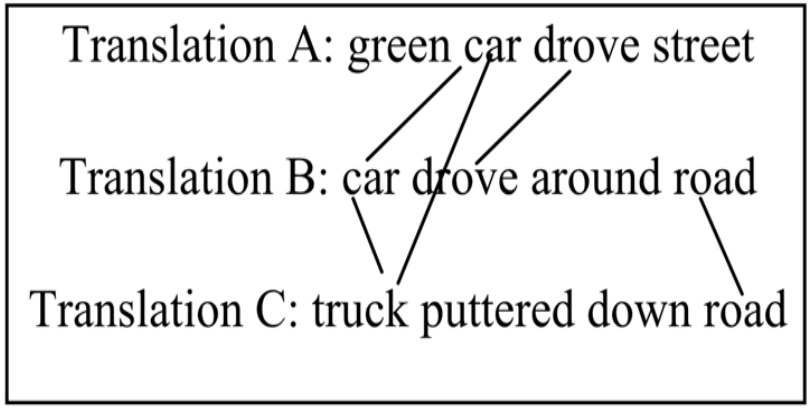
- At hypothesis is consider “complete” if one or more original translation strings propose to end the sentence.
- In each iteration, the decoder extends a hypothesis by choosing the first unused word from one of the original translation.
- If select word w then marks w as used and marks the alignment words as uesd.

Basic hypothesis generation(3/4)

- They attempt to identify a word that is likely to be a of the source language word that corresponds to w.
- The decoder tries to create an “artificial alignment“ between w and a word in the original translation, where there was no alignment found by the matcher.

Basic hypothesis generation(4/4)





- The decoder produces a hypothesis for each one of the artificial alignments.
- The decoder continues to iterate until all expandable hypotheses.

Part-of-speech based matching

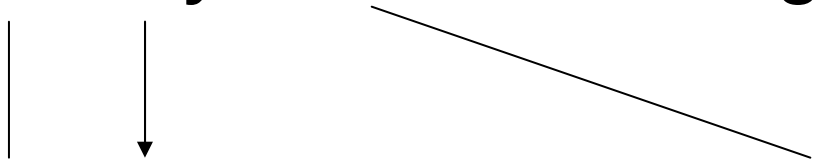
- The algorithm to only allow artificial matching between words that have the same part of speech.
- The system uses a dictionary that lists possible parts of speech for words in the target language.

Matching window(1/2)

- The matching window parameter that restricts how far ahead the decoder can look to artificially align a chosen word.

Matching window(2/2)

