A REAL-TIME MT SYSTEM FOR TRANSLATING BROADCAST CAPTIONS[1]

Eric Nyberg and Teruko Mitamura
Center for Machine Translation
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

ABSTRACT

This presentation demonstrates a new multi-engine machine translation system, which combines knowledge-based and example-based machine translation strategies for real-time translation of business news captions from English to German.

1. Introduction

Broadcast captioning is derived from a textual transcription of a television broadcast, and is typically prouduced in real-time by a human operator using a stenography machine. The ASCII transcription is then encoded and transmitted in the vertical blanking interval (VBI) portion of the video signal. Some broadcasts (such as entertainment shows in syndication), are captioned off-line, but many shows of interest (such as "headline news" programs) are always captioned in real-time, with only a few seconds delay between the audio signal and the transcribed captions.

Real-time translation of broadcast captioning poses several challenges for machine translation. The vocabulary and grammar of the source text are uncontrolled. It is not feasible to employ either pre- or post-editing during the broadcast. Multinational broadcasts typically must convey timely information about business news, politics, etc., with little or no margin for translation error. To produce adequate translations in this environment, these challenges are met through an appropriate synthesis of MT techniques.

Our system combines a knowledge-based component and an example-based component to translate business news captions from English to German. In the remainder of this paper, we briefly describe the requirements for caption translation and give an overview of the system architecture and current status.

2. Requirements for Caption Translation

Successful translation of broadcast captioning must meet the following requirements:

- *The system must deal with transcription of spoken language*. Unlike translation of textual material (technical manuals, etc.), translation of broadcast captioning must address the difficulties of unknown words (proper names, company names, etc.), fragmented input (repetitions, partial sentences, run-on sentences), transcription errors (typographical errors, omissions), metaphoric and idiomatic use of language, etc. Since professional captioners try to render as literal a transcription as

[1] Proceedings of Machine Translation Summit VI, San Diego, CA, Oct. 29 – Nov. 1, 1997.

possible, spoken elements of the broadcast can contain language phenomena which are difficult to translate.

- *The system must perform in real-time*. The captions (and their translations) must follow the spoken broadcast by no more than a few seconds, and response time should be constant, regardless of variations in input complexity.

- *The system cannot make use of pre- or post- editing*. Pre-editing is not feasible, because there isn't enough time available to correct errors in the captioning stream. Pre- and/or post-editing is possible only when the broadcast will be time-delayed.

- *The system must produce texts which are as accurate as possible*. Most broadcasts which are captioned in real-time relate timely information (e.g., business news, stock market reports, etc.) which should be translated accurately to avoid misinformation on a critical topic.

Although these requirements are challenging for MT, broadcast captioning does have characteristics which make it amenable to automatic machine translation:

- *The domain of discourse is narrow*. For some broadcasts, the content of the most important, timely information (e.g., corporate takeovers, stock closings, etc.) varies very little.

- *It is possible to build up a large corpus of data from daily broadcasts*. It is straightforward to capture data from the targeted broadcast, through use of a caption decoder connected to the video signal.

- *It is not necessary to translate every utterance in the broadcast*. Typical broadcast segments have a fairly well-defined discourse structure (greetings, 'headline' sentences, descriptive text, short utterances conveying raw data, etc.). It is not necessary to translate those portions of the broadcast which don't translate cross-culturally (e.g., metaphoric use of language, such as "*The bulls outrunning the bears today on Wall Street"*). The focus is on accurate translation of the facts, rather than the mode of presentation, stylistic quality or variation.

3. System Architecture

Because we are targeting timely broadcasts in narrow domains (e.g., business news update), and because accurate translation is a priority, our implemented solution contains a knowledge-based component which allows full understanding of important pieces of the broadcast (e.g., stock closings, mergers, buy-outs, etc.). The repetitive nature of the broadcasts, coupled with the variation in the phrasing of various stock utterances (like stock closings), is well-suited to an example-based approach, both to handle predictable variation in syntax and vocabulary (e.g., proper/place names), and to provide more efficient processing than a full parse/generate approach. And finally, because of the uncontrolled and errorful nature of the input stream, it is necessary to build these two modules into a multi-engine architecture, which can combine the output of the two modules to provide the best possible translation of each input. The architecture of the system is shown in Figure 1.

