

Linear B System Description for the 2005 NIST MT Evaluation Exercise

Chris Callison-Burch
Linear B Ltd.
39 B Cumberland Street
Edinburgh, EH3 6RA
chris@linearb.co.uk

Abstract

This document describes Linear B's entry for the 2005 NIST MT Evaluation exercise. Linear B examined the efficacy of human-aided statistical machine translation by looking at the improvements that could be had by involving non-Arabic speakers in the translation process. We examined two conditions: one in which non-Arabic speakers edited the output of a statistical machine translation system, and one in which they were allowed to select phrasal translations from a chart of possible translations for an Arabic sentence, and then edit the text.

1 Introduction

With a few exceptions, the primary objective of the computational linguistics community's research into translation has been the creation of fully automatic machine translation systems. Some projects have been directed at developing aids for bilingual human translators (Kay, 1980; Foster et al., 2002; Och et al., 2003; Civera et al., 2004). Here we are instead interested in ways of improving translation with *monolingual* human intervention, specifically through involving speakers of the target language.

In this paper we

- Describe our experimental setup which tested two different methods for allowing monolingual English speakers to improve the translation of Arabic documents;

- Show example translations produced with human intervention, compare them to fully-automatic machine translations;
- List the questions that we intend to investigate when analyzing our translations, their evaluations, and the data that we collected when creating them.

2 Experimental Design

We translated each of the 100 newswire articles in the Arabic-English track in one of two ways:

- By simply post-editing the output of a fully-automatic machine translation system
- By visualizing the most probable phrasal translations for an Arabic sentence, manually choosing among them, and then post-editing the composite results.

Each article was presented to the human subjects as a whole ordered document. They were instructed that the first sentence would be a headline, and that the next sentence would contain a byline that indicated the city, the date, and the news agency. They were instructed to try not to reduce the lengths of the translation (although this was not enforced), and that if something was unclear, that they should try to use the other sentences to guess at what it could be. They were further instructed that if a name were untranslated they could try to look it up using Google.

Each article was presented to subjects on a web page where each of the sentences was given its own text box. In the simple post-editing condition the automatic translation was put in the text box, and

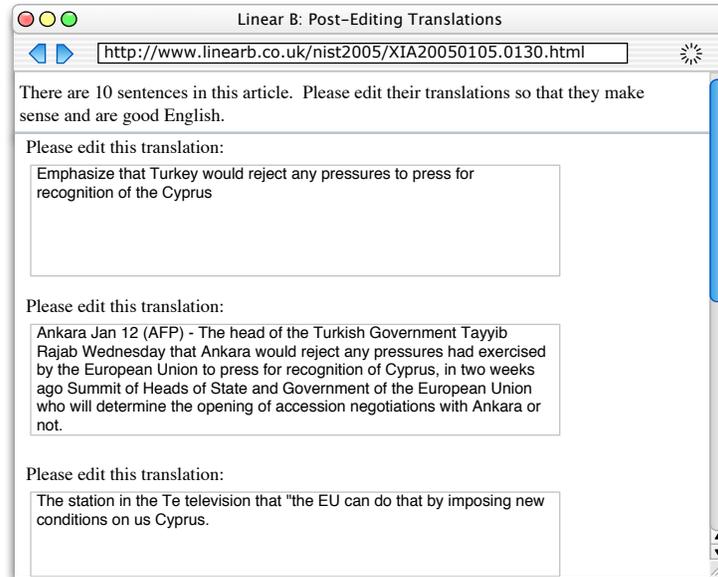


Figure 1: In the simple editing condition subjects simply edited the output of a fully-automatic statistical machine translation system

directly edited (shown in Figure 1). In the visualization condition the text box was initially empty, and the subjects created the text by selecting phrasal translations from a chart (shown in Figure 2).

In both cases timing information was collected, and the editing was logged so that we can analyze how many changes were made. In the visualization condition we logged additional information about which phrasal translations were selected, including their span and their rank.

2.1 Baseline

The automatic translations that were manually corrected in the simple editing condition were produced for Edinburgh University's entry for this year's NIST MT Evaluation exercise. We choose to use Edinburgh's entry as a starting point for two reasons:

1. Philipp Koehn's system, which the Edinburgh entry is based on, provided state-of-the-art results during last year's NIST MT Eval (Koehn, 2004a; Przybocki, 2004), and thus represents a strong baseline for improvement.
2. Edinburgh's entry will be evaluated separately in this year's exercise, and thus the scores for Linear B's human-aided entry can be directly

compared to its scores.

