Hiero

Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

A Motivating Example from Chiang 2007

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

The English translation:

Australia is one of the few countries that have diplomatic relations with North Korea

Why the difference in word order?

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

The English translation:

Australia is one of the few countries that have diplomatic relations with North Korea

shaoshu guojia zhiyi $\Leftrightarrow$ one of the few countries

Why the difference in word order?

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

The English translation:

Australia is one of the few countries that have diplomatic relations with North Korea

shaoshu guojia zhiyi ⇔ one of the few countries

yu Beihan you bangjiao de ⇔
that have diplomatic relations with North Korea

Why the difference in word order?

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

The English translation:

Australia is one of the few countries that have diplomatic relations with North Korea

shaoshu guojia zhiyi ⇔ one of the few countries

yu Beihan you bangjiao de ⇔
that have diplomatic relations with North Korea

A Motivating Example from Chiang 2007

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

The English translation:

Australia is one of the few countries that have diplomatic relations with North Korea

Output from a phrase-based system:

[Aozhou] $[\text{shi}]_1$ $[\text{yu Beihan}]_2$ [you] [bangjiao] [de shaoshu guojia zhiyi]

[Australia] [has] [dipl. rels.] $[\text{with North Korea}]_2$ $[\text{is}]_1$ [one of the few countries]

A Solution: Hierarchical Phrases

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

Hierarchical phrases needed for this example:

$\langle$ yu $\boxed{1}$ you $\boxed{2}$, have $\boxed{2}$ with $\boxed{1}$ $\rangle$

$\langle$ $\boxed{1}$ de $\boxed{2}$, the $\boxed{2}$ that $\boxed{1}$ $\rangle$

$\langle$ $\boxed{1}$ zhiyi, one of $\boxed{1}$ $\rangle$

(We'll see how to formalize this next.)

Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

Synchronous CFGs

- Rules in a synchronous CFG (s-CFG) take the form:

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle$$

  where:

  - X is any non-terminal in the grammar
  - γ and α are strings of terminals and non-terminals
  - $\sim$ is a one-to-one correspondence between non-terminal occurrences in γ and α.

- An important constraint: Non-terminals that correspond to each other *must be the same*.

Examples of s-CFG Rules (from Chiang 2007)

$X \rightarrow \langle$ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}} \rangle$

$X \rightarrow \langle$ $X_{\boxed{1}}$ de $X_{\boxed{2}}$, the $X_{\boxed{2}}$ that $X_{\boxed{1}} \rangle$

$X \rightarrow \langle$ $X_{\boxed{1}}$ zhiyi, one of $X_{\boxed{1}} \rangle$

Note: these rules make use of a single non-terminal, X

We use subscripts such as $\boxed{1}$, $\boxed{2}$ to specify which non-terminals correspond to each other.

Examples of s-CFG Rules

Another valid s-CFG rule:

$VP \rightarrow \langle$ $PP_{\boxed{1}}$ you $NP_{\boxed{2}}$, have $NP_{\boxed{2}}$ $PP_{\boxed{1}} \rangle$

In this case three non-terminals, NP, PP, and VP are used. The above rule is perfectly valid in an s-CFG. However, Chiang's grammar only makes use of two non-terminals: X and S.

Examples of s-CFG Rules

An **invalid** s-CFG rule:

$VP \rightarrow \langle\ PP_{\boxed{1}}$ you $NP_{\boxed{2}}$, have $NP_{\boxed{2}}\ X_{\boxed{1}}\ \rangle$

This rule is invalid because a PP corresponds to an X. *Non-terminals that correspond to each other must be the same.*

More Examples of s-CFG Rules from Chiang 2007

$X \rightarrow \langle$ Aozhu, Australia $\rangle$

$X \rightarrow \langle$ Beihan, North Korea $\rangle$

$X \rightarrow \langle$ shi, is $\rangle$

$X \rightarrow \langle$ bangjiao, diplomatic relations $\rangle$

$X \rightarrow \langle$ shaoshu guojia, few countries $\rangle$

$S \rightarrow \langle\ S_{\boxed{1}}\ X_{\boxed{2}}\ ,\ S_{\boxed{1}}\ X_{\boxed{2}}\ \rangle$

$S \rightarrow \langle\ X_{\boxed{1}},\ X_{\boxed{1}}\ \rangle$

Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

Intuition Behind Translation with an s-CFG

- *First step:* we can read off a CFG for Chinese from the s-CFG, and parse the Chinese with this CFG
- For example,