- The boy walked the large dog

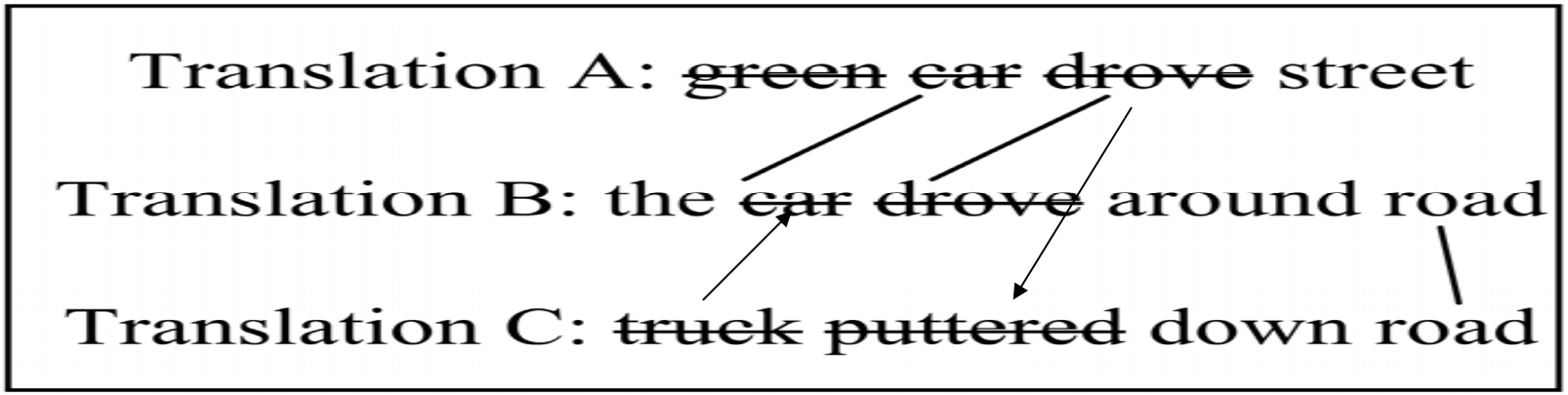


- The child who is male walked a big pet
  - Partial hypothesis : The boy walked
  - The matching window is set to 1.
  - When chosen “the”, then we can consider “who” “is” “male” “a” “big” as alignment.

Alignment horizon(1/2)

- Which is defined as the number of words behind the current word in the hypothesis, for which a word can still be considered for incorporation into the hypothesis.
- The horizon doesn’t affect words that are aligned to words in other original systems and that are within the horizon.

Alignment horizon(2/2)



- Partial hypothesis : Green truck drove
- Alignment horizon is set to 1.
- When after align, then delete the word “the”

Scoring of hypotheses

- Includes two components – a language model and a confidence score assigned to each word that is included in the hypothesis.

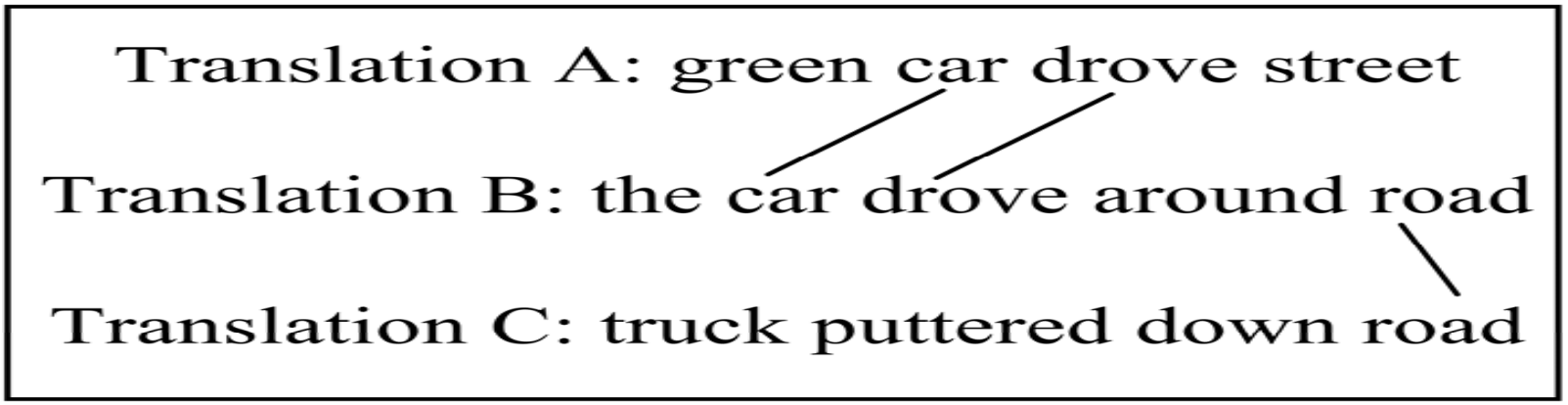
The confidence score(1/2)

- Each word supplied by any one of the original system is given a confidence score equal to the confidence score associated with the system which produced it.
- Two way to set the confidence scores of the original systems.

The confidence score(2/2)

