DIPLOMAT

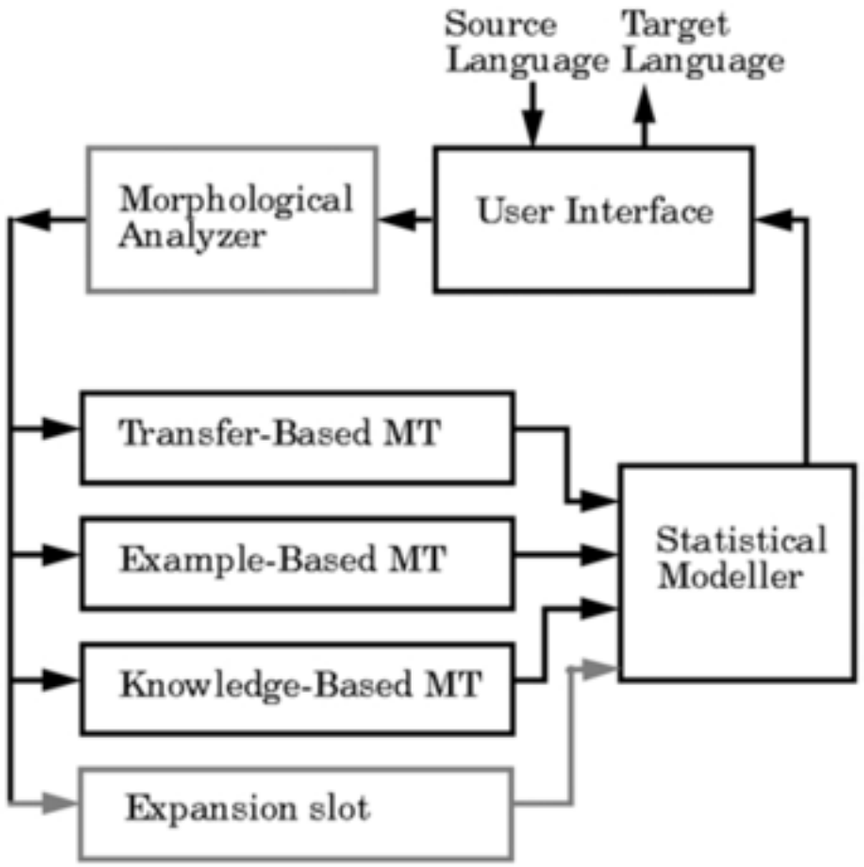
A Speech-to-Speech MEMT System

Introduction

- Its underlying translation system is actually Pangloss
  - EBMT
  - KBMT
  - LTMT
- Pangloss is the kernel and Diplomat is the shell
- On top of the translation system, Diplomat presents a user interface and a speech recognition system which in turn feeds the analyzed speech as input text to its translation component.

MEMT

- Exploit the differences between MT technologies
- Each engine attempts to translate the entire input text
- Each segment is given a score indicating the engine's internal assessment of the quality of the output segment.

