Projection of Trees across Parallel Texts

Daniel Zeman, Rudolf Rosa

April 17, 2020
Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, Okan Kolak (2004). Bootstrapping Parsers via Syntactic Projection across Parallel Texts

- Source: English
- Target: Spanish, Chinese
- Dependency trees (not phrase structure)
Projection System Architecture

- Bilingual corpus
  - English
  - Non-English

- English Dependency Parser
- Word Alignment Model

- Projected Non-English Dependency Treebank
  - Train

- Dependency Parser

- New Non-English sentences

- Dependency trees for new sentences

- Projection
- Transformation
- Filtering
Direct Projection

Given sentence pair $(E, F)$ and a set of syntactic relations for $E$, where $E = e_1, ..., e_n$ is an English sentence and $F = f_1, ..., f_m$ is its non-English parallel, syntactic relations $R(x, y)$ are projected from English as follows:

- **one-to-one** – $e_i$ aligned with a unique $f_x$ and $e_j$ aligned with a unique $f_y$ – then $R(e_i, e_j) \Rightarrow R(f_x, f_y)$
Direct Projection

Given sentence pair \((E, F)\) and a set of syntactic relations for \(E\), where \(E = e_1, ..., e_n\) is an English sentence and \(F = f_1, ..., f_m\) is its non-English parallel, syntactic relations \(R(x, y)\) are projected from English as follows:

- **one-to-one** – \(e_i\) aligned with a unique \(f_x\) and \(e_j\) aligned with a unique \(f_y\) – then \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\)

- **unaligned English** – \(e_j\) not aligned with any word in \(F\) – create new empty word \(f_y\) so that for any \(e_i\) aligned with a unique \(f_x\), \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\) and \(R(e_j, e_i) \Rightarrow R(f_y, f_x)\)
Direct Projection

Given sentence pair \((E, F)\) and a set of syntactic relations for \(E\), where \(E = e_1, \ldots, e_n\) is an English sentence and \(F = f_1, \ldots, f_m\) is its non-English parallel, syntactic relations \(R(x, y)\) are projected from English as follows:

- **one-to-one** – \(e_i\) aligned with a unique \(f_x\) and \(e_j\) aligned with a unique \(f_y\) – then 
  \[ R(e_i, e_j) \Rightarrow R(f_x, f_y) \]

- **unaligned English** – \(e_j\) not aligned with any word in \(F\) – create new empty word \(f_y\) so that for any \(e_i\) aligned with a unique \(f_x\), 
  \[ R(e_i, e_j) \Rightarrow R(f_x, f_y) \] and 
  \[ R(e_j, e_i) \Rightarrow R(f_y, f_x) \]

- **one-to-many** – \(e_i\) aligned with \(f_x, \ldots, f_y\) – then create new empty \(f_z\), parent of \(f_x, \ldots, f_y\), and set \(e_i\) to align to \(f_z\) instead

- **many-to-one** – \(e_i, \ldots, e_j\) uniquely aligned to \(f_x\) – then keep the head of \(e_i, \ldots, e_j\) aligned to \(f_x\), and delete other alignments

- **many-to-many** – decompose: first one-to-many, then many-to-one

- **unaligned foreign** – leave them out of the projected tree
Given sentence pair \((E, F)\) and a set of syntactic relations for \(E\), where \(E = e_1, \ldots, e_n\) is an English sentence and \(F = f_1, \ldots, f_m\) is its non-English parallel, syntactic relations \(R(x, y)\) are projected from English as follows:

- **one-to-one** – \(e_i\) aligned with a unique \(f_x\) and \(e_j\) aligned with a unique \(f_y\) – then \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\)

- **unaligned English** – \(e_j\) not aligned with any word in \(F\) – create new empty word \(f_y\) so that for any \(e_i\) aligned with a unique \(f_x\), \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\) and \(R(e_j, e_i) \Rightarrow R(f_y, f_x)\)

- **one-to-many** – \(e_i\) aligned with \(f_x, \ldots, f_y\) – then create new empty \(f_z\), parent of \(f_x, \ldots, f_y\), and set \(e_i\) to align to \(f_z\) instead

- **many-to-one** – \(e_i, \ldots, e_j\) uniquely aligned to \(f_x\) – then keep the head of \(e_i, \ldots, e_j\) aligned to \(f_x\), and delete other alignments
Direct Projection

Given sentence pair \((E, F)\) and a set of syntactic relations for \(E\), where \(E = e_1, ..., e_n\) is an English sentence and \(F = f_1, ..., f_m\) is its non-English parallel, syntactic relations \(R(x, y)\) are projected from English as follows:

- **one-to-one** – \(e_i\) aligned with a unique \(f_x\) and \(e_j\) aligned with a unique \(f_y\) – then \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\)

- **unaligned English** – \(e_j\) not aligned with any word in \(F\) – create new empty word \(f_y\) so that for any \(e_i\) aligned with a unique \(f_x\), \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\) and \(R(e_j, e_i) \Rightarrow R(f_y, f_x)\)

- **one-to-many** – \(e_i\) aligned with \(f_x, ..., f_y\) – then create new empty \(f_z\), parent of \(f_x, ..., f_y\), and set \(e_i\) to align to \(f_z\) instead

- **many-to-one** – \(e_i, ..., e_j\) uniquely aligned to \(f_x\) – then keep the head of \(e_i, ..., e_j\) aligned to \(f_x\), and delete other alignments

- **many-to-many** – decompose: first one-to-many, then many-to-one
Direct Projection

Given sentence pair \((E, F)\) and a set of syntactic relations for \(E\), where \(E = e_1, \ldots, e_n\) is an English sentence and \(F = f_1, \ldots, f_m\) is its non-English parallel, syntactic relations \(R(x, y)\) are projected from English as follows:

- **one-to-one** – \(e_i\) aligned with a unique \(f_x\) and \(e_j\) aligned with a unique \(f_y\) – then \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\)
- **unaligned English** – \(e_j\) not aligned with any word in \(F\) – create new empty word \(f_y\) so that for any \(e_i\) aligned with a unique \(f_x\), \(R(e_i, e_j) \Rightarrow R(f_x, f_y)\) and \(R(e_j, e_i) \Rightarrow R(f_y, f_x)\)
- **one-to-many** – \(e_i\) aligned with \(f_x, \ldots, f_y\) – then create new empty \(f_z\), parent of \(f_x, \ldots, f_y\), and set \(e_i\) to align to \(f_z\) instead
- **many-to-one** – \(e_i, \ldots, e_j\) uniquely aligned to \(f_x\) – then keep the head of \(e_i, \ldots, e_j\) aligned to \(f_x\), and delete other alignments
- **many-to-many** – decompose: first one-to-many, then many-to-one
- **unaligned foreign** – leave them out of the projected tree
Direct Projection Example

He took a picture of my daughter

Vyfotili si moji dceru
He took a picture of my daughter.

Vyfotil si moji dceru.
He took a picture of my daughter.
He took a picture of my daughter.
He took a picture of my daughter.
He took a picture of my daughter.
He took a picture of my daughter.
He took a picture of my daughter Vyfotil simoji dceru
He took a picture of my daughter.
He took a picture of my daughter

Projection of Trees across Parallel Texts
He took a picture of my daughter
Many-to-One Assumption:

\( e_i, \ldots, e_j \) Is a Phrase with One Head
Many-to-One Assumption:

$e_i, \ldots, e_j$ is a phrase with one head. What if not?

He took a picture of my daughter.
Experiments with Direct Projection

- 100 gold trees projected from English to Spanish
- 88 gold trees projected from English to Chinese

- Word alignments are gold-standard too!
  - The goal is just to check the direct correspondence assumption.
Experiments with Direct Projection

- 100 gold trees projected from English to Spanish
- 88 gold trees projected from English to Chinese

- Word alignments are gold-standard too!
  - The goal is just to check the direct correspondence assumption.

- Compared with target gold-standard trees
  - Spanish unlabeled F-score = 37%
  - Chinese unlabeled F-score = 38%
Problems

- Many-to-one deletes alignments ⇒ tree is not connected
  - Possible solution: transitive closure?

He took a picture of my daughter

\[ f_1 \quad \text{Vyfotil} \quad \text{si} \quad f_6 \quad \text{moji dceru} \]
Problems

- Many-to-one deletes alignments \(\Rightarrow\) tree is not connected
  - Possible solution: transitive closure?

He took a picture of my daughter

Projection of Trees across Parallel Texts
Problems

- Many-to-one deletes alignments $\Rightarrow$ tree is not connected
  - Possible solution: transitive closure?

- Unaligned foreign words remain unattached
  - Possible solution: postprocessing with target language knowledge
A few dozen rules, less than a month work

Spanish example
  A reflexive clitic should modify the verb to its left.

Chinese example
  An aspectual marker should modify the verb to its left.
100 gold trees projected from English to Spanish
88 gold trees projected from English to Chinese

Word alignments are gold-standard too!