  $X \rightarrow \langle$ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}}\ \rangle$

  implies the Chinese-only context-free rule

  $$X \rightarrow \text{yu}\ \ X\ \ \text{you}\ \ X$$

  and

  $X \rightarrow \langle$ bangjiao, diplomatic relations $\rangle$

  implies the Chinese-only context-free rule

  $$X \rightarrow \text{bangjiao}$$

Intuition Behind Translation with an s-CFG

The resulting CFG for Chinese:

$X \rightarrow$ yu X you X

$X \rightarrow X$ de X

$X \rightarrow X$ zhiyi

$X \rightarrow$ Aozhu

$X \rightarrow$ Beihan

$X \rightarrow$ shi

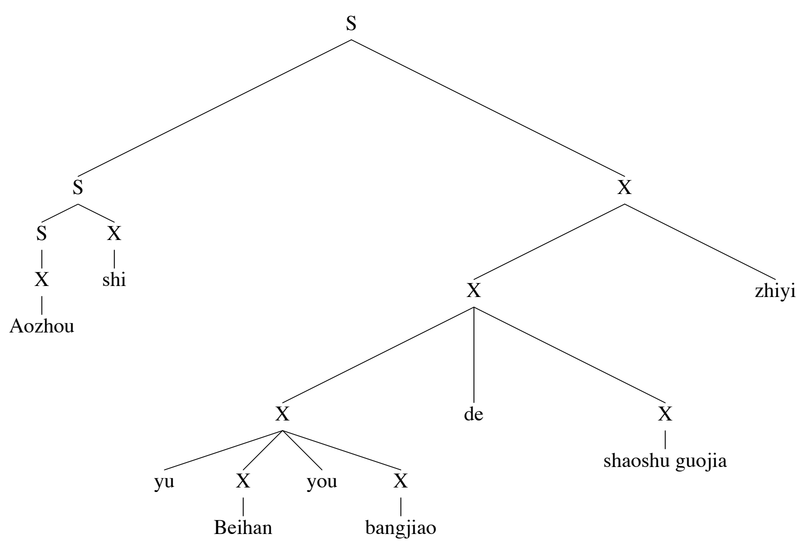
$X \rightarrow$ bangjiao

$X \rightarrow$ shaoshu guojia

$S \rightarrow S\ X$

$S \rightarrow X$

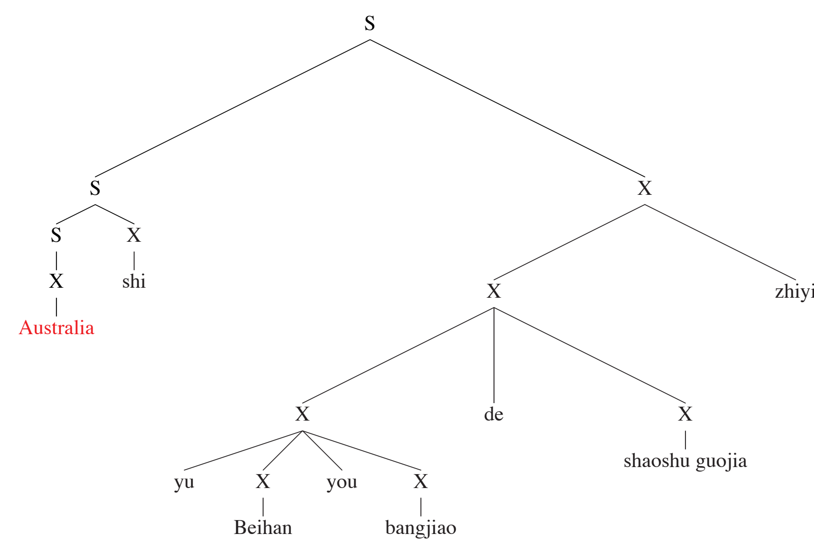
A parse tree for our example:



