# Improving Word Alignment With Bridge Languages

Shankar Kumar
Joint Work with Franz Och and Wolfgang Macherey
The Language Translation Team
Google, Inc.
Nov 8, 2007

# Statistical Approach to MT

Goal: High quality translation of natural language text

- Viewpoint of statistical Machine Translation (MT):
  - In machine translation we have to make decisions under uncertainty.
  - Lets try to make optimal decisions.
- Advantages:
  - General framework for handling ambiguities, combining unreliable knowledge sources and integrating prior knowledge
  - Measure of success: performance on unseen test data
  - Automatic training methods
    - We are already doing this: Chinese, Arabic, Russian from/to English
    - Excellent performance in NIST '05 and '06 MT evaluations



Google

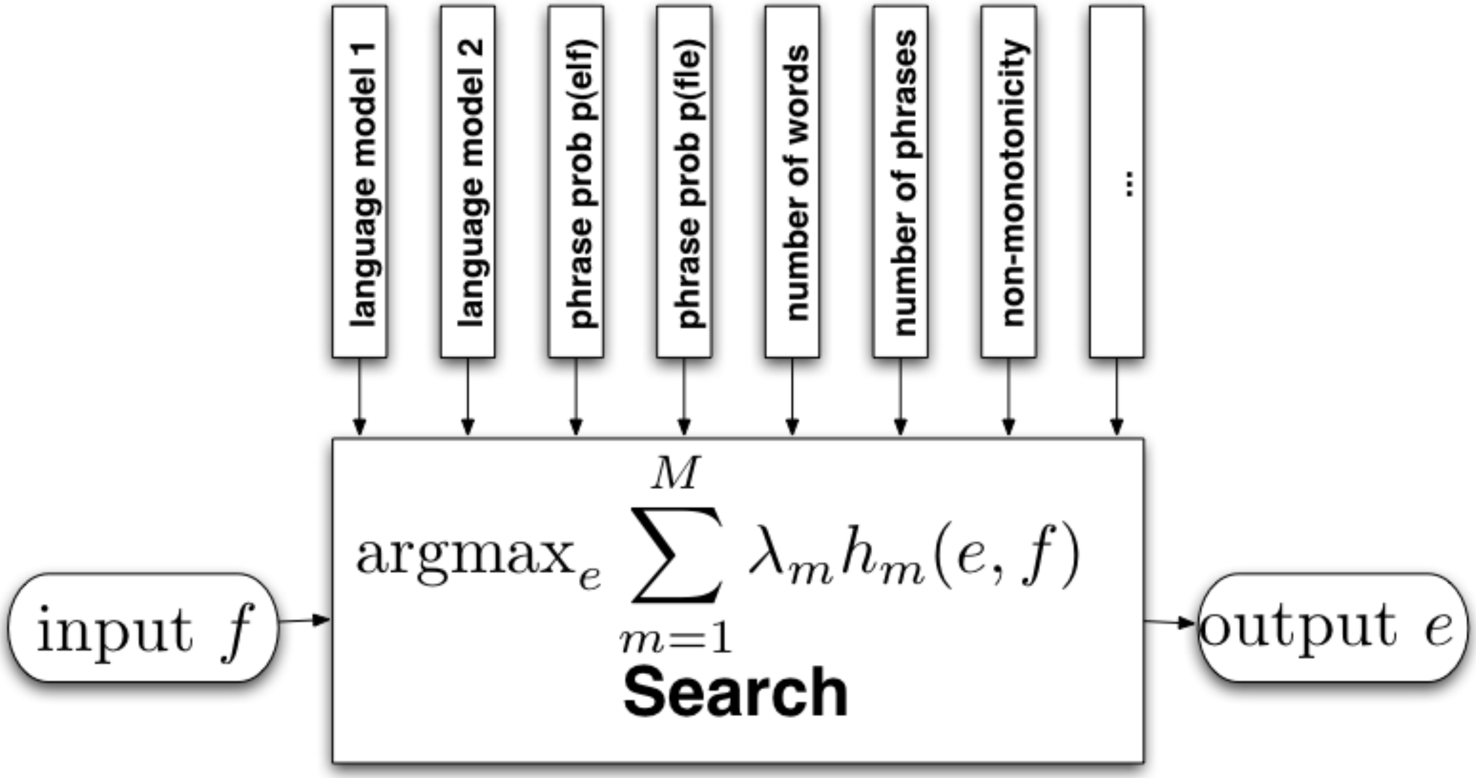
# Evaluation of MT

- Problem: Evaluation by humans is expensive, slow, subjective
- Goal: automatic, objective evaluation of MT
  - Crucial during system development
  - Much progress in research due to systematic use of automatic evaluation criteria
- Approach: compare MT output with human references
- BLEU metric (Papineni)
  - Compute precision of uni-, bi-, tri-, fourgram
  - Average + brevity penalty
  - 0.0: no overlap with references
  - 1.0: perfect overlap
- BLEU is highly correlated with subjective judgments
- Introduction of BLEU in 2001 had huge positive impact on MT