The visualizations of the phrasal translations were created using Linear B's suffix array-based data structure (Callison-Burch et al., 2005b), which allows us to load the entire Arabic-English data set into memory and look up phrase alignments on-the-fly. We used Giza++ (Och and Ney, 2003) created the word-alignments for the data set for last year's Arabic-English large data track, and extracted the phrases using the phrase extraction algorithm (Koehn, 2004b) used in the Edinburgh entry.

3 Example Translations

Figures 3 and 4 give a side by side comparison of the translations created by Linear B's human-aided method, and Edinburgh's fully automatic machine translation. Figure 3 was done using the visualization tool. Figure 4 was started using the simple editing method, but was switched to the visualization method because the subject complained that "The Group Include Blue Whiting" was impossible to decipher.

On the whole, subjects doing the visualization method found it to be time consuming, but felt more confident in the translations that they pro-

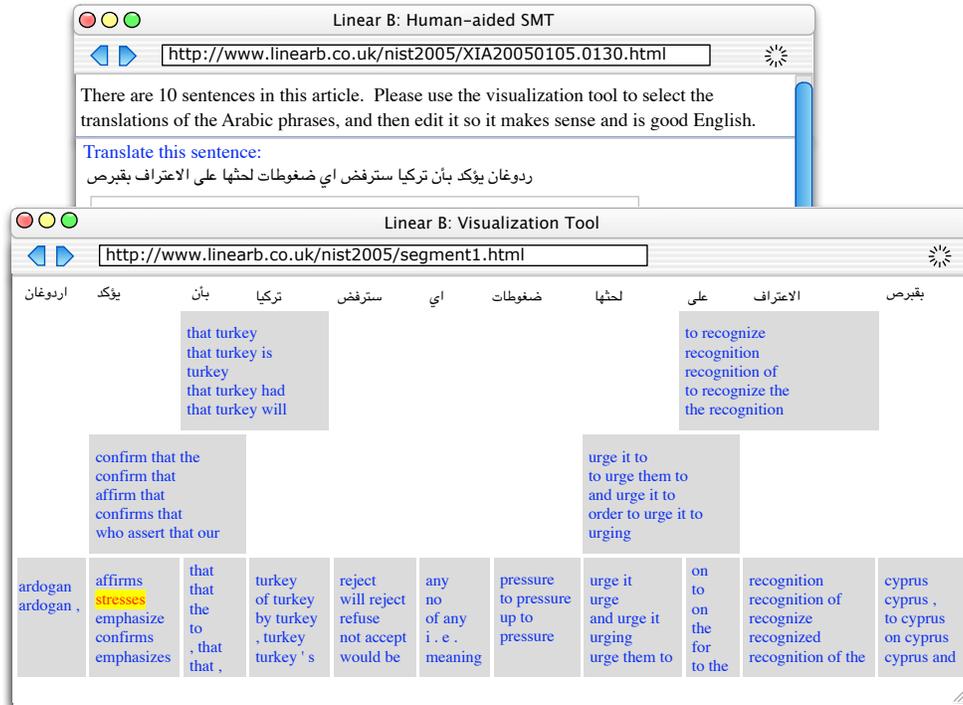


Figure 2: In the visualization condition subjects first constructed the translations by selecting from a set of probable translations of the phrases in each Arabic sentence

duced, since they were selecting from a set of possible alternatives.

4 Questions for Analysis

We plan to analyze the data that we collected to answer the following questions:

- Is the quality of the translations produced using the visualization technique any different than the quality of the translations produced by simply editing the output?
- What is the relative gain in translation quality that can be achieved using monolingual human intervention when compared to fully automatic machine translation?
- How long does it take on average to do translations in this fashion?
- How does this compare to the time it takes for a bilingual individual to do the translation from scratch?

- What is the average length and rank of the phrases selected by a human from the chart used for visualization?
- How does this compare to the average length and rank of the phrases selected by the decoder from its chart?
- How much editing happens on average?
- Can we generalize the things that are fixed through human editing, and incorporate them into our fully automatic systems?

We are further interested in comparing relative ranking of our entry when scored using the Bleu evaluation metric against ranking of our entry when scored in the human evaluation.