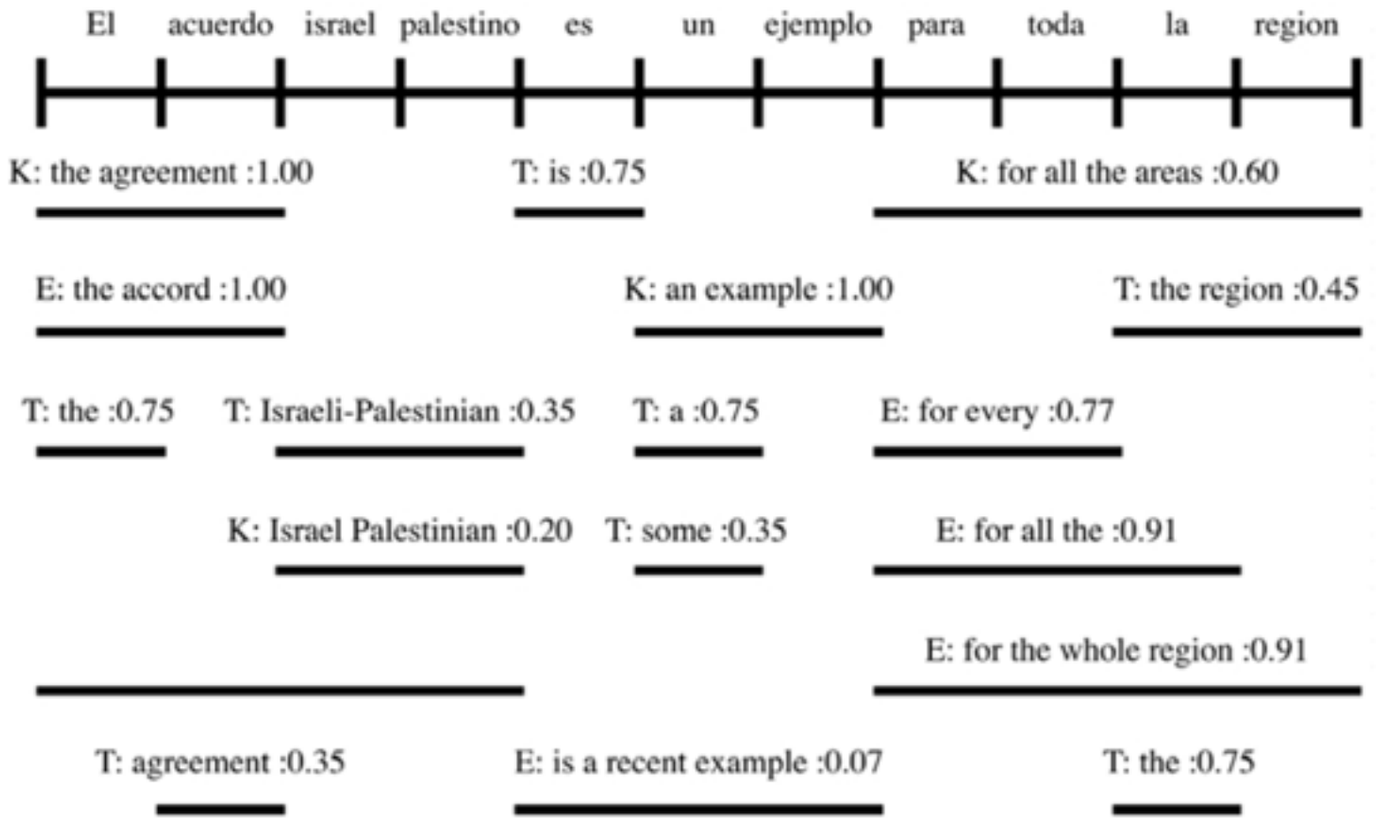

Source Language
Target Language
Morphological Analyzer
User Interface
Transfer-Based MT
Example-Based MT
Knowledge-Based MT
Expansion slot
Statistical Modeller

Translation Chart

- The chart contains multiple, possibly overlapping, alternative translations

- Use statistical language modeling techniques adapted from speech recognition research to select the best overall set of outputs

- Taking the probability of transitions between segments into account as well as modifying the quality scores of individual segments.


El acuerdo israel palestino es un ejemplo para toda la region
K: the agreement :1.00
T: is :0.75
K: for all the areas :0.60
E: the accord :1.00
K: an example :1.00
T: the region :0.45
T: the :0.75
T: Israeli-Palestinian :0.35
T: a :0.75
E: for every :0.77
K: Israel Palestinian :0.20
T: some :0.35
E: for all the :0.91
E: for the whole region :0.91
T: agreement :0.35
E: is a recent example :0.07
T: the :0.75

Scoring

- The KBMT and EBMT engines provide a score for each output element
- For the Lexical Transfer Engine, the score for each glossary is a constant based on the reliability of the glossary.
- The scores are also normalized so as to be comparable
- Finally, during chart construction, the base score produced by the scoring functions is multiplied by the length of the candidate in words, on the assumption that longer items are better.

Chart Walk

- Once the edges are scored, the cover is produced using a simple dynamic programming algorithm.

- The chart-walk algorithm used to produce a single, best, non-overlapping, contiguous combination of the available component translations.

- A recursive divide-and-conquer procedure

  - For each position within a segment, the sentence is split into two parts

  - The best possible cover for each part is recursively found and the two scores are combined to give a score for the chart-walk containing the two best subwalks.