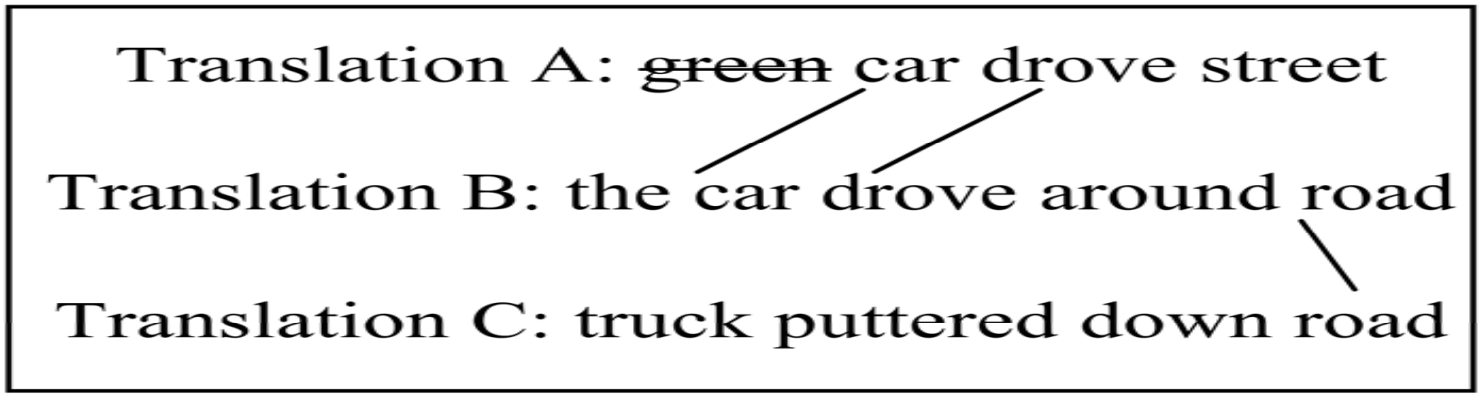
- Two way to set the confidence scores of the original systems
  - Score the original systems on a common test set then calculate confidences that are proportional to the relative ratios of the resulting systems scores.
  - Sets the scores to be relative to percentage of time that any one system produces the best translation for a sentence.

A complete example



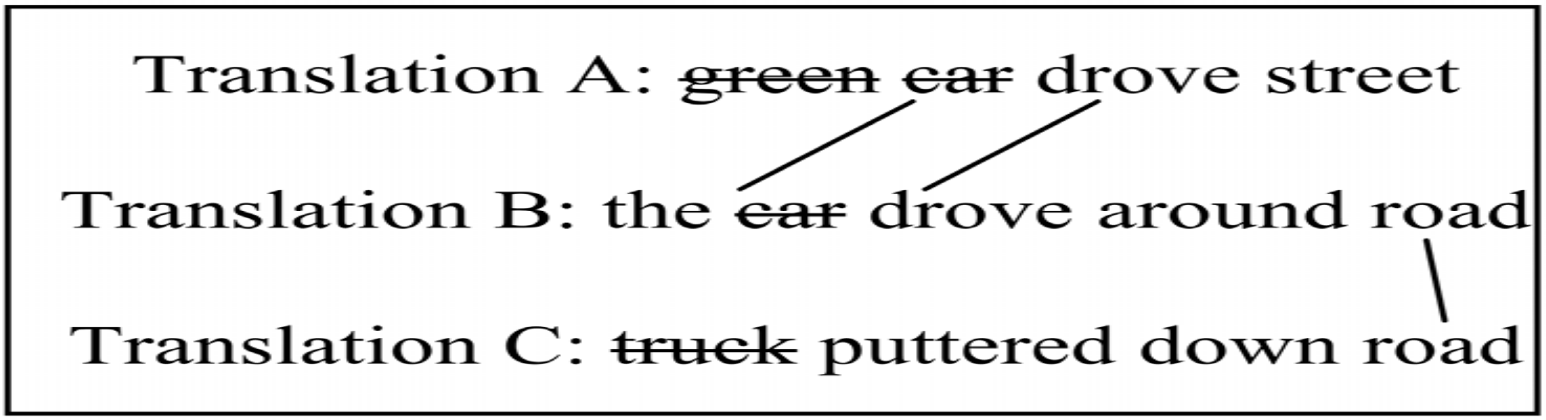
- The alignment horizon is set to two.
- The matching window is set to one.
- The MEMT system chooses the word “green”
- It wasn’t align to other words
- Attempts to find artificial alignments in translation B and C
- Matching window set to 1, so the boundary of translation B and C is car.