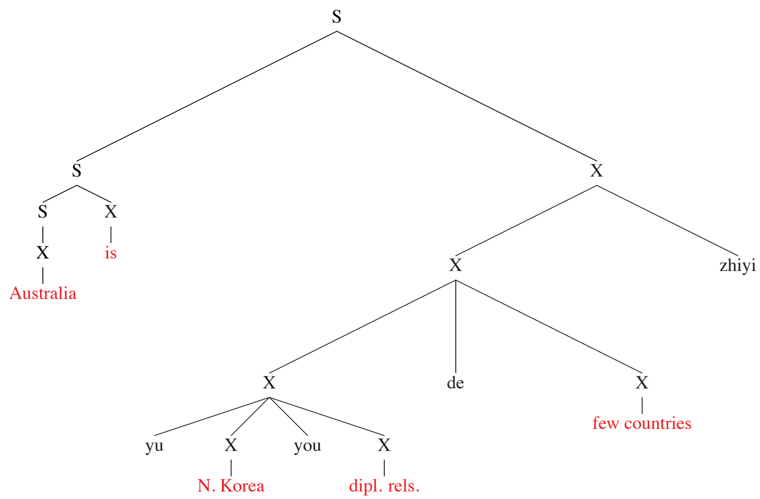
Intuition Behind Translation with an s-CFG

- *First step:* we can read off a CFG for Chinese from the s-CFG, and parse the Chinese with this CFG
- *Second step:* we use the synchronous rules to map the Chinese parse tree to an English parse tree

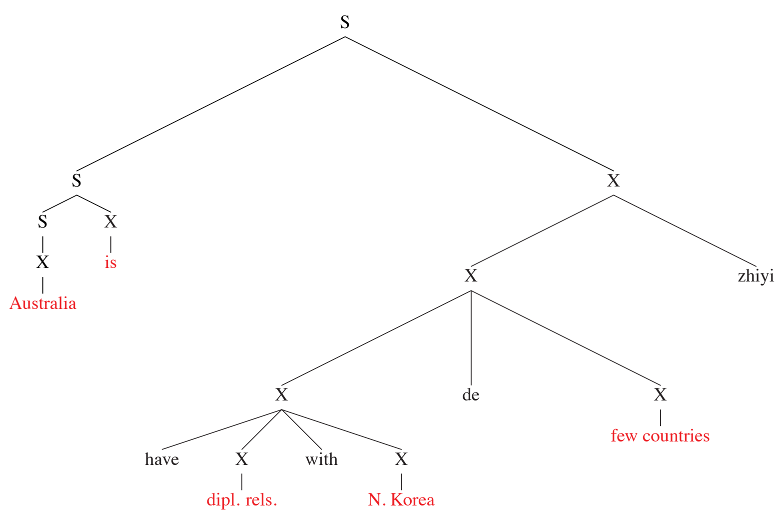
Start bottom-up. For example, $X \rightarrow \langle$ Aozhu, Australia $\rangle$ gives:



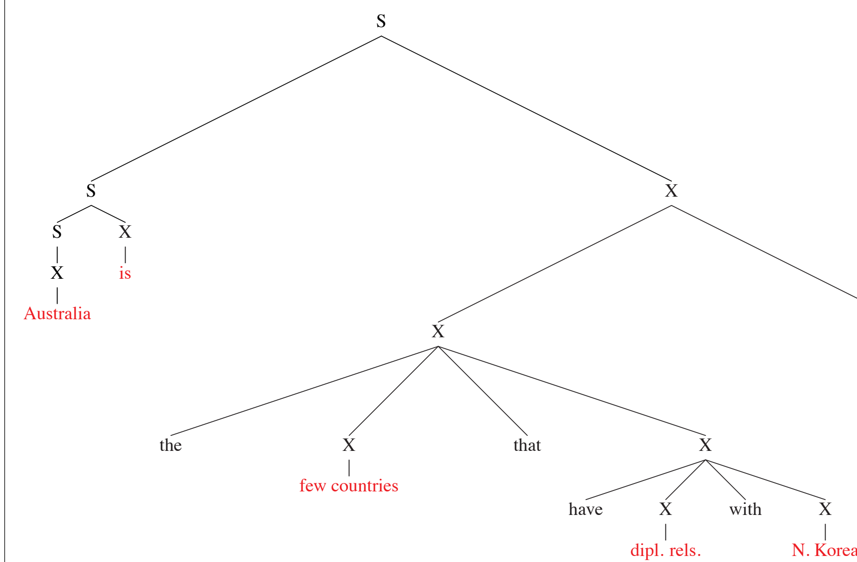
The tree after all the lowest-level rules are applied:



