Hybrid Machine Translation

EBMT

Kurt Eberle

April 19, 2018
Outline

Translation Memories

EBMT
Outline

Translation Memories

EBMT
Motivation

- Human Translators
- build Translation Memories and reuse the information stored
- (Parallel texts)
- CAT tools
- IBM, SAP, Taus...
Workflow I

find TM entries by similarity to sentence to translate
manage TM information in the CAT tool and show it to the user
CAT-Tools II

Workflow ... Example

• CAT-workbench... e.g. TRADOS

1) open text
2) statistics about TM/non-TM cost
3) translation by TM
4) copy for HT/MT
5) translate manually by using parts of TM suggestions
5) translate automatically

27.01.2014 Universität Mainz 48
Problem 1

Human translator must . . .

- identify relevant parts
- put pieces together
- manually
Problem I

Reinhard Schäler (Localization Research Center)

- http://www.localisation.ie/
- TM resources are underused
- → consider parts of entries (‘phrases’)

- Lit: Schäler, R.: Beyond Translation Memories (EAMT 2001)
  Becker et al: The Phrasal Lexicon
What are 'Phrases'? 

Becker’s theory (on human translation)
What are 'Phrases'? II

Becker’s Theory of Language Production

Information to be conveyed

Attitudes to be expressed or invoked

Phrasal Lexicon

Phrase Adaptation
Stitch phrases together
Fill in the blanks
Modify phrases

Language Production

Generative Gap-Filling
Smooth over the transitions
Fill in conceptual holes
Idea 1

Automatize process

- identify relevant parts and correspondences in TMs
- store (sub-sentential correspondences)
- find best cover of the source sentence by such source parts
- combine corresponding target parts
Example I
Example II


(1)  
a. EN: The bullets move to the new paragraph.  
   DE: Die Blickfangpunkte rücken in den neuen Abschnitt.  
b. EN: The title moves to the center of the slide.  
   DE: Der Titel rückt in die Mitte des Dias.  

Despite the fact that the TM entries in (1) are good matches for the new sentence in (2), it would not be translated automatically by the TM system:

(2)  
The bullets move to the center of the slide.  

At most, the system would be capable of identifying one (or both) of the two source sentences in the TM in (1) as fuzzy matches. If so, these would be displayed together with their translations, which would then have to be adapted by a translator to generate the final output string in (3):³

(3)  
Die Blickfangpunkte rücken in die Mitte des Dias.
Phrasal lexicon I

(Becker et al. 1975: units 'stitched' together in language production)
Phrasal lexicon II

Figure 1.2. Overview of the Phrasal Lexicon
Phrasal lexicon I

First task

- analyze sentences into parts
- identify correspondences in target sentences
- store 'phrases'
- then ...
Outline

Translation Memories

EBMT
Cover I

(Watanabe 1992)
Problem: I

Different possibilities
Problem: II

Figure 1: Selecting the most appropriate set of rules.
(Maruyma/Watanabe 1992)

"For instance, assume that the input sentence is represented as (a) in Figure 1, and that (b), (c), (d), and (e) are translation examples. The problem is to find the most appropriate combination of the examples that completely covers the input structure (a). In this case, (b),(e) and (c),(d) are the candidates of covers. Suppose that matching cost of (b) is 2, (c) is 1, (d) is 5, and (e) is 1, then matching cost of (b),(e) is smaller than that of (c),(d)."
Task 1

(Maruyama/Watanabe 1992)

- Similarity Calculation - To calculate the similarity between an input and a source part of a translation example.
- Example Selection - To select the most appropriate set of translation examples.
- Target Construction - To construct a target sentence by combining the target parts of the chosen examples.
Suggestion 1

- Tree Cover Search Algorithm for Example-Based Translation