Pangloss

1. Text input via standard input or sockets

2. Morphological analysis

3. Translation: results of morphological analysis passed to each MT engine; scored outputs placed into chart

4. Language modeller selects "best" edges, and adds results to chart

5. Output: either text composed of "best" edges or entire chart

EBMT

1. align corpus at sentence level

2. find chunks from the source language part of corpus which are best candidates for matching an input chunk (intra-language matching)

3. find the target language chunk corresponding to the chunk from the source language part of the corpus (inter-language matching)

4. combine chunk-level results to obtain the "cover" for the entire text

EBMT Extension

- Diplomat extends the EBMT approach in several ways, the most important being word-class substitution which permits words within certain classes to be substituted for one another.

EBMT

- *The government gave no food supplies to Ethiopia for the period January 1 to December 31,1972.*

- *Le gouvernement na pas envoy de vivres en thiopie durant la priode du 1er janvier au 31 dcembre 1972.*

- *<np-m > gave no food supplies to <country > <prep > <np-f > <month > 1 to <date >.*

- *<np-m > ne a pas envoy de vivres en <country > <prep > <np-f > de le 1er <month > au <date >.*

- Such a sentence pair would then allow partial matches, as well as more complete matches.

- *he gave no food supplies to Zimbabwe*

- *March 1 to March 20th*

KBMT

- The mainline engine of PANGLOSS is the knowledge-based engine.

- It consists of an analyzer, called the PANGLYZER and a generation module centered on the generator called PENMAN.

- The approach has been to develop the system in a bottom up manner, providing layer after layer of increasingly abstract analysis in a multi-pass process

Lexical Transfer MT

- PANGLOSS uses a simple and traditional lexical transfer MT engine as a safety net.
- Lexical transfer is carried out using a number of bilingual resources:
  - the lexicons developed as an aid in the KBMT engine
  - a machine readable dictionary (Spanish-English Collins)
  - a set of manually produced glossaries.

Pangloss System Details

- As of 2000, versions of PanLite (Pangloss Lite) existed for translating unrestricted Spanish to English, Serbo-Croatian to English, English to Spanish, and English to Serbo-Croatian.

- The code is the same for each version, with just databases and configuration files changing.

Code

- PanLite main program: 4,500 lines of code
- EBMT/glossary: 12,300 lines of code
- LM: 9,700 lines of code
- FramepaC: 50,600 lines of code (used by all three programs)
- Total object code size: about 1200K for SunOS and 900K for Windows NT.

Data

- PanLite:
  - 39,800-word Serbo-Croatian stem list
  - 12,300-word English root list
  - 41,300-word Spanish root list

- EBMT:
- 280M Spanish-English corpus
- 280M English-Spanish corpus (inverse of S-E)
- 2.3M SerboCroatian-English corpus
- 2.3M English-SerboCroatian corpus (inverse of SC-E)
- 19,700-word English root/synonym list
- 56,900-word Spanish-Eng association dictionary
- 21,300-word Eng-SCro association dict
- 51,100-word SCro-Eng association dict

- Glossaries:
  - 193,000-entry Spanish-English glossary
  - 85,000-entry SerboCroatian-English glossary
  - 129,000-entry English-SerboCroatian glossary

- Language Modeller:
  - 13M Serbo-Croatian model (from about 12M text)
  - 60M English model (from about 450M text)
  - 41M Spanish model (from about 135M text)

Pangloss Implementation

- Pangloss and its EBMT translation engine PanEBMT [5] were written in C++

- Using the FramepaC library for accessing Lisp data structures stored in files or sent from the main Pangloss module via Unix pipes.

Pangloss Performance

- Hogan et al. [6] evaluated the performance of MEMT systems.

- What they were inspecting in general was the architecture itself, rather than Pangloss translation system.

- They tested English to Croatian and Croatian to English and used 4 human evaluators 2 of them being Croatian native speakers and the other two English native speakers.

Table 1: Initial statistics

		cro1	cro2	eng1	eng2	Total
EBMT	mean	2.08	1.98	2.59	2.99	2.20
	stddev	1.13	1.19	1.29	1.20	1.23
LEX	mean	1.66	1.53	2.30	2.68	1.82
	stddev	0.93	0.91	1.32	1.29	1.10
MEMT	mean	1.92	1.86	2.58	2.96	2.10
	stddev	1.12	1.20	1.42	1.29	1.27
Total	mean	1.80	1.70	2.42	2.80	1.96
	stddev	1.03	1.06	1.34	1.28	1.18

- Another reason these numbers are in favor of EBMT is due to the way means are computed.

