Searchable Translation Memories

Chris Callison-Burch
LINEAR B
What's wrong with translation memories?

• A very valuable tool:
  - Productivity gains
  - Increased consistency
  - Simplified quality control
  - Improved terminology management

• However, limited to sentence / term retrieval
The Future of TMs

- "Certainly there are improvements that can be made to current TM applications and their underlying technologies ... the next new wave of language technology innovation will come from extending the recycling of translated material to different levels of granularity, and to combine them. Today we mainly recycle on the term level, sentence level, or document level. In the future, we want to extend this to recycling on the phrase level. This will require its own smart linguistic algorithms."

(Trados CTO, article in Clientside News)
Searchable Translation Memories

• Tools that allow Google-style searching of translation memories.

• Find past translations which are not normally retrieved in current TMs

• Better exploit knowledge in TM. Further increase consistency, improved productivity, etc.
Talk Overview

• Motivation

• Linear B's Statistical Machine Translation

• Technology for Searchable Translation Memories

• Demo

• Summary

• Sneak Peek
Linear B: Who we are

• Linear B was founded in October 2002 by Informatics PhD students at the University of Edinburgh

• Funded by venture capital and a SMART technology grant from UK government

• Linear B creates statistical machine translation software
Statistical Machine Translation

- Data-driven - Automatically learns to translate from example translations
- Eliminates the need for a staff of linguists to hand-craft rules
- Can be applied to any language pair
what is more, the relevant cost dynamic is completely under control.

sooner or later we will have to be sufficiently progressive in terms of own resources as a basis for this fair tax system.

we plan to submit the first accession partnership in the autumn of this year.

it is a question of equality and solidarity.

the recommendation for the year 1999 has been formulated at a time of favourable developments and optimistic prospects for the European economy.

that does not, however, detract from the deep appreciation which we have for this report.

im übrigen ist die diesbezügliche kostenentwicklung völlig unter kontrolle.

früher oder später müssen wir die notwendige progressivität der eigenmittel als grundlage dieses gerechten steuersystems zur sprache bringen.

wir planen, die erste beitrittspartnerschaft im herbst dieses jahres vorzulegen.

hier geht es um gleichberechtigung und solidarität.

die empfehlung für das Jahr 1999 wurde vor dem hintergrund günstiger entwicklungen und einer für den kurs der europäischen wirtschaft positiven perspektive abgegeben.

im übrigen tut das unserer hohen wertschätzung für den vorliegenden bericht keinen abbruch.
What is learned

<table>
<thead>
<tr>
<th>im</th>
<th>übrigen</th>
<th>ist</th>
<th>die</th>
<th>diesbezügliche</th>
<th>kostenentwicklung</th>
<th>völlig</th>
<th>unter</th>
<th>kontrolle</th>
</tr>
</thead>
<tbody>
<tr>
<td>what</td>
<td>is</td>
<td>more</td>
<td>the</td>
<td>relative</td>
<td>cost</td>
<td>dynamic</td>
<td>is</td>
<td>completely</td>
</tr>
</tbody>
</table>

is completely under control
Advantages of SMT

- Can be applied to any language pair
- Quick to develop
- Improves as more data becomes available
- (Recently) High-quality
Honorables sénateurs, tandis que la guerre en Irak entre dans sa troisième semaine, nous ne devons pas oublier qu'il faut prendre des mesures pour éviter une crise humanitaire dans la population civile. À cet égard, il y a de nombreux domaines dans lesquels les Canadiens doivent faire preuve de leadership.

Maintenant que la guerre fait rage en Irak et que des pénuries de produits alimentaires et de fournitures médicales commencent à se produire, les pays qui ont accès à des ressources ont la responsabilité d'essayer de minimiser les effets négatifs du conflit sur la population irakienne. Je crois personnellement que le Canada devrait jouer un plus grand rôle à cet égard.

Au Canada, nous disposons en abondance de blé et d'autres produits alimentaires. Nous pouvons également fournir des articles médicaux aux citoyens de l'Irak. Le Canada a une fière réputation pour ce qui est d'offrir une aide humanitaire aux gens quand ils en ont besoin.
Honourable senators, while the war in Iraq between in his third week, we must not forget that it must take of measures to avoid a crisis humanitarian in the people calendar. In many areas which Canadians must show leadership in this regard.

