

Hybrid Machine Translation Applied to Media Monitoring

Hassan Sawaf, Braddock Gaskill, Michael Veronis

AppTek Inc.
6867 Elm Street #300
McLean, VA 22101, USA

{hsawaf,bgaskill,mveronis}@apptek.com

Abstract

In this paper, a system is presented that recognizes spoken utterances in Arabic Dialects which are translated into text in English. The input is recorded from a broadcast channel and recognized using automatic speech recognition that recognize Modern Standard Arabic and Iraqi Colloquial Arabic. The recognized utterances are normalized into Modern Standard Arabic and the output of this Modern Standard Arabic interlingua is then translated by a hybrid machine translation system, combining statistical and rule-based features.

1 Introduction

There has long been a need and desire for better quality translation. Hearing the spoken word and translating it correctly are two separate processes. Recognizing speech and converting it to its written form is one. The other is taking that transcript and translating it into another language. Early success in news monitoring applications was considered to be the ability to achieve a consistent 60% or better accuracy in recognition and translation. What happens if the speech recognition is flawed and does not detect everything needed for an accurate transcript and then the machine translation tries to process that transcript based on errors or missed words?

In some cases, automatic machine translation (MT) can close the gap by using additional information: rule-based machine translation systems (RBMT) aid in correcting these errors by using

semantic information while statistical machine translation systems (SMT) use graph-oriented decoding mechanisms on multiple hypotheses from the automatic speech recognition (ASR) to correct for the ASR mistakes (Ney et.al. 2000).

These disparate systems each have their own strengths and weaknesses so, independently, they were able to contribute to a part of a solution. Hybrid machine translation (HMT) elevated the improvement in the final output of an ASR or media monitoring system by combining the key qualities of RBMT and SMT to generate a more readable and reliable translated transcript.

In comparison with written language, speech and especially spontaneous speech poses additional difficulties for the task of MT. Typically, these difficulties are caused by errors of the recognition process, which is carried out before the translation process. As a result, the sentence to be translated is not necessarily well-formed from a syntactic point-of-view.

Even without ASR errors, speech translation has to cope with a lack of conventional syntactic structures because the structures of spontaneous speech differ from those of written language. A prime motivation for creating a hybrid machine translation system is to take advantage of the strengths of both rule-based and statistical approaches, while mitigating their weaknesses.

Thus, for example, a rule that covers a rare word combination or construction should take precedence over statistics that were derived from sparse data (and therefore not very reliable). Additionally, rules covering long-distance dependen-



Missed or omitted words in the source will multiply errors in the target translation.

Hybrid MT can correct certain errors and omissions to produce a complete transcript and better translation.

Figure 1: MediaSphere Screenshot of a Broadcast Transmission: Left the original Arabic text, the Names are marked in red; Right the translation using AppTek's Hybrid Machine Translation

cies and embedded structures should be weighted favorably, since these constructions are more difficult to process in statistical machine translation.

Conversely, a statistical approach should take precedence in situations where large numbers of relevant dependencies are available, novel input is encountered or high-frequency word combinations occur.

An aspect that is extremely important, especially in regards to processes that extract information from text (e.g. a “distillation engine”, prepar-

ing audio and textual data for question answering) is the weakness that SMT sometimes has in “informativeness” (the accurate translation of information) and “adequacy” (how well the meaning of the test translation matches the meaning of the reference translation) due to the influence of the target language model. For example single words that may make a disproportionately heavy contribution to informativeness and adequacy, such as terms indicating negation or important content words may be missing.

On the other hand, rule-based systems may excel with respect to informativeness. For exam-

ple, a Lexical Functional Grammar (LFG) rule-based system almost equaled the capability of novice translators, and was not far behind expert translators, in respect to informativeness in a previous evaluation (Doyon et.al., 1999).

It can be expected that the use of rule-based machine translation in conjunction with statistical machine translation would greatly improve informativeness, by imbuing statistical machine translation with all necessary features of a good rule-based machine translation system, ensuring high fluency (a definitive strength of the statistical approach) and increasing adequacy and informativeness (using embedded rule-based machine translation features).

A brief comparison of the two systems will help us illustrate how the hybridized MT approach unites very valuable features to form a comprehensive system:

RBMT provides fidelity, meaning that informativeness and adequacy are greater than fluency. This means that RBMT output might not read as well, but it is usually more accurate.

SMT systems strive for accuracy, as well, but are more noted for their fluency. MT output reads well and that gives a more immediate appearance of correctness. Both systems have attractive and useful qualities to generate an output that is useful.