# Outline

- Overview of Statistical Machine Translation at Google
- Improving Word Alignment with Bridge Languages

# SMT: Translation as a search problem



# Phrase Translation Model: Training Steps

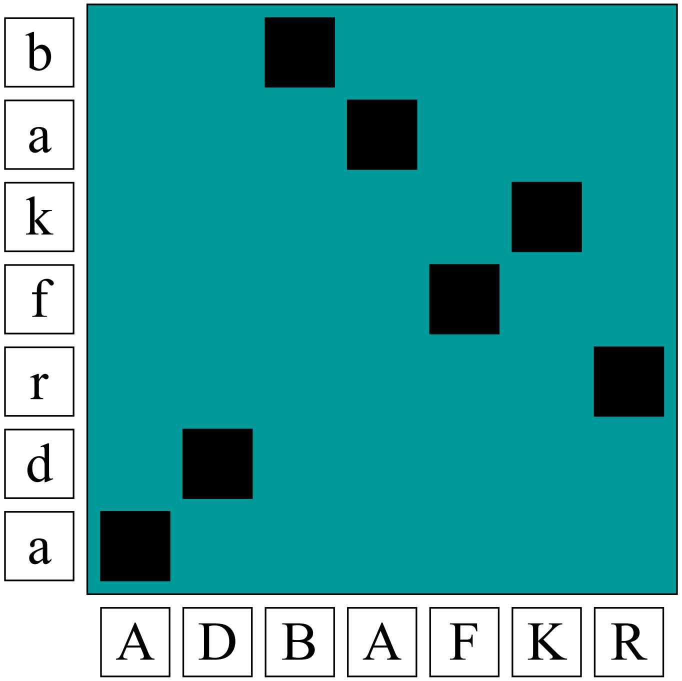
1. Find parallel data
2. Document alignment
3. Preprocessing/tokenization
4. Sentence/chunk alignment
5. Word alignment
6. Phrase-Pair extraction

# TM Training: Sentence/chunk alignment

Goal: Find corresponding sentence/chunks in aligned documents

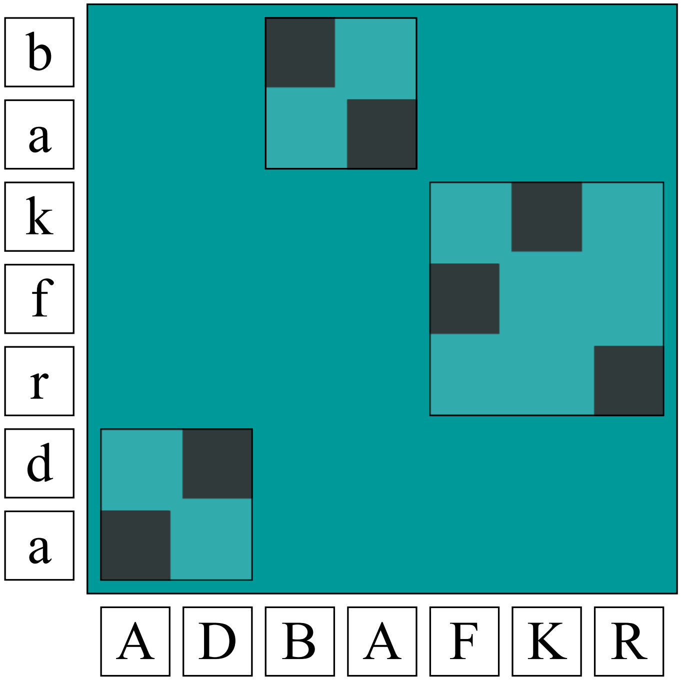
- Score sentence alignments using
  - Dictionary overlap
  - Sentence length mismatch
- Assumption
  - Monotone translation of sentences
  - Alignment Possibilities: 1-1, 2-1, 1-2
  - Dynamic programming search for optimal alignment