Compared with target gold-standard trees
- Spanish unlabeled F-score = 70%
- Chinese unlabeled F-score = 67%
Collins Model2 (1997) English parser trained on Penn Treebank / WSJ

Converted to dependencies (Magerman 1994, Xia and Palmer 2001)

Word alignments computed with GIZA++ (Och and Ney 2003)
- 100K en-es sentence pairs (Bible, Federal Broadcasting Information Service, United Nations Parallel Corpus)
- 240K en-zh sentence pairs (Federal Broadcasting Information Service)

Project trees using direct correspondence + postprocessing
Real-World Setting

- Collins Model2 (1997) English parser trained on Penn Treebank / WSJ
- Converted to dependencies (Magerman 1994, Xia and Palmer 2001)

- Word alignments computed with GIZA++ (Och and Ney 2003)
  - 100K en-es sentence pairs (Bible, Federal Broadcasting Information Service, United Nations Parallel Corpus)
  - 240K en-zh sentence pairs (Federal Broadcasting Information Service)

- Project trees using direct correspondence + postprocessing
- Aggressive filtering: discard projected trees of poor quality
Real-World Setting

- Collins Model2 (1997) English parser trained on Penn Treebank / WSJ
- Converted to dependencies (Magerman 1994, Xia and Palmer 2001)

- Word alignments computed with GIZA++ (Och and Ney 2003)
  - 100K en-es sentence pairs (Bible, Federal Broadcasting Information Service, United Nations Parallel Corpus)
  - 240K en-zh sentence pairs (Federal Broadcasting Information Service)

- Project trees using direct correspondence + postprocessing
- Aggressive filtering: discard projected trees of poor quality
- Train Collins dependency parser (1999) on remaining trees
- Apply the parser to unseen target-language sentences
Pruning Criteria

- Based on tuning on development set, discard if...
  - > 20% of the English words have no Spanish counterpart
Pruning Criteria

- Based on tuning on development set, discard if...
  - > 20% of the English words have no Spanish counterpart
  - > 30% of the Spanish words have no English counterpart

Additional criteria for English-Chinese:
- Crossing dependencies
- Number of unattached nodes after postprocessing
- Number of words with unknown POS category
- 20K projected Spanish trees after filtering
- 50K projected Chinese trees after filtering
Pruning Criteria

- Based on tuning on development set, discard if...
  - > 20% of the English words have no Spanish counterpart
  - > 30% of the Spanish words have no English counterpart
  - > 4 Spanish words were aligned to the same English word
Pruning Criteria

- Based on tuning on development set, discard if…
  - > 20% of the English words have no Spanish counterpart
  - > 30% of the Spanish words have no English counterpart
  - > 4 Spanish words were aligned to the same English word
- Additional criteria for English-Chinese:
  - Crossing dependencies
  - Number of unattached nodes after postprocessing
  - Number of words with unknown POS category
Pruning Criteria

- Based on tuning on development set, discard if...
  - > 20% of the English words have no Spanish counterpart
  - > 30% of the Spanish words have no English counterpart
  - > 4 Spanish words were aligned to the same English word
  - Additional criteria for English-Chinese:
    - Crossing dependencies
    - Number of unattached nodes after postprocessing
    - Number of words with unknown POS category

- 20K projected Spanish trees after filtering
- 50K projected Chinese trees after filtering
Experiments

- **Spanish**
  - Baseline (left-to-right) unl F-score = 33.8%
  - Parser on unfiltered data (98K) F = 67.3%
  - Parser on filtered data (20K) F = 72.1%
  - Commercial parser F = 69.2%

- **Chinese**
  - Baseline (left-to-right) F = 35.1%
  - Baseline + postprocessing F = 44.3%
  - Parser on filtered data (50K) F = 53.9%
  - Parser on PennChineseTB (10K) F = 64.3%

- Learning curve: projected parser = about 2K manual sentences
Experiments

- **Spanish**
  - Baseline (left-to-right) unl F-score = 33.8%
  - Parser on unfiltered data (98K) F = 67.3%
  - Parser on filtered data (20K) F = 72.1%
  - Commercial parser F = 69.2%

- **Chinese**
  - Baseline (left-to-right) F = 35.1%
  - Baseline + postprocessing F = 44.3%
  - Parser on filtered data (50K) F = 53.9%
  - Parser on PennChineseTB (10K) F = 64.3%
Experiments

- Spanish
  - Baseline (left-to-right) unl F-score = 33.8%
  - Parser on unfiltered data (98K) F = 67.3%
  - Parser on filtered data (20K) F = 72.1%
  - Commercial parser F = 69.2%

- Chinese
  - Baseline (left-to-right) F = 35.1%
  - Baseline + postprocessing F = 44.3%
  - Parser on filtered data (50K) F = 53.9%
  - Parser on PennChineseTB (10K) F = 64.3%

- Learning curve: projected parser = about 2K manual sentences