Heuristic word alignment with parallel phrases

Maria Holmqvist
marho@ida.liu.se
Department of Computer and Information Science
Linköping University, Sweden

LREC May 19, 2010
Outline

- Word alignment with parallel phrases
- English-Swedish gold-standard word alignments
- Alignment experiments
- Comparison with Giza++
- Conclusions and future work
Word alignment

- Identify corresponding words in a text and its translation

**I do not think it is necessary for classic cars to be part of** the directive.

**Jag anser det inte nödvändigt att veteranbilar skall utgöra en del av** direktivet.

I – jag
I do not think – jag anser inte
I do not think it is necessary – jag anser det inte nödvändigt
...

- Applications:
  - Statistical machine translation
    - parallel segments (phrases) extracted from a word aligned parallel corpus
  - Bilingual dictionaries
  - Annotation projection
Word alignment with parallel phrases

- Parallel phrases from manually word aligned texts can be used to align new text.

- Extract all **parallel phrases** from a manually aligned parallel text
  - I do not think – jag anser inte – 0-0 2-2 3-1
  - do not think – anser inte – 1-1 2-0
  - ...

- Match parallel phrases to new sentence pairs

- If source and target phrase match the new sentence, add links

- Generalize phrases with parts-of-speech to increase recall

| Phrase          | Alignment                       | Example |
|-----------------|---------------------------------|---------|
| the European economy | den europeiska ekonomin        | 0-0 1-1 2-2 |
| DET European economy  | DET europeiska ekonomin        | 0-0 1-1 2-2 |
| the A economy     | den A ekonomin                 | 0-0 1-1 2-2 |
| DET A economy     | DET A ekonomin                 | 0-0 1-1 2-2 |
Word alignment with parallel phrases cont.

- in this N, V
- i det N, V

0-0 1-1 2-2 4-4 5-3

Heuristics for matching phrases to new sentences

- Prefer longer phrases over shorter
  - more context -> better alignments

- Prefer phrases with words over phrases with POS
  - DET European economy, DET Europeiska ekonomin

- Skip phrases that match more than once in the sentence
Manual word alignments

- English-Swedish Europarl corpus (Koehn, 2003)

| English                                                                 | Swedish                                                                 |
|------------------------------------------------------------------------|------------------------------------------------------------------------|
| But we must support our tourism!                                       | Men vi måste stödja vår turism!                                       |
| These are events and situations that can not be tolerated.             | Detta är händelser och situationer som vi inte kan acceptera.          |
| Well in simple economic terms, Europe's culture adds to genuine prosperity. | Jo, i enkla ekonomiska termer främjar europeisk kultur verkligt välstånd. |

- Training data set (1000 sentence pairs)
  - Extract parallel phrases
  - One annotator

- Reference set (200 sentence pairs)
  - Evaluation
  - Two annotators
  - Confidence labels for links
The English-Swedish reference word alignment

- Guidelines for English-Swedish reference alignment similar to Spanish-English guidelines (Lambert et al, 2005)
- Link types
  - Sure links
  - Possible links
  - Null links

- Two annotators
  - Agreement 85.8%
  - Alignments combined into the final reference alignment
  - Different link types -> Possible link

- The final reference alignment
  - 73% sure links
  - 27% possible links
Experiment 1:
Generalized phrases

- Europarl English-Swedish training data
  - 1000 sentence pairs with manual alignments
  - parallel phrases 2-7 words were extracted from 900 sentence pairs
  - word alignment evaluated on 100 sentence pairs

| Method                  | Precision | Recall | F-score |
|-------------------------|-----------|--------|---------|
| phrases                 | 92.25     | 16.85  | 28.50   |
| generalized phrases     | 48.81     | 55.20  | 51.81   |

- Some generalized phrases propose incorrect word links
  - Ex. PREP N -> PREP N
    N -> N N
    of DET -> DET
Experiment 2: Constraining generalization

- Which generalized phrases produce correct word links?
- Thresholds for phrase length and generalization
  - P: Phrase length
  - L: Minimum phrase length for generalization
  - G: Max number of words generalized with POS

- P=1, L=3, G=1
  - Phrases of length 1 and up
  - Phrases have to be at least 3 words long to be generalized with POS
  - At most one word is generalized

| Thresholds (P-L-G) | Precision | Recall | F-score |
|-------------------|-----------|--------|---------|
| 1-3-1             | 86.8      | 42.8   | 57.3    |
| 1-4-7             | 72.0      | 49.8   | 58.9    |
| 2-5-2             | 95.4      | 17.5   | 29.6    |
| 2-2-7             | 48.8      | 55.2   | 51.8    |
Experiment 3: Evaluating phrases on training data