- EBMT does not guarantee to provide a translation for every input sentence.

- The averages were computed for those translations EBMT could produce output.

- This is disadvantageous for the other two systems (MEMT and LEX) since they do not have the luxury of producing empty output.

- Assuming we assign the score 1 - Totally incomprehensible to missing EBMT translations, EBMT scores can be updated as

Table 2: Updated EBMT statistics

		cro1	cro2	eng1	eng2	Total
EBMT	mean	1.79	1.72	2.07	2.35	1.87
	stddev	1.08	1.11	1.30	1.36	1.17

- Now the average scores are 1.82 for LEX, 2.10 for MEMT, and 1.87 for EBMT

- LEX and EBMT are very close to each other whereas MEMT is significantly better than the two.

- Another interesting issue is evaluator agreement, since evaluators may assign different scores along the scale to the same translation.

- All evaluators compromise when we inspect the scores comparatively for systems.

- That is,

  - MEMT = EBMT > LEX

  - MEMT > EBMT = LEX

- depending on the scoring mechanism you use for EBMT.

- The MEMT translation system does no translation itself

- But tries to choose the best translation out of the outputs of its translation engines

- The real performance of an MEMT can be measured by checking whether the system can identify the best translation

- For that purpose, we need define a metric that answers the following question: how often can an MEMT system choose the best translation available?

$$d(\tau_m) = \frac{\sum_{i=1}^{N} \delta(\sigma_{im}, \max\{\sigma_{i1}, \ldots, \sigma_{iM}\})}{N}$$

Where $\delta(i, j)$ is a Kronecker function and is given by the following:

$$\delta(i, j) = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases}$$

Table 3: Percentages

	cro1	cro2	eng1	eng2	Total
EBMT	55.80%	55.60%	49.69%	46.58%	53.86%
LEX	72.80	69.00	69.57	66.46	70.20
MEMT	73.20	73.00	70.19	70.81	72.47

- This metric is also independent of the translation engines used in the MEMT system, since it evaluates the architecture itself.
- If different translation engines were used, the metric would not change since it is measuring the capability of the architecture to choose the best translation among the ones produced by the translation engines.

- Lastly, these numbers present a better evaluator agreement than the mean.

- For the same material, evaluator difference is no more than 3%

- For MEMT systems, this number goes as low as 1%, indicating a higher evaluator agreement.

Diplomat Speech Recognition

- Contemporary speech-recognition systems derive their power from corpus-based statistical modeling, both at the acoustic and language levels.

- Building a speech-recognition system for a target domain or language requires models at three levels (assuming that a basic processing infrastructure for training and decoding is already in place): acoustic, lexical and language.

- After the speech recognition step, the best overall hypothesis is displayed as text on the screen.