First iteration



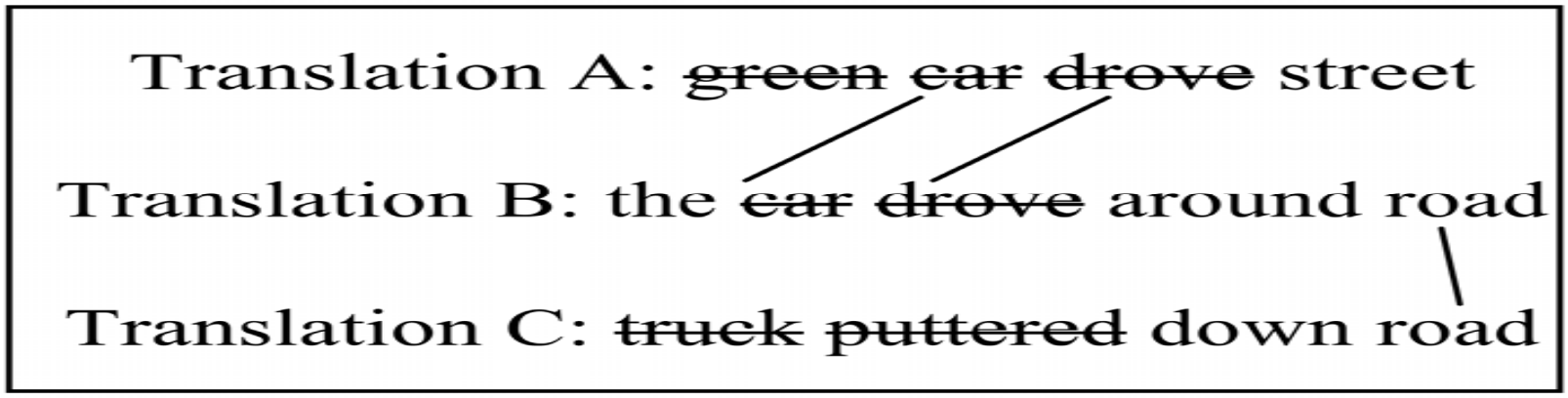
- The MEMT system chooses “truck” from translation C.
- Mark the “car” be used in the translation A.
- Mark the “car” be used in the translation B.

Second iteration



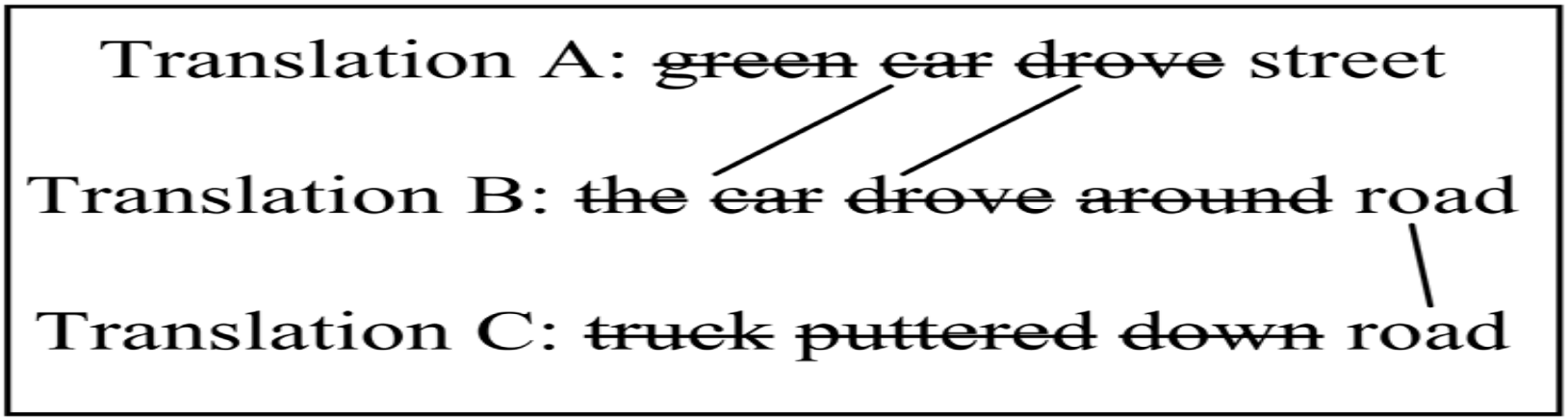
- The MEMT system chooses “drove” from translation A.
- Attempts to find artificial alignments in translation C.
- Mark the “puttered” be used.