Now that the war fact raging in Iraq and that of shortages of products food and of supplies medical beginning to be produce, the country which have access to of resources have the responsibility of try of minimize the effects negative of conflict on the people irakienne. I personally believe this regard great role Canada should play more.

In other food products wheat Canada have abundant We can provide the citizens medical articles Iraq The offer Canada has a reputation proud.
Honourable senators, while the war in Iraq extending into a third week, the minister should not forget the one we must take some steps to prevent a crisis humanitarian face in the civilian populations. In this regard, there are a number of areas in which the Canadians must concentrate to show leadership.

That the war is raging in Iraq and that a shortage of food and medical supplies are beginning to take place, those countries that have access to the resources are the responsibility for doing try to minimize the negative effects on workers on the people Irakienne. I personally believe that Canada should play a more significant role in this regard.

In Canada, we have in abundance of wheat, as have other food products. We can also provision of medical articles to the people on the other Iraq. Canada has a proud record in so far, as would be to give humanitarian assistance to people when they need it.
| Je crois personnellement que le Canada devrait jouer un plus grand rôle à cet égard. |
| I personally believe that Canada should play a more significant role in this regard. |
Synergy Between SMT and TMs

- Quality of statistical MT improves when large TMs are available
- Better chance of retrieving larger blocks of human translated text
- Techniques used in statistical MT can be used to align phrases within TMs
- Might be used to increase usefulness of TMs
Statistical MT to Searchable TMs

• The technology allows us to build an index of phrasal translations

• Like a concordance, but also lets us identify the translation of a phrase
<table>
<thead>
<tr>
<th>im übrigen ist die diesbezügliche kostenentwicklung völlig unter kontrolle</th>
<th>what is more the relative cost dynamic is completely under control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kostenentwicklung</td>
<td>what is more the relative cost dynamic is completely under control</td>
</tr>
<tr>
<td>relational</td>
<td>what is more the relative cost dynamic is completely under control</td>
</tr>
<tr>
<td>dynamic</td>
<td>what is more the relative cost dynamic is completely under control</td>
</tr>
<tr>
<td>is</td>
<td>what is more the relative cost dynamic is completely under control</td>
</tr>
<tr>
<td>completely</td>
<td>what is more the relative cost dynamic is completely under control</td>
</tr>
<tr>
<td>control</td>
<td>what is more the relative cost dynamic is completely under control</td>
</tr>
</tbody>
</table>
Index of Phrases 2

wir sind es den steuerzahlern schuldig die Kosten unter Kontrolle zu haben we owe it to the taxpayers to keep the costs in check
Highlighting Translations

• An index of phrasal translation allows us to highlight the translation that someone is looking for:

Search for translations of *paiement initial*

<table>
<thead>
<tr>
<th>Down Payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ils n’ont pas les quelques milliers de dollars nécessaires pour</td>
</tr>
<tr>
<td>le <em>paiement initial</em> afin d’acheter leur propre maison.</td>
</tr>
<tr>
<td>They do not have the several thousand dollars needed for</td>
</tr>
<tr>
<td>a <em>down payment</em> on their own shelter.</td>
</tr>
</tbody>
</table>
Ranking Translations

• Knowing how many times a translation occurred in the index lets us rank results by frequency:

<table>
<thead>
<tr>
<th>Search for translations of paiement initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Payment - 15 sentences matched</td>
</tr>
<tr>
<td>Down Payment - 8 sentences matched</td>
</tr>
<tr>
<td>Initial Payments - 1 sentence matched</td>
</tr>
</tbody>
</table>
Demo
Evaluation

• Created a "gold standard" set of correct translations for 120 phrases that occurred in a translation memory containing 50,000 German-English sentences

• Evaluated the correctness of the suggestions that our software presents, and how many of the correct translations it managed to find
Results

• 78% of phrases were correct

• 81% of the "gold standard" translation were retrieved

• The top translation that our system returned for a phrase was correct 87% of the time
Conclusion

• Presented Linear B's first foray into aides to the human translation process

• Technology developed for statistical machine translation can be used to create searchable translation memories

• Allows us to better exploit the knowledge contained within TMs
Future Directions

• Preview of Linear B's coming technology:
  - Improved machine translation through post-editing
  - Human-aided machine translation
Improved Translation Through Post-editing
1. Please send a proofreading manuscript by the E-mail / Fax / mailing.  
2. I announce a visitor an estimate and time for delivery.