Our system is based on the Multi-Engine Machine Translation (MEMT) architecture (Frederking, et al., 1994). The system will integrate the following modules:

- **Segmentation**. The segmentation module will break the raw captioning stream up into individual sentences to be translated. The performance of this module will depend directly on how well the captioners follow simple training guidelines for segmenting the input (e.g., always end a sentence with a period and newline).

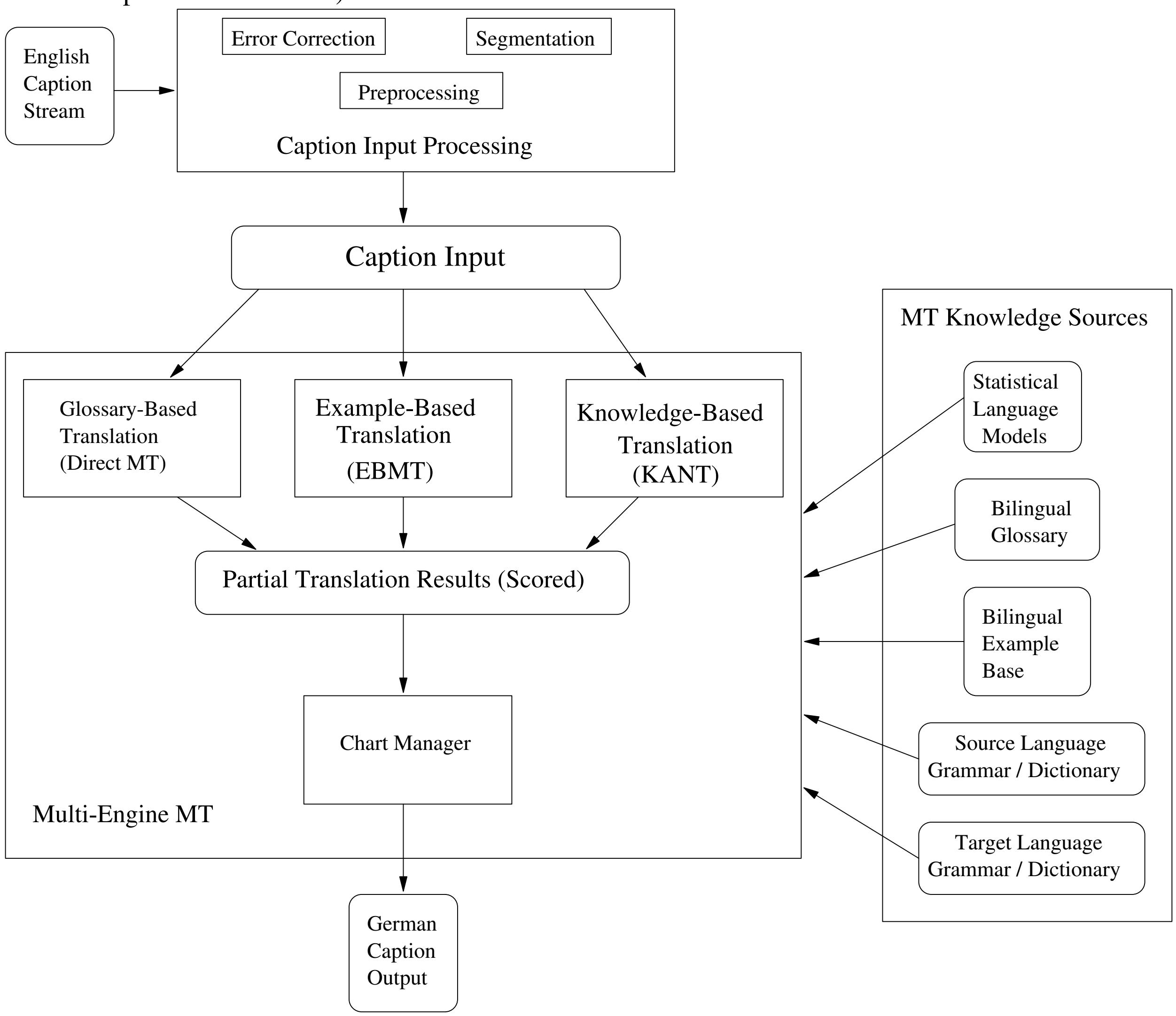

Figure 1: Multi-Engine MT Architecture for Caption Translation

- **Error Correction**. The error correction module attempts to recover sentences or fragments which contain obvious spelling errors, missing / redundant punctuation, etc. We intend to model and correct only those frequently-occurring errors which can be quickly corrected. The usefulness of this module will also depend on the quality of the input captioning stream.