- Apply phrases to the training data and collect
  - precision for each phrase
  - number of matches
- Perform word alignment with a subset of phrases
  - E.g., phrases with a training data precision above 95%

| Setting | Training precision threshold | Frequency threshold | Precision  | Recall  |
|---------|-----------------------------|---------------------|------------|---------|
| 1-3-1   | 0.95                        | 5                   | 99.6       | 12.2    |
| 1-3-1   | 0.95                        | 3                   | 99.4       | 15.1    |
| 1-3-1   | 0.95                        | 2                   | 99.1       | 18.4    |
| 1-3-1   | 0.90                        | 3                   | 98.5       | 17.9    |
| 1-3-1   | 0.90                        | 2                   | 98.3       | 20.8    |
| 1-3-1   | 0.85                        | 2                   | 98.1       | 23.2    |
| 1-3-1   | 0.80                        | 1                   | 95.8       | 30.6    |
Creating final phrase-based alignments

- We have sets of reliable phrases
  - Generalization thresholds
  - Training data precision
  - Precision on the devtest set

- Alignment of test data
  - Combination of reliable phrases
  - Add linksets in sequence according to their precision on the devtest set

| Method       | Precision | Recall  | F-mått |
|--------------|-----------|---------|--------|
| Combination1 | 95.85     | 28.27   | 43.66  |
| Combination2 | 90.61     | 41.73   | 57.14  |
Giza++

- Giza++ (Och and Ney, 2003)
  - State-of-the-art system for statistical word alignment
  - Produces one-to-many alignments

- Method for symmetrization
  - Intersection – high precision
  - Union – high recall
  - Grow-diag – best AER

- Results for Giza++ trained on 700K sentences:

| Method         | Precision | Recall | F-score | AER  |
|----------------|-----------|--------|---------|------|
| intersect      | 94.77     | 57.05  | 71.22   | 16.31|
| union          | 70.09     | 77.17  | 73.46   | 21.77|
| grow-diag      | 82.35     | 73.30  | 77.56   | 15.46|
Results for Giza++ and phrase-based word alignment

| Method                        | Precision | Recall | AER  |
|-------------------------------|-----------|--------|------|
| Combination 1                 | 95.8      | 28.3   | 45.7 |
| Combination 2                 | 90.6      | 41.7   | 31.5 |
| Giza++ grow-diag 700K         | 82.3      | 73.3   | 15.5 |
| Giza++ grow-diag 5K           | 71.4      | 62.0   | 26.6 |
| Giza++ intersect 700K         | 94.8      | 57.1   | 16.3 |
| Giza++ intersect 5K           | 93.3      | 42.8   | 28.7 |

- Evaluation on test data (200 sentences)
  - Precision is comparable
  - Recall is lower than Giza++
A combination of Giza++ and phrase-based alignments

- Combine phrase based word alignment with statistical alignment
  - Add links from Giza++ for unaligned words
  - Results in improved AER

| Method                     | Precision | Recall | AER  |
|----------------------------|-----------|--------|------|
| Giza++ 700k intersect      | 94.77     | 57.05  | 16.31|
| + Combination1             | 93.41     | **60.10** | **14.81** |
| Giza++ 700K grow-diag      | 82.35     | 73.30  | 15.46|
| + Combination1             | **84.60** | 71.64  | **14.22** |
A combination of Giza++ and phrase-based alignments

- Larger improvements in AER for Giza++ trained on a small data set (5K)

| Method           | Precision | Recall | AER  |
|------------------|-----------|--------|------|
| Giza++ 5K intersect | 93.3      | 42.8   | 28.7 |
| + Combination1   | 92.2      | 48.2   | 24.4 |
| Giza++ 5K grow-diag | 71.4      | 62.0   | 26.6 |
| + Combination1   | 75.1      | 62.2   | 23.9 |
Conclusions

- Word alignment with parallel phrases creates alignments with high precision (90-95%)
- Generalizing phrases with part-of-speech increased recall
- We have shown two methods to select reliable subsets of phrases that produce high-precision word alignments
- A combination of phrase-based word alignment and Giza++ produced a word alignment with lower AER than the best Giza++ alignment
Future work

- Other factors to find good phrase matches
  - relative position of matches in a sentence pair

- New language pairs

- Different merging strategy with Giza++. Perhaps during symmetrization.

- Better machine translation quality?
Thanks for listening!

Questions?