5 Related Work

There are two papers from Linear B that describe other ways of integrating people into the translation process. Callison-Burch et al. (2004) describes a framework for using post-edited output to improve

Human-aided MT	Fully automatic MT
Erdogan Confirms That Turkey Will Reject Any Pressure To Urge It To Recognize Cyprus	Emphasize that Turkey Would Reject Any Pressures to Press for Recognition of the Cyprus
Ankara 1-12 (AFP) The president of the Turkish government Recep Tayyip Erdogan said today that Ankara will reject any pressure from the European Union that urges it to recognize Cyprus before the two week summit with the heads of state of the governments of the European Union who will decide whether or not to open negotiations for Ankara joining the EU.	Ankara Jan 12 (AFP) - The head of the Turkish Government Tayyip Rajab Wednesday that Ankara would reject any pressures had exercised by the European Union to press for recognition of Cyprus, in two weeks ago Summit of Heads of State and Government of the European Union who will determine the opening of accession negotiations with Ankara or not.
Erdogan told the "NTV" television station that "the European Union cannot begin addressing us by imposing new conditions about Cyprus.	The station in the Te television that "the EU can do that by imposing new conditions on us Cyprus."
We shall be discussing this issue during the negotiations for joining the EU".	We will discuss this issue during accession negotiations.
He added: "Let me to be clear from the start that I must act as the arm of Turkey and this is an issue that cannot be accept by it".	He added: "Let me be clear, may I arm of Turkey, this is cannot accept it."
It is expected that on December 14 the European Parliament and heads of state will recommend negotiations with Ankara for joining the EU but with specific conditions.	It is expected to recommend to the European Parliament on December 14 Heads of State and Government of the approval of the opening of accession negotiations with Ankara, but with conditions.
Greece considers that the recognition by Ankara of Cyprus, which itself has become a member of the European Union to be a "self-evident condition" to grant before the European Union will open accession negotiations with Ankara.	The Greece's recognition of Cyprus, which has Ankara is also a member of the European Union, requirement of self-evident to give the European Union green to open accession negotiations with Ankara.
Still Turkey refused to recognize the Republic of Cyprus which became a member of the European Union on the first of May last year, but it is the only state in the world which has recognized the Turkish Republic of North Cyprus which declared its independence, and is self-governance in the North of the island.	Turkey still refuses to recognize the Republic of Cyprus, which has become a member of the European Union since the beginning of May, but the only State in the world acknowledged the "B" Turkish Republic of Northern Cyprus, declared independence self-government in the north of the island.
Erdogan also rejected the possibility of open negotiations to establish a distinct partnership between Turkey and the European Union as an alternative to negotiating accession.	He also rejected premise reopen negotiations to distinctive partnership between Turkey and the European Union as an alternative to accession negotiations.
Erdogan said, "This proposal cannot be accepted and that it is incompatible with the seriousness of an organization such as the European Union."	The in this subject this proposal could not accept that, in line with the seriousness of such as the European Union.

Figure 3: A comparison of Linear B's human-aided and Edinburgh University's fully automatic translation for article AFP20041201.0189