HMT brings those key features into one system to deliver output that reads well and is true to the context of the spoken word. This has improved the ability of our automated media monitoring system to capture live feeds that contain both broadcast quality speech and unrehearsed interviews from the field, transcribe them despite dialect and other spoken nuances, and create an English translation that captures the true meaning of each word that was spoken.

1.1 Statistical MT Module

Statistical Machine Translation (SMT) systems have the advantage of being able to learn translations of phrases, not just individual words, which permits them to improve the functionality of both example-based approaches and translation memory. Another advantage to some SMT sys-

tems, which use the Maximum Entropy approach, is that they combine many knowledge sources and, therefore give a good basis for making use of multiple knowledge sources while analyzing a sentence for translation.

1.2 Rule-Based MT Module

For the presented rule-based module, an LFG system (Shihadah&Roochnik, 1998) is employed which is used to feed the hybrid machine translation. The LFG system contains a richly-annotated lexicon containing functional and semantic information.

1.3 Hybrid MT

In the hybrid machine translation (HMT) framework introduced in this paper, the statistical search process has full access to the information available in LFG lexical entries, grammatical rules, constituent structures and functional structures. This is accomplished by treating the pieces of information as feature functions in the Maximum Entropy.

Incorporation of these knowledge sources both expand and constrain the search possibilities. Areas where the search is expanded include those in which the two languages differ significantly, as for example when a long-distance dependency exists in one language but not the other.

2 Description of HMT Approach

Statistical Machine Translation is traditionally represented in the literature as choosing the target (e.g., English) sentence with the highest probability given a source (e.g., French) sentence.

Originally, and most-commonly, SMT uses the “noisy channel” or “source-channel” model adapted from speech recognition (Brown et.al.,1990;Brown et.al.,1993).

While most SMT systems used to be based on the traditional “noisy channel” approach, this is simply one method of composing a decision rule that determines the best translation. Other methods may be employed and many of them can even be combined if a direct translation model using a Maximum Entropy is employed.

Such a method enables improved language modeling, for it allows postulating proper independence assumptions that reflect the knowledge of causality in the real world. In the field of statistical parsing, (Collins, 1999) and (Charniak, 2000) place in their work place a large emphasis on effective parameterization of their models in order to model real-world causality.

2.1 Translation Models

The translation models introduced for the system which is described herein is a combination of statistically learned lexicons interpolated with a bilingual lexicon used in the rule-based LFG system.

2.2 Language Models

In this paper the use of lexical and grammatical feature functions in a statistical framework is introduced. The incorporation of rich lexical and structural data into SMT helps accelerate an emerging trend in SMT, which is the use of linguistic analysis. Analogous to the work of (Sawaf et.al., 2000; Charniak et.al., 2003; Och&Ney, 2004) to improve MT quality language model feature functions of the following form used: the language model feature functions cover standard 5-gram, POS-based 5-gram and time-synchronous CYK-type parser, as described in (Sawaf et.al., 2000). The m-gram language models (word and POS class-based) are trained on a corpus, where morphological analysis is utilized.

Then a hybrid translation system is trained to translate the large training corpus in non-dialect language into the targeted dialect language. After that, the new “artificially” generated corpus is utilized to train the statistical language models. For the words, which do not have a translation into the target language, are transliterated, using a transliteration engine, conceptionally borrowed from Grapheme-to-Phoneme converter like (Bisani&Ney, 2002; Wei, 2004). Besides this corpus, the original text corpus is used for the training of the final language models.

2.3 Functional Models

The use of functional constraints for lexical information in source and target give a deeper

syntactic and semantic analytic value to the translation. Functional constraints are multiple, and some of these functions are language dependent (e.g. gender, polarity, mood, etc.).

These functions can be cross-language or within a certain language. A cross-language function could be the tense information but also the function “human”, describing that the concept to be generally a human being (“man”, “woman”, “president” are generally “human”, but also potentially concepts like “manager”, “caller”, “driver”, depending on the semantic and syntactic environment).

A “within-language” function could be gender, as objects can have different genders in different languages (e.g. for the translation equivalents of “table”, in English it has no gender, in German it is masculine, in French and Arabic it is feminine).

2.4 Translation of MSA into English

The translation from Modern Standard Arabic into English is done using the above described system using lexical, functional and syntactical features which were used in a LFG based system.

The statistical models are trained on a bilingual sentence aligned corpus for the translation model and the alignment template model. The language models (POS-based and word-based) are being trained on a monolingual corpus.

2.5 Translation of Dialect Arabic into English

Translation of dialect Arabic is implemented using a hybrid MT system that translates the dialect into Modern Standard Arabic (MSA). For the presented translation system, a bilingual corpus is used which consists of sentences in Iraqi dialect (Iraqi Colloquial Arabic, ICA) and Modern Standard Arabic. Also feature functions built out of rules built to translate (or rather: convert) the dialect into non-dialect are used.