# TM Training: Word alignment



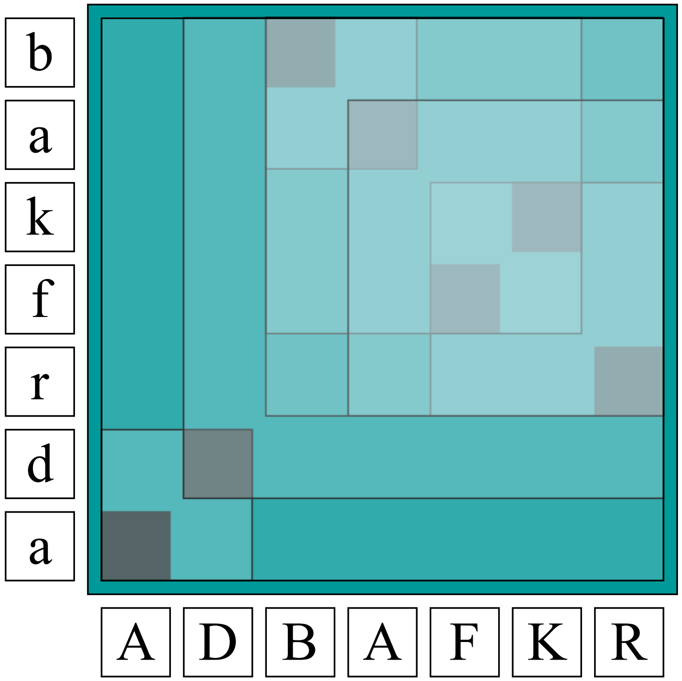

- Treat word alignment as a hidden variable in a probabilistic model
- Maximum Likelihood training using EM algorithm (more later)

# TM Training: Phrase Extraction




- Find all aligned phrase pairs in word alignment

# TM Training: Phrase Extraction




- Find all aligned phrase pairs in word alignment
- Provide various quality signals for assessing 'quality' of phrase
  - Phrase translation probability $p(f|e), p(e|f)$
  - Word translation probability
  - ...

# Search: The actual translation process

$$\hat{e}(f, \lambda_1^M) = \underset{e}{\operatorname{argmax}} \sum_{m=1}^{M} \lambda_m h_m(e, f)$$

- For each input sentence
  - Get candidate phrase for each source language substring
  - Search for optimal translation according to the log-linear model
- Algorithm
  - Dynamic programming beam-search
- Reordering constraints
  - Local reordering up to 7 words

# Recent work

[1] Stefan Riezler, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu Mittal, Yi Liu. Statistical Machine Translation for Query Expansion in Answer Retrieval In *ACL 2007*.

[2] Thorsten Brants, Ashok C. Popat, Peng Xu, Franz Och, Jeffrey Dean. Large Language Models in Machine Translation In *EMNLP 2007*

[3] Wolfgang Macherey and Franz Och. An Empirical Study on Computing Consensus Translations from Multiple Machine Translation Systems In *EMNLP 2007*

[4] Shankar Kumar, Franz Och, Wolfgang Macherey. Improving Word Alignment with Bridge Languages In *EMNLP 2007*

# Outline

- Overview of SMT at Google
- Improving Word Alignment with Bridge Languages

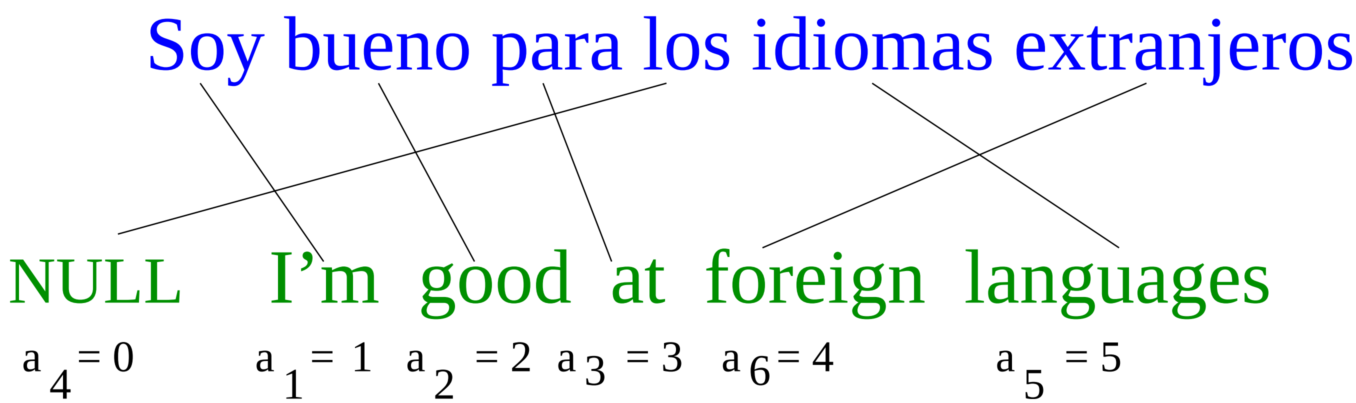
# Improving Word Alignment with Bridge Languages

