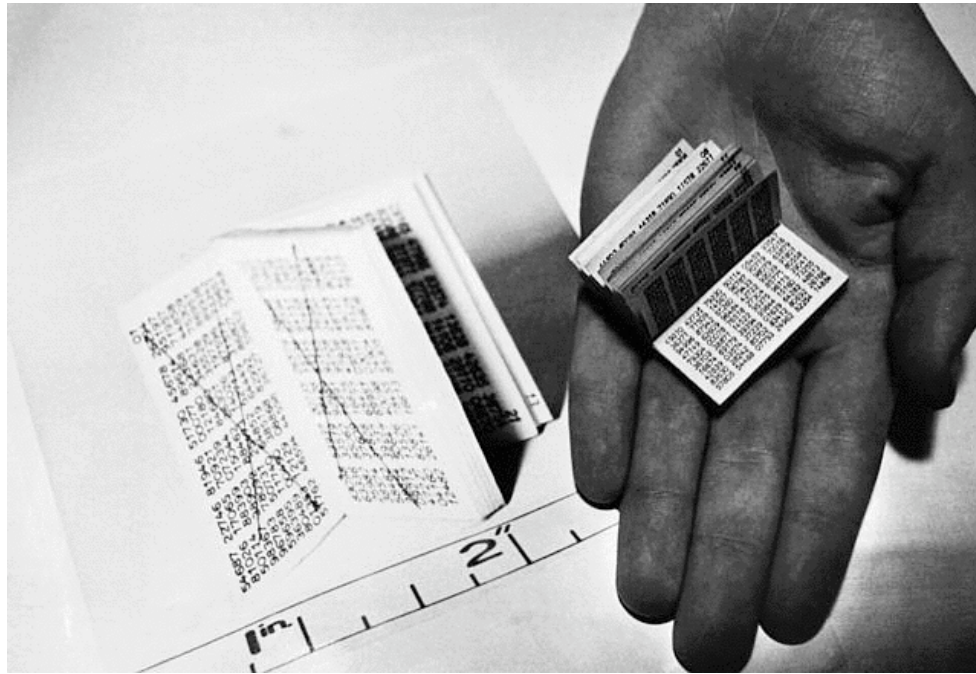


Human-aided Statistical Machine Translation



Chris Callison-Burch
LINEAR B LTD

Linear B

- Edinburgh based startup company
- "LanguageWeaver of Europe"
- Fully automatic statistical machine translation systems
- Tools for human translators

Human-aided SMT

- Most research focuses on fully automatic MT
- Some on aids for *bilingual* human translators
- Here we examine *monolingual* human involvement
- How well can English speakers translate Arabic?

Experimental Design

- Monolingual English speakers / non-Arabic speakers to improve translations of 100 Arabic documents translated into English
- Conditions:
 - 1) Simple post-editing of FAMT
 - 2) Visualization, selection, then editing