Human-aided MT	Fully automatic MT
Mussab Group Claims Credit For The Assassination Of The President Of The Diyala Provincial Assembly And His Brother	The Group Include Blue Whiting by the Assassination of the Council Governorate and His Brother
Abu Dhabi - January 19 / Xinhua / The organization “Tanzim Qa’idat Al-Jihad in Bilad al-Rafidayn” under the leadership Abu Musab al-Zarqawi announced on Saturday its responsibility for the killing of the the president of Diyala provincial assembly and brother in the city, according a statement that was released in a declaration on the web site of organization.	Abu Dhabi January 1 / Xinhua / The organization of Al-Jihad Base in the country “led by Abu include blue whiting on Saturday claimed responsibility for the killing of the Diyala province, and his brother in the city of death, as stated in a statement broadcast on the Organization on the Internet site.”
The statement said “the military wing of Tanzim Qa’idat Al-Jihad in Bilad al-Rafidayn has assassinated of one of the criminals in the city of Baquba”.	The statement said, “in the military wing of Al-Jihad Base in the country have assassination of one of criminals in the capital city.”
A medical source had announced on Saturday that gunmen had murdered president provincial assembly of Diyala and his brother.	It was a medical source has announced that they killed gunmen on Saturday, President of the governorate and his brother.
A doctor at the hospital told reporters that “Al-husayn arrived in the hospital after his death, while his brother died while they were attempting to save his life”.	He said in a hospital doctor death journalists that “Abd al-Husayn arrived in hospital after his death while his brother died during attempts to.”
An official source for the Iraqi police said that the president provincial assembly of Diyala in the east of the Iraq was subjected to attack on Saturday evening in the exclusion zone of the city when he was fired upon by persons unknown who shot him dead. He was immediately killed. The source pointed out that brother, who was accompanying him was also seriously injured was transferred to the hospital in the city.	He said an official source in the Iraqi police said the President of the governorate Iraq on Saturday evening attack in the area of the liberalization of capital, where they were shot dead by unidentified in the case, saying that the brother of the, who was accompanying him was also seriously injured transferred to hospital to death.
He had taken up this post several months earlier after the resignation of the outgoing president who had also been subject to an assassination attempt.	The post take several months ago after resigned Hisham of the former for over an assassination attempt.
In the latest statement on its web site, the organization announced on Saturday that they claimed responsibility for a car bomb in a Baiji refinery to 200 kilometers north of Baghdad.	The latest in a statement on the web site, reorganization announced on Saturday claimed responsibility for a car bomb in Baiji refinery to 200 kilometres north of Baghdad.
The statement said that “a battalion of the military wing of Tanzim Qa’idat Al-Jihad set off a car bomb in the Baiji region.”	The statement said that “of battalion in the military wing of Al-Jihad Base in the country sailed detonated a car bomb in the Baiji.”
The car bomb exploded on Friday near the Iraqi national guard patrol in Baiji, and caused the deaths of a number of civilians and wounded six members of the national guard, according to a spokesman on behalf of the U.S. army.	The car bomb exploded Friday near the Iraqi National Guard patrol in Baiji, which led to the killing of civilians and wounded six members of the National Guard, in accordance with on behalf of the US Army.

Figure 4: A comparison for article XIA20050101.0119

the quality of automatic translation. We plan to test this framework with our NIST MT Eval 2005 data to see whether post-editing only the first few sentences in an article could be used to improve the automatic translation of the rest of the article. Callison-Burch et al. (2005a) describes our “searchable translation memories” that rely on the same technology that we used to create the visualization, but go further and show the context of each possible translation of a phrase.

Acknowledgments

Thanks go to Craig Baracco, Trevor Cohn, Dave Hawkey, Sarah Luger, Danielle Matthews, Brian Milch, Matt Prowse, Phil Shilane, Andrew Smith, David Talbot, Maarika Traat, and Valerie Wiersma for helping out with the experiments, and inexpressible gratitude goes to Colin Bannard, Alexandra Birch Mayne, Alexa Callison-Burch, and Annabel Harrison for their exceptional efforts. You guys rock.

References

- Chris Callison-Burch, Colin Bannard, and Josh Schroeder. 2004. Improved statistical translation through editing. In *European Association for Machine Translation*.
- Chris Callison-Burch, Colin Bannard, and Josh Schroeder. 2005a. A compact data structure for searchable translation memories. In *European Association for Machine Translation*.
- Chris Callison-Burch, Colin Bannard, and Josh Schroeder. 2005b. Scaling phrase-based statistical machine translation to larger corpora and longer phrases. In *Proceedings of ACL*.
- Jorge Civera, Elsa Cubel, Antonio L. Lagarda, David Pic, Jorge Gonzalez, Enrique Vidal, Francisco Casacuberta, Juan M. Vilar, and Sergio Barrachina. 2004. From machine translation to computer assisted translation using finite-state models. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP04)*.
- George Foster, Philippe Langlais, Elliott Macklovitch, and Guy Lapalme. 2002. Transtype: Text prediction for translators. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Demonstration Description.
- Martin Kay. 1980. The proper place of men and machines in language translation. Technical Report CSL-80-1, Xerox PARC. Reprinted in *Machine Translation* 12(1):3–23, 1997.
- Philipp Koehn. 2004a. The foundation for statistical machine translation at MIT. In *Proceedings of Machine Translation Evaluation Workshop 2004*.
- Philipp Koehn. 2004b. Pharaoh: A beam search decoder for phrase-based statistical machine translation models. In *Proceedings of AMTA*.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, March.
- Franz Och, Richard Zens, and Hermann Ney. 2003. Efficient search for interactive statistical machine translation. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 387–393.
- Mark Przybocki. 2004. NIST 2004 machine translation evaluation results. Confidential e-mail to workshop participants, May.