For a language pair such as Arabic-English, a third language such as Spanish is a bridge language if word-alignments for Arabic-English are derived using Arabic-Spanish and Spanish-English alignments

- Multi-lingual parallel corpora are richer than bilingual corpora
  - Word-alignment errors in Arabic-English are somewhat orthogonal to the errors in Arabic-Spanish or Spanish-English
  - Can we correct Arabic-English alignment errors given Arabic-Spanish and Spanish-English alignments?
- Translation systems derived from bridge language alignments provide a diverse pool of hypotheses for system combination
- Can use language-pairs (e.g. Spanish-English) which can be trained on lots of training data and have high alignment accuracy
  - Not the focus of this work
  - We train all systems on the exact same sentence-pairs

# Word Alignment definitions

*An English-Spanish Sentence Pair:* $(e_1^I, f_1^J)$




- Full alignment space: $\{(j, i)\}$
- Constraints: Each Spanish word aligns to exactly one English word
  - $f_j$ is aligned to $e_{a_j} \rightarrow$ Alignment : $a_1^J$
  - Empty (NULL) word accounts for unaligned Spanish words

# Word Alignment Framework

- Sentence Pair: $\mathbf{f} = f_1^J, \mathbf{e} = e_1^I$
- Alignment as a hidden variable $\mathbf{a} = a_1^J$

$$P(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} P_\theta(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

- Maximum Likelihood or Viterbi Alignment

$$\hat{\mathbf{a}} = \operatorname*{argmax}_{\mathbf{a}} P(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

- Maximum A Posteriori (MAP) Alignment (Ge ’04, Matusov ’04)

$$\begin{aligned} P(a_j = i|\mathbf{e}, \mathbf{f}) &= \sum_{\mathbf{a}} P(\mathbf{a}|\mathbf{f}, \mathbf{e})\delta(i, a_j) \\ a_{\text{MAP}}(j) &= \operatorname*{argmax}_{i} P(a_j = i|\mathbf{e}, \mathbf{f}) \end{aligned}$$

# Constructing Word Alignment Using a Bridge Language

- Get alignment for FE sentence-pair $(\mathbf{f}, \mathbf{e})$ using a translation $\mathbf{g}$ in a third language G.
- Express FE word alignments using FG and GE alignments

$$P(a_j^{FE} = i|\mathbf{e}, \mathbf{f}) = \sum_{k=0}^{K} P(a_j^{FG} = k|\mathbf{g}, \mathbf{f}) P(a_k^{GE} = i|\mathbf{g}, \mathbf{e})$$

- Matrix Multiplication of FG and GE posterior probability matrices
  - Prepend an extra column in the GE matrix

$$P(a_k^{GE} = i|k = 0) = \begin{cases} \epsilon & i = 0 \\ \frac{1-\epsilon}{I} & i \in \{1, 2, ..., I\} \end{cases}$$

  - Higher $\epsilon \rightarrow$ more empty alignments; for the experiments, $\epsilon = 0.5$

# Word Alignment Combination: Multiple Bridge Languages

Suppose we have $N$ bridge languages $G_1, G_2, ...., G_N$

$$P(a_j^{FE} = i|\mathbf{e}, \mathbf{f}) = \sum_{l=0}^{N} P(B = G_l) P(a_j^{FE} = i|G_l, \mathbf{e}, \mathbf{f})$$

- $G_0$ corresponds to direct alignment without a bridge (None)
- Weight each language uniformly with probability $\frac{1}{N+1}$
- Linear Interpolation of Posterior Probability Matrices

# Experiments

Goal: Improve an Arabic-English system

- Training Data: ODS United Nations corpus: 6 languages
  - English(En)/French(Fr)/Chinese(Zh)/Spanish(Es)/Russian(Ru)/Arabic(Ar)
- All other components from Google's 2006 NIST Unlimited Track system
- Development Data for Minimum Error Rate Training (MERT)
  - 2007 sents from NIST '01-'05
- Test Data: test(1610 sents from NIST '01-'05), blind(nist06/1797 sents)