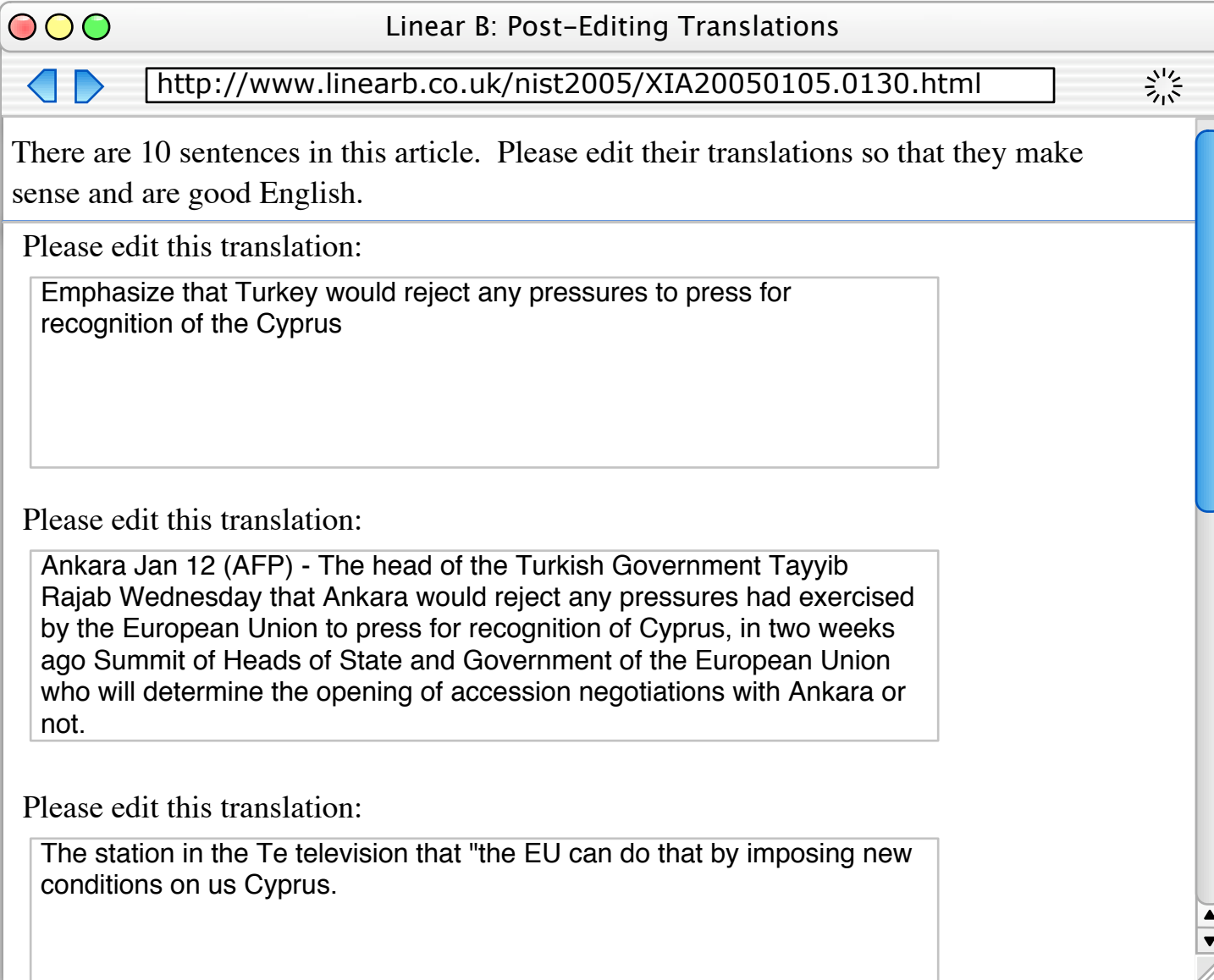
Experimental Design

- Baseline: Edinburgh's FAMT system
- Strong baseline, will be evaluated separately, so direct comparisons can be made
- Data collection:
 - Timing information
 - Complete editing information
 - Selection info (length of phrase, rank, etc.)

Experimental Design

- Subjects: 17 grad students, most native English speakers
- Instructions:
 - 1) Correct English
 - 2) Don't decrease length of translation
 - 3) Use context to help
 - 4) Look up names on Google

Simple post editing



Linear B: Post-Editing Translations

<http://www.linearb.co.uk/nist2005/XIA20050105.0130.html>

There are 10 sentences in this article. Please edit their translations so that they make sense and are good English.

Please edit this translation:

Emphasize that Turkey would reject any pressures to press for recognition of the Cyprus

Please edit this translation:

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.

Please edit this translation:

The station in the Te television that "the EU can do that by imposing new conditions on us Cyprus.