- **Pre-Processing.** The pre-processing module performs various “chunking” functions (grouping together words / phrases that should be treated as a single translation unit, recognizing unknown company/place names) and rewriting functions (e.g., expanding contracted forms, abbreviations, etc.).

- **Knowledge-Based Translation**. The KBMT module performs a full analysis and generation of each sentence, or coherent portion thereof. For each translated sentence or fragment, it assigns a score indicating the estimated quality of the translation (this scoring mechanism is used by the Multi-Engine module to select the best translation overall). The KBMT module is realized through use of the KANT software (Mitamura, et al., 1991). KANT uses explicitly coded lexicons, grammars, and semantic rules to perform translation. KANT is also an interlingual system, making use of an explicit intermediate representation which acts as a "pivot" between the source and target languages. KANT has three main characteristics which make it useful for this application:

- *Increased accuracy of translation*. By allowing the creation of lexicons and grammars that can be as simple or as complex as the application requires, KANT supports a high degree of accuracy in both the source analysis and target generation phases.

- *Efficient support of multiple target languages*. Through use of an intermediate interlingua representation, KANT's architecture allows straightforward extension to new target languages – in fact, the interlingua produced from a single source text may be used to generate several target texts.

- *Separation of code and knowledge bases*. All of the software in KANT (analysis and generation modules) is independent of the language pair being translated. When a new output language or language pair is desired, all that is required is "plugging in" the new knowledge sources; the system code remains constant. Unlike transfer-based systems, the addition of a new target language does not require any replication or redesign in the analysis phase of translation; rather, new grammars and lexicons are added for each new target language, which use the intermediate interlingua representation as their input.

KANT is a good fit for broadcast captioning, at least where these criteria apply:

- The texts to be translated focus on an area of technical information (such as business news), such that it is feasible to design a domain-specific terminology and lexicon;

- A high degree of output accuracy is required, such that the benefits of knowledge-based translation are cost-justified;

- The volume of text to be translated (and the impossibility of manual translation) justify the effort to develop an automatic machine translation system.

Of course, KANT is designed for use with well-defined grammars and input languages, so its performance degrades in the presence of ungrammatical or ill-formed input. Since the captioning input stream is subject to a fairly high error rate (and various speech phenomena such as restarts, interjections and run-ons), KANT must be supplemented by a more robust approach for proper overall handling of difficult input.

- **Example-Based Translation**. The EBMT module is used to match portions of the input which can be translated from a memory of past translations, using a pre-stored database of example patterns and translations. For each translated sentence or fragment, it will assign a score indicating the estimated quality of the translation. This scoring mechanism is used by the Multi-Engine module to select the best translation overall. It is expected that performance of EBMT will increase over time as translation memory grows.

- **Statistical Language Model**. We are now gathering a large corpus of German text, which is used to build a statistical n-gram model of word occurrences. The translation modules use the probabilities of various co-occurrences to inform decisions regarding the relative quality of competing translations.

- **Multi-Engine Module**. The Multi-Engine module is responsible for monitoring the progress of each translation technique on a given input, terminating the computation(s) when the alloted time for a given input has been used up. It uses the scores assigned by the KBMT/EBMT modules to each full or partial translation they produce, in order to piece together the best possible translation for the input. If the best possible translation does not receive an aggregate score which is above a certain threshold of acceptability, no translation will be produced. It is expected that garbled sentences and long, complex sentences may not be translated fully, but that shorter sentences containing relevant, crucial facts will almost always receive a good translation.

The multi-engine approach is crucial in a domain where relevant information is sometimes embedded in complex or ungrammatical constructions. In cases where the input is garbled, the EBMT module (in partial/constituent mode) can identify and translate those portions of the input which are deemed relevant. On the other hand, the KBMT module will have the ability to translate much longer and more varied inputs, improving the quality of the system overall. The two techniques are combined by the multi-engine approach to make the best of both.