As much of the Arabic input can be either MSA or ICA at the same time, the quality of translation can be increased by using dialect feature functions for both the MSA and ICA dialect variants and allow the Generative Iterative Scal-



Figure 2: MediaSphere Screenshot of a Media Content Repository

ing (GIS) algorithm to change the weighting of these features during the training process.

3 Description of Media Monitoring System

MediaSphere is a software solution providing multilingual transcripts of various TV and Radio stations for many domestic and international news bureaus as well as transcripts for conversational telephony. MediaSphere supports video, audio and telephony for text processing with advanced linguistic capabilities such as machine translation of transcribed text, information retrieval with query translation (cross-language information retrieval; XL-IR), automated name and entity recognition (NER) and translation (NET), automatic summarization (SUM) and automatic topic detection (TD).

MediaSphere is a state of the art solution to facilitate the process of generating transcription from TV transmissions and telephony, translating transcribed text into and from English and deliver rich media content online. The multilingual MediaSphere solution utilizes video and audio logging technologies, telephony platforms as well as ASR integrated with MT, XL-IR, NER/NET, SUM and TD.

The solution automatically captures and indexes television, video and audio content in near real-time making it fully searchable and accessible. Once audio is captured and indexed, Knowledge Management tools facilitate intelligent content search capabilities for users.

For the speech recognition of dialectal Arabic speech, the main problem is that it is very difficult to estimate a statistical language model from the very small corpus, especially in comparison to the size of the vocabulary.

In addition to the challenge of dialects, are the nuances of various topical areas of speech that present unique terms specific to those areas or commonly used terms with different meanings

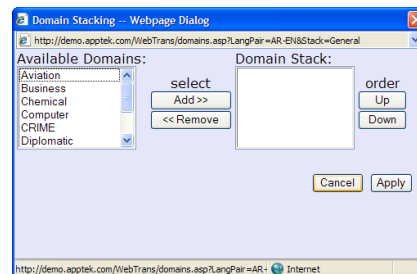


Figure 3: MediaSphere Dialogue for Selection/Stacking of Domains and/or Micro-dictionaries

Table 1: Machine Translation output of AppTek’s Hybrid Machine Translation in comparison to a state-of-the-art 3rd party Statistical Machine Translation

| News & General | AppTek HMT | 3 rd Party SMT | Remarks |
|--|--|--|--|
| مشرف يدعو للمصالحة ويتجاهل مساعي إبعاده عن السلطة | Musharraf calls for reconciliation and ignores the efforts his expulsion from the authority | Musharraf calls for reconciliation efforts and ignores the removal from power | For fluency the SMT removed the pronoun and created ambiguity about who is being removed |
| عملية بالقلب للطالباني والمشهداني بعمان لإجراء فحوص | Process in the heart to Jalal Talabani and Al Mshhdani in Amman for examinations | The heart of Amman, al-Mashhadani, Talabani and for tests | The correct translation is a heart operation in Amman for Talabani and tests for Mashadani |
| ووقع التفجير حوالي الساعة الثامنة والنصف صباحا بالتوقيت المحلي وهو موعد وصول المراجعين إلى دائرة الجوازات. وأسفر عن احتراق خمس سيارات مدنية وإلحاق أضرار مادية بالمباني والمحال التجارية المجاورة. | The explosion occurred about eight thirty a.m. local time, an arrival date of auditors to the circle passports, and resulted in the burning five civil cars and causing material damages to the buildings and neighboring shops. | The explosion occurred about eight o'clock in the morning local time, when the arrival of auditors to the Passport Service, and resulted in the burning of five civilian cars and causing material damage to buildings and shops nearby. | The correct translation of the time is eight thirty (8:30 am) not 8 |
| وذكر المصدر أن شاحنة مفخخة من طراز (كيا) كانت مركونة بالقرب من مرآب للسيارات تابع لدائرة جوازات الأعظمية الواقعة في شارع المغرب انفجرت موقعة 12 قتيلا مدنيا وعشرين جريحا. | The source said a truck bomb of the type (a) was parked near the garage for the cars belonging to the service passports-in the street of morocco exploded killing 12 civilians and twenty injured. | The source said that a truck bomb aircraft (Kia) was near Marconi garage for cars continued to circle passports Alaazemih located in the street exploded Morocco signed 12 civilians were killed and twenty injured. | All words in red are absolutely wrong translations. It is a truck not an aircraft. It is a parking garage not Marconi garage. It is Morocco street not Morocco |

specific to those areas. For example, a news broadcast might report on the year’s flu season and then a malicious attack spread by a hacker. Both topics might use the term “virus” with different meanings. The transcript created by the ASR and the subsequent translation could be inaccurate without the proper understanding of the context of the term “virus” in each instance. The use of domain-specific information resolves this potential problem.