Third iteration



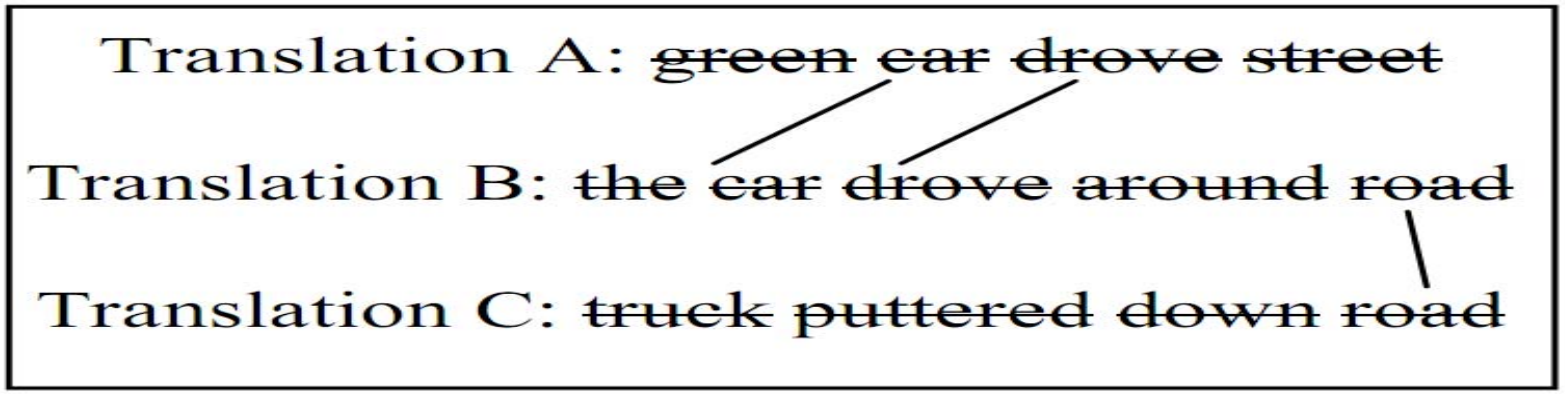
- The MEMT system chooses “down” from translation C
- Attempts to find artificial alignments in translation A and B.
- Mark the “around” be used.
- Alignment horizon is 2 so delete the word “the”.

Fourth iteration



- The MEMT system chooses “road” from the translation B.
- Mark the “street” be used.

Fifth iteration



- The MEMT system would mark “green truck drove down road” ad a completed translation and score it.

Experimental setup

- The outputs of three online translation systems outputs (Systran, Netat, and Wordlingo) on the TIDES 2002 Chinese and TIDES 2003 Chinese evaluation sets.
- 900 sentences of news wire text in simplified Chinese.
- Chinese to English.

Results(1/2)

System	METEOR Score
Online Translator A	.5225
Online Translator B	.5309
Online Translator C	.5225
Oracle (best original)	.5740
MEMT System	**.5673**

System	METEOR Score
Online Translator A	.5314
Online Translator B	.5453
Online Translator C	.5321
Oracle (best original)	.5821
MEMT	**.5762**
Oracle (best hyp)	.6268

- METEOR
  - Combination of unigram precision, unigram recall and an explicit penalty related to the average length of matched segments between the evaluated translation and its reference.

Results (2/2)

System	METEOR Score
Online Translator A	.4886
Online Translator B	.5047
Online Translator C	.4855
Oracle (best original)	.5440
MEMT System	**.5347**

System	METEOR Score
Online Translator A	.4917
Online Translator B	.4859
Online Translator C	.4910
Oracle (best original)	.5381
MEMT System	**.5301**
Oracle (best hyp)	.5840

Conclusion and future work

- The system achieves an improvement of about 6% over the best original system.
- The scoring algorithm is not yet capable of selecting the best generated hypothesis.
- Hope to identify salient features that will help further improve scoring and hypothesis selection.