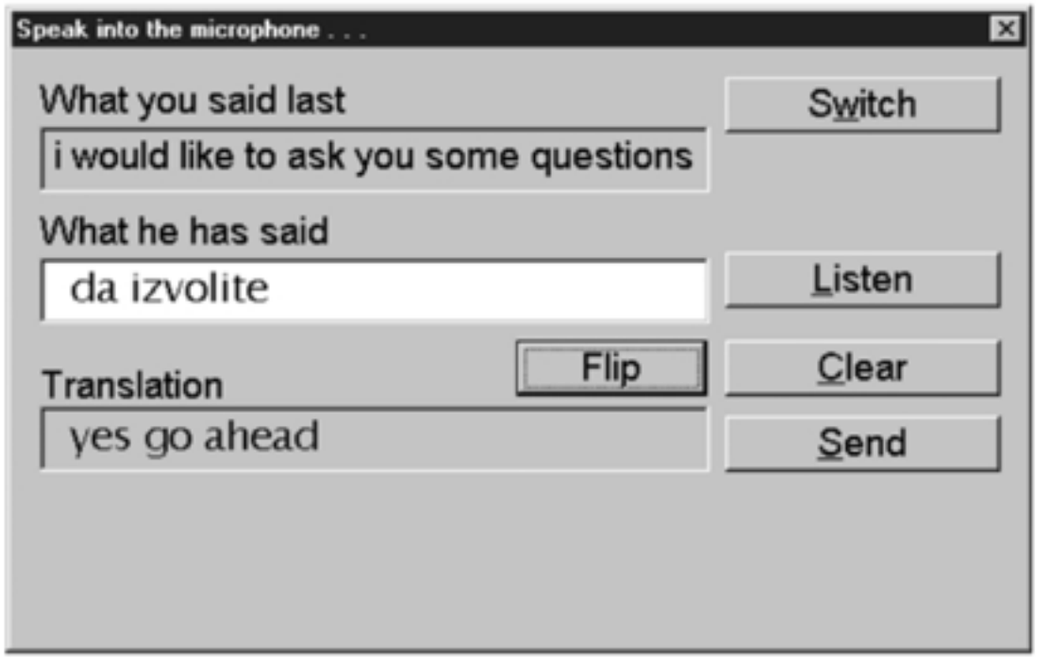
- The user can highlight an incorrect portion using the touchscreen, and respeak or type it

User Interface

- The primary potential use for Diplomat identified so far is to allow English speaking soldiers on peace-keeping missions to interview local residents.

- While it is possible to train the interviewer to use a restricted vocabulary, the interviewees usage of language is much more difficult to control or predict.

- An initial system has been developed to run on a either a laptop or a wearable computer, with each speaker taking turns using a graphical user interface (GUI) on a single display screen


Speak into the microphone . . .
What you said last
i would like to ask you some questions
Switch
What he has said
da izvolite
Listen
Translation
Flip
Clear
yes go ahead
Send

Languages Deployed

- Diplomat currently works with the following language pairs:

- Croatian - English

- Spanish - English

- Haitian Creole - English

- All translations are bi-directional. Preliminary work was also done for Korean-English and Arabic-English.

Tongues

- Due to computational resource and training data limitations, the Diplomat system was never expected to perform at impressive levels of quality compared to systems running on large machine working with huge corpora

- A field test did not occur during the project.

- It was expected to be field-tested by US soldiers in Bosnia or Haiti but that never happened

Tongues

- As of 2002, the Diplomat project was moved into the project Tongues [8] led by Lockheed-Martin, which is basically an extension of the Diplomat project.

References

1. R. Frederking, A. Rudnicky, C. Hogan, and K. Lenzo, "Interactive speech translation in the diplomat project," Machine Translation , vol. 15, no. 1, pp. 27-42, 2000.

2. R. Frederking and R. D. Brown, "The pangloss-lite machine translation system," in Expanding MT Horizons: Proceedings of the Second Conference of the Association for Machine Translation in the Americas , pp. 268-272, 1996.

3. R. Frederking, S. Nirenburg, D. Farwell, S. Helmreich, E. Hovy, K. Knight, S. Beale, C. Domashnev, D. Attardo, D. Grannes, et al. , "Integrating translations from multiple sources within the pangloss mark iii machine translation system," in Technology Partnerships for Crossing the Language Barrier: Proceedings of the First Conference of the Association for Machine Translation in the Americas , pp. 73-80, 1994.

4. M. Nagao, "A framework of a mechanical translation between japanese and english by analogy principle," Artificial and human intelligence , pp. 351-354, 1984.

5. R. D. Brown, "Example-based machine translation in the pangloss system," in Proceedings of the 16th conference on Computational linguistics-Volume 1 , pp. 169-174, Association for Computational Linguistics, 1996.

6. C. Hogan and R. E. Frederking, "An evaluation of the multi-engine mt architecture," in Conference of the Association for Machine Translation in the Americas , pp. 113-123, Springer, 1998.

7. F. Chen and K. Jokinen, Speech Technology, Theory and Applications . Springer, 2010.

8. A. W. Black, R. Brown, R. Frederking, R. Singh, J. Moody, and E. Steinbrecher, "Tongues: Rapid development of a speech-to-speech translation system," in Proceedings of the second international conference on Human Language Technology Research , pp. 183-186, Morgan Kaufmann Publishers Inc., 2002.