Linear B Machine Translation

they post-edit

Manuscript sent to client

Parallel text sent to Linear B for retraining
Human-aided Statistical Machine Translation
after ten (0.75)
detonated (0.25)
days of (1)
visit (0.7143)
his visit (0.1429)
visit had (0.0714)
visit undertaken (0.0714)
foreign minister (1)
Ahmad Maher (0.4286)
Ahmed Maher (0.4286)
Ahmad Mahir (0.1429)

visit after (1)
ten days (0.7222)
10 days (0.2778)
from visiting (0.8333)
from seeing (0.0667)
undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)
minister (0.8796)
Foreign Minister (1)
Ahmad (1)

comes (0.3636)
coincides (0.1364)
is (0.1364)

after (0.45)
post (0.1096)
yet (0.0829)
ten (0.6391)
10 (0.2561)
days (0.7595)
day (0.1393)
of (0.642)
from (0.3076)
visit (0.7583)
visits (0.0554)
, (0.238)
has (0.0686)
by (0.0657)
by (0.3873)
undertaken (0.0901)
minister (0.8273)
Foreign (0.5241)
external (0.3355)
egypt (0.7153)
egyptian (0.0945)
Ahmad (0.5823)
Ahmed (0.3844)
Mahir (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)

Maher Ahmed (0.5823)
Ahmad (0.5387)
Mahir (0.2232)
Selection

ten days after (1)

after ten (0.75)
detonated (0.25)

days of (1)

visit (0.7143)
his visit (0.1429)
visit had (0.0714)
visit undertaken (0.0714)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)

foreign minister (0.8273)

egyptian foreign (1)

ahmad maher (0.4286)
ahmed maher (0.4286)
ahmad mahir (0.1429)

comes (0.3636)
his (0.1454)
post (0.1096)
yet (0.0829)

ten (0.6391)
days of (0.1393)
of (0.642)

visit (0.7583)
visits (0.0554)

by (0.3873)
undertaken (0.0901)

minister (0.8273)

foreign (0.5241)
external (0.3355)
egypt (0.7153)

egyptian (0.0945)
ahmad (0.5823)
ahmed (0.3844)
mahir (0.5387)

coincides (0.1364)
is (0.1364)

Maher Ahmad al-Musri al-Kharji

Selection

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)

visit after (1)

ten days (0.7222)
10 days (0.2778)

from visiting (0.8333)
from seeing (0.0667)

undertaken (0.273)
carried out by (0.1546)
made by (0.0987)
conducted by (.07)

foreign minister (0.8796)

egyptian foreign minister (1)

egyptian foreign minister ahmad maher (1)
comes (0.3636) visit (0.7285) his (0.1454)

egyptian foreign minister ahmad maher (1)

ten days after (1) visit (0.958)
of (0.642) the

Manual re-ordering

Maher Ahmed Al-Masri, the Egyptian foreign minister, arrived ten days after his visit, completing ten days after his visit to the country.
Ramallah (West Bank) 1/1 (AFP) An Agence France-Presse correspondent said that the Egyptian presidential advisor Osama Baz arrived on Thursday in Ramallah for talks with Palestinian authority president Yasser Arafat. Minister Hun and Authority Chairman Katif participated in talks and negotiations with Interior Minister Saib Ariqat. Baz refused to make statements immediately upon arrival to the province. *His visit comes ten days after the visit of Egyptian foreign minister Ahmed Maher.* Baz began his visit to Ramallah last August with a new truce idea for Palestinians and Israel, however Palestinian factions have not agreed to this Egyptian sponsored proposal. It is expected that Egyptian endeavours will continue next week with a planned visit by the head of general intelligence.