3.1 Domains and Micro-dictionaries

The introduction of special domain dictionaries is readily available. Multiple domain specific on-line dictionaries include, in addition to the general dictionary, the following micro-dictionaries:

- Military;
- Special Operations;
- Mechanical;
- Political & Diplomatic;
- Nuclear;
- Chemical;
- Aviation;
- Computer & Technology;
- Medical;
- Business & Economics;
- Law Enforcement;
- Drug Terms.

4 Examples

The tests were performed on input from broadcast news using the open domain as well as input from the military domain. The experiments show that the hybrid MT performs better for this task than either the purely statistical or purely rule-based MT approaches.

The hybrid machine translation approach introduced here shows a very high accuracy in all categories: fluency, informativeness and adequacy. The approach shows that information units which need to be translated are processed correctly, moreover the output of the translation reads fluently.

5 Conclusion

This paper shows that the proposed Hybrid Machine Translation approach shows better results than a pure rule-based and a pure corpus-based approach for both written and especially for spoken input. It also introduces an approach to increase language model quality for dialect language speech recognition by using non-dialect, non-spontaneous language resources.

Future work will incorporate further integration of other features into the translation process. Also the use of full morphosyntactic analysis will be helpful, as Arabic is highly morphologically inflected. (Nießen, 2002) shows that the utilization of morpho-syntactic analysis promises to have an impact on languages, especially morphologically complex languages like Finnish, Arabic and Hungarian, but also languages like German. In the context of this hybrid approach, the utilization of deep morphology should have an even higher impact, as syntactical analysis is performed within the core translation procedure as opposed to just being preprocessing step for a purely corpus-based translation (e.g. statistical translation).

Acknowledgments

The work and experiments in this paper are partly funded by DARPA under Contract No. HR0011-07-C-0023 and HR0011-08-C-0110. The authors would like to thank Hermann Ney and the

RWTH's ASR team and the AppTek's MT team for fruitful discussions and support.

References

- Bisani, M., H. Ney. 2002. *Investigations on Joint-Multigram Models for Grapheme-to-Phoneme Conversion*. International Conference on Spoken Language Processing (ICSLP), pp. 105–108. Denver, CO.
- Brown, P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, J. Lafferty, R. Mercer, and P. Roosin. 1990. *A Statistical Approach to Machine Translation*. Computational Linguistics, 16, pp. 79–85. Cambridge, MA.
- Brown, P., S. Della Pietra, V. Della Pietra, and R. Mercer. 1993. *The Mathematics of Statistical Machine Translation: Parameter Estimation*. Computational Linguistics, 19(2), pp. 263–311. Cambridge, MA.
- Charniak, E., K. Knight, and K. Yamada. 2003. *Syntax-Based Language Models for Statistical Machine Translation*. In Proceedings of MT Summit IX, 23–27. New Orleans, LA.
- Charniak, E. 2000. *A Maximum-Entropy-Inspired Parser*. In Proceedings of the 2000 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). Seattle, WA.
- Collins, M. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Thesis, Department of Computer and Information Science, University of Pennsylvania. Philadelphia, PA.
- Doyon, J., K. Taylor and J. White. 1999. *Task-Based Evaluation for Machine Translation*. In Proceedings of the Machine Translation Summit VII. Singapore.
- Ney, H., S. Nießen, F. J. Och, H. Sawaf, C. Tillmann, S. Vogel. 2000. *Algorithms for statistical translation of spoken language*. In IEEE Transactions on Speech and Audio Processing. 8(1):24–36. Piscataway, NJ.
- Nießen, S. 2002. *Improving Statistical Machine Translation Using Morpho-syntactic Information*. Thesis, Aachen University of Technology. Aachen, Germany.
- Och, F. J., and H. Ney. 2004. *The Alignment Template Approach to Statistical Machine Translation*. In Computational Linguistics, 30(4), pp. 417–449. Cambridge, MA.
- Sawaf, H., K. Schütz, and H. Ney. 2000. *On the Use of Grammar-Based Language Models for Statistical Ma-*

chine Translation. In Proceedings of the Sixth International Workshop on Parsing Technologies (IWPT), pp. 231–241. Trento, Italy.

Shihadah, M., P. Roohnik. 1998. *Lexical-Functional Grammar as a Computational-Linguistic Underpinning to Arabic Machine Translation*. In Proceedings of the 6th International Conference and Exhibition on Multi-lingual Computing. Cambridge, UK.

Wei, G. 2004. *Phoneme-based Statistical Transliteration of Foreign Names for OOV Problem*. Thesis. Hong Kong, China.