4. Current Status

We have completed a working KBMT prototype of the system, which has been demonstrated on a portion of a headline news broadcast (see Appendix 1). We are in the process of scaling up the example base, the statistical language model, and the MEMT scoring engine. After the system is complete, it will undergo extensive field testing, during which the performance of the system will be evaluated with respect to practical use in a real-time environment.

Acknowledgements

The authors would like to thank Dieter Waeltermann, Bob Igo, Remi Chadel and Krzysztof Czuba for their collaboration on the KBMT prototype for this system. We also acknowledge the ongoing contributions of Bob Frederking, Ralf Brown, and Jaime Carbonell to the system architecture and software design.

References

Frederking, F., et al., "Integrating Translations from Multiple Sources within the PANGLOSS Mark III Machine Translation System," *Proceedings of the First Conference of the Association for Machine Translation in the Americas*, Columbia, MD, 1994.

Mitamura, T., et al., "An Efficient Interlingua Translation System for Multi-lingual Document Production," *Proceedings of the Third Machine Translation Summit*, Washington D.C., 1991.

Nyberg, E., et al., "The KANT Machine Translation System: From R&D to Initial Deployment," *Proceedings of June 1997 LISA Workshop on Integrating Advanced Translation Technology*, Washington, D.C.

Appendix 1: Sample Caption Translations

Note that some of the captions in this brief sample are not translated; for example, sentence 3 is too complicated to be translated by the present system, so no output is produced. In other cases (e.g., sentence 2) the German translation is either partial or not of the best quality. Nevertheless, the quality of translation produced by the prototype system is understandable and useful.

1. ON WALL STREET THE DOW INDUSTRIALS MISSED "RECORD NUMBER NINE" BY JUST ONE POINT.

   Der Dow-Jones-Index verpaßte 'Rekord Nummer Neun' an der Wall Street um einen Punkt.

2. IT WAS A SCHIZOPHRENIC KIND OF DAY OF TRADING ALL DAY LONG:

   Es war ein verrückt Art der Handelstag:

3. THE BLUE CHIPS WERE UP AS MUCH AS 24 POINTS, AND DOWN AS MUCH AS 20, FINALLY CLOSING ON THE DOWNSIDE.

   (no translation)

4. LET'S RUN DOWN THE FINAL NUMBERS NOW FROM WALL STREET.

   Gehen wir jetzt die Abschlußzahlen von der Wall Street durch.

5. AND TOPPING TODAY'S BUSINESS HEADLINES:

   Und ganz oben in den Schlagzeilen der Wirtschaft von Heute:

6. IT CERTAINLY WAS ANOTHER MERGER MONDAY.

   Es war sicherlich ein weiterer Montag der Fusionen.

7. TOSCO SEALED A DEAL TO BUY THE WEST COAST OPERATIONS OF UNOCAL ALSO KNOWN AS "76 PRODUCTS" COMPANY FOR ABOUT $1.4 BILLION.

   Tosco brachte einen Handel, um die Westküsten-Betriebe von Unocal, das auch als '76 Products' bekannt ist, für ungefähr 1,4 Milliarde Dollar zu kaufen, zum Abschluß.

8. TOSCO WILL BECOME THE NATION'S LARGEST INDEPENDENT REFINERY.

   Tosco wird zur größten unabhängigen Raffinerie des Landes werden.

9. UNOCAL SAYS IT WILL USE SOME OF THE PROCEEDS TO PARE DOWN DEBT.

   Unocal sagt, daß es einen Teil der Einnahmen verwenden wird, um Verschuldung abzubauen.

10. UNOCAL WILL ALSO CONSIDER A STOCK REPURCHASE PLAN.

    Unocal wird auch einen Aktienrückkaufplan erwägen.

11. SHARES OF TOSCO CLOSING THE DAY UP 12 3/8.

    Aktien von Tosco beschlossen den Handelstag höher mit 12 3/8.

12.UNOCAL SHARES ALSO HIGHER ON THE SESSION.

Unocal Aktien beschließen den Handelstag auch höher.

13.THEY ADDED 1 1/4 POINTS.

Sie gewannen 1 1/4 Punkte.

14.MATTEL BUYING TYCO TOYS FOR \$755 MILLION.

Mattel kauft Tyco Toys für 755 Millionen Dollar.

15.THAT DEAL WILL BE IN STOCK.

Der Handel wird in Form von Aktien stattfinden.