# Experiments (Continued)

Need Aligned Sentence Pairs in all 6 languages

- Run sentence-alignment for Ar-En/Ar-X/X-En (9 pairs)
- Find the common subset of sentence-pairs: 1.8M/7.0M
  - 55M Arabic tokens/58 M English tokens
- Train models for all language pairs with same recipe: Model1-6, HMM-6
- Generate 6 word alignments for Ar-En
  - No bridge language (None)
  - Bridge Languages Es/Fr/Ru/Zh
  - Alignment Combination (AC) using None/Es/Fr/Ru
- In each case, we obtain Ar→En and En→Ar word alignments
- Paper has two more sets of experiments where we relax the constraint that each sentence-pair be present in all languages

# Alignment Performance

Precision/Recall/Alignment Error Rate on 94 sentences with human alignments

| Bridge Language | Metrics(%) | | | | | |
|---|---|---|---|---|---|---|
| | AE | | | EA | | |
| | Prec | Rec | AER | Prec | Rec | AER |
| None | 74.1 | 73.9 | 26.0 | 67.3 | 57.7 | 37.9 |
| Es | 61.7 | 56.3 | 41.1 | 50.0 | 40.2 | 55.4 |
| Fr | 52.9 | 48.0 | 49.7 | 42.3 | 33.6 | 62.5 |
| Ru | 57.4 | 50.8 | 46.1 | 40.2 | 31.6 | 64.6 |
| Zh | 44.3 | 39.3 | 58.3 | 39.7 | 29.9 | 65.9 |
| AC | 70.0 | 65.0 | 32.6 | 56.8 | 46.4 | 48.9 |

Spanish is the best bridge language for alignment

# Translation Results

- Build Phrase-based SMT systems from each word alignment
- Each system produces 2 translations
  1. with a fixed set of log-linear weights (-MERT)
  2. with MERT
- MBR-like Consensus Decoding by combining translations from 6 systems.

| Bridge Lang | test | blind |
|---|---|---|
| None | <u>52.1</u> | <u>40.1</u> |
| Es | 51.7 | 39.8 |
| Fr | 51.2 | 39.5 |
| Ru | 50.4 | 38.7 |
| Zh | 48.4 | 37.1 |
| AC | 52.1 | 40.3 |

# Translation Results

- Build Phrase-based SMT systems from each word alignment
- Each system produces 2 translations
  1. with a fixed set of log-linear weights (-MERT)
  2. with MERT
- MBR-like Consensus Decoding by combining translations from 6 systems.

| Bridge Lang | test | blind |
|---|---|---|
| None | <u>52.1</u> | <u>40.1</u> |
| Es | 51.7 | 39.8 |
| Fr | 51.2 | 39.5 |
| Ru | 50.4 | 38.7 |
| Zh | 48.4 | 37.1 |
| AC | 52.1 | 40.3 |

| Consensus Decoding | | |
|---|---|---|
| None | 51.9 | 39.8 |
| +Es | 52.2 | 40.0 |
| +Fr | 52.4* | 40.5* |
| +Ru | 52.8* | 40.7* |
| +Zh | 52.6* | 40.6* |
| +AC | **53.0*** | **40.9*** |

# Why do bridge languages help ?

- Spanish is the best bridge language in isolation
- For consensus: French, Russian and combination are also useful !

Inter-system BLEU scores relative to the None/-MERT System

| None/-MERT | None/+MERT | Es | Fr | Ru | Zh | AC |
|---|---|---|---|---|---|---|
| 100.0 | 85.7 | 60.0 | 59.8 | 59.7 | 59.5 | 58.7 |

Bridge Language systems have lower correlation than a discriminatively trained system

# Conclusions

- Overview of the phrase-based statistical machine translation system at Google

- A simple approach to use bridge languages to improve word alignments for SMT: matrix multiplication and linear interpolation
- Advantages
  - No need for human word alignments to train combination weights though such weights might help
  - Can use any underlying word alignment model: word-based, syntax-based
- Disadvantage: Requires sentence-aligned data. Might be possible to create bridge language systems using automatic translations
- Possible Implications
  - Use interpolation for combining word alignments from other sources: HMM/Model-4
  - Identify language families without linguistic knowledge if we have a large multi-parallel corpus!

# Thank you!