Visualization

Linear B: Visualization Tool

<http://www.linearb.co.uk/nist2005/segment1.html>

اردوغان	يؤكد	بأن	تركيا	سترفض	اي	ضغوطات	لحثها	على	الاعتراف	بقبرص
		that turkey that turkey is turkey that turkey had that turkey will						to recognize recognition recognition of to recognize the the recognition		
	confirm that the confirm that affirm that confirms that who assert that our						urge it to to urge them to and urge it to order to urge it to urging			
ardogan ardogan ,	affirms stresses emphasize confirms emphasizes	that that the to , that that ,	turkey of turkey by turkey , turkey turkey ' s	reject will reject refuse not accept would be	any no of any i . e . meaning	pressure to pressure up to pressure	urge it urge and urge it urging urge them to	on to on the for to the	recognition recognition of recognize recognized recognition of the	cyprus cyprus , to cyprus on cyprus cyprus and

Question to be investigated

- How much improvement in translation quality do we get?
- Can we identify things to improve FAMT?
- How long does it take monolinguals to translate?

Improvement Quantified

- Edinburgh's FAMT = 3.97 Bleu
- Linear B's HAMT = 4.3 Bleu
- Rank among other entries: 6th place
- Expect higher rank in human eval
- *Is Bleu adequate?*

Example Translation

FAMT

Emphasize that Turkey Would Reject Any Pressures to Press for Recognition of the Cyprus

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.

The station in the Te television that "the EU can do that by imposing new conditions on us Cyprus." We will discuss this issue during accession negotiations.

He added: "Let me be clear, may I arm of Turkey, this is cannot accept it."

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.

Example Translation

HAMT

Erdogan Confirms That Turkey Will Reject Any Pressure To Urge It To Recognize 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.

Erdogan told the "NTV" television station that "the European Union cannot begin addressing us by imposing new conditions about Cyprus. We shall be discussing this issue during the negotiations for joining the EU".

He added: "Let me 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".

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.

Example Translation Reference

Erdogan Confirms That Turkey Will Reject Any Pressures To Encourage It To Recognize Cyprus

Ankara 12-1 (AFP) - The Turkish head of government, Recep Tayyip Erdogan, announced today (Wednesday) that Ankara would resist any pressure the European Union might exercise in order to force it into recognizing Cyprus. This comes two weeks before a summit of European Union heads of state and government, who will decide whether or not to open membership negotiations with Ankara.

Erdogan said to the NTV television channel: "The European Union cannot engage with us through imposing new conditions on us with regard to Cyprus. We shall discuss this issue in the course of the membership negotiations."

He added: "Let me be clear - I cannot confine Turkey. This is something we do not accept."

It is expected that on the 14th December the European Parliament will recommend that the heads of state and government of the European Union agree to open membership negotiations with Ankara, but on clearly defined conditions.

Analysis

- Very good translation, close to reference
- But contains 2 factual errors:
 - Not "president" but "prime minister"
 - "two week summit"
- Poor word choice in "arm of Turkey"

Observations for FAMT

- Low hanging fruit:
- Names (easy to find with titles)
- Deleted / inserted words
- Better strategies for OOV items

Questions for further analysis

- Difference between simple editing and viz?
- Avg length and rank of selected phrases?
- How much editing happens on avg?
- Cost v. bilingual translation?

Ways of integrating human knowledge

- Incorporate changes into translation model
- Searchable translation memories
- Bilingual involvement: correct misalignments
- Use corrected alignments in training

Conclusions / Broader questions

- Do we want it fast, or good?
- If we have labor available, how do we use it?
- SMT buzzwords: fully automatic, language independent, knowledge lean, low overhead
- How about:
 - easy to use?
 - simple to identify errors?
 - straightforward to improve?

Thank you!

Related Publications

- ACL-05: Scaling SMT to longer phrases and larger corpora
- ACL-04: SMT with word- and sentence-aligned parallel corpora
- EAMT-05: Searchable translation memories (demo at <http://linearb.co.uk/>)
- EAMT-04: Improved SMT through post-editing