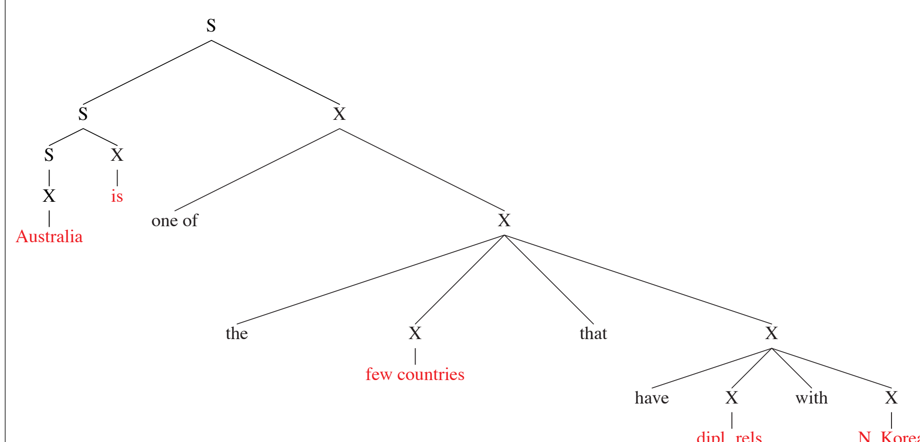
Next, apply higher-level rules. For example, use
$X \rightarrow \langle$ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}} \rangle$ **to get:**



Use $X \rightarrow \langle X_{\boxed{1}}$ de $X_{\boxed{2}}$, the $X_{\boxed{2}}$ that $X_{\boxed{1}} \rangle$ **to get:**



Use $X \rightarrow \langle X_{\boxed{1}}$ zhiyi, one of $X_{\boxed{1}} \rangle$ **to get:**



Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

Derivations in s-CFGs

- A derivation in a conventional CFG generates a single string, using the rules in the grammar
- A derivation in an s-CFG generates a *pair* of strings

Derivations in s-CFGs

- We always start with the following pair of strings in an s-CFG derivation:

$$\langle S_{\boxed{1}}, S_{\boxed{1}} \rangle$$

  (S is the start symbol in the grammar)

- We'll call the left-hand string in the derivation the *foreign* string, the right-hand string the *English* string

Derivations in s-CFGs

- At each step in the derivation:
  - We pick the left-most non-terminal in the foreign string
  - We find the corresponding non-terminal in the English string
  - We expand both non-terminals using some rule in the grammar

An Example of a Derivation Step

- Assume our current foreign/English string pair is as follows:

  $\langle$Aozhou shi $X_{\boxed{8}}$ de $X_{\boxed{9}}$ zhiyi, Australia is one of the $X_{\boxed{9}}$ that $X_{\boxed{8}}\rangle$

- In this case, the left-most non-terminal in the foreign string is $X_{\boxed{8}}$

- We can apply the rule
  $X \rightarrow \langle$ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}} \rangle$
  to derive:

  $\langle$Aozhou shi yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$ de $X_{\boxed{9}}$ zhiyi,
  Australia is one of the $X_{\boxed{9}}$ that have $X_{\boxed{2}}$ with $X_{\boxed{1}}\rangle$

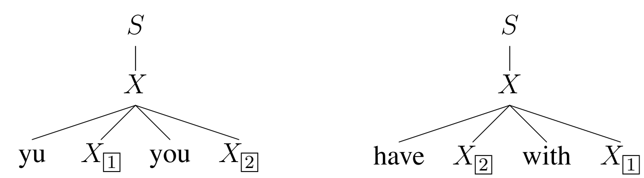
- See Figure 1 from Chiang (2007) for a full derivation

An Example Derivation

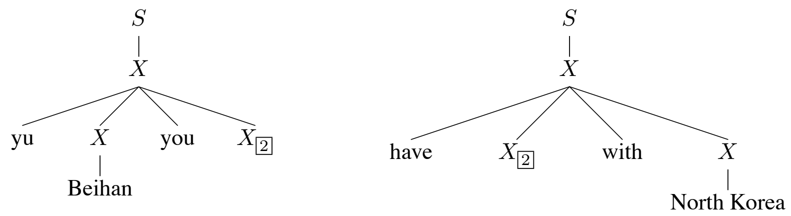
$S_{\boxed{1}}$ $\qquad$ $S_{\boxed{1}}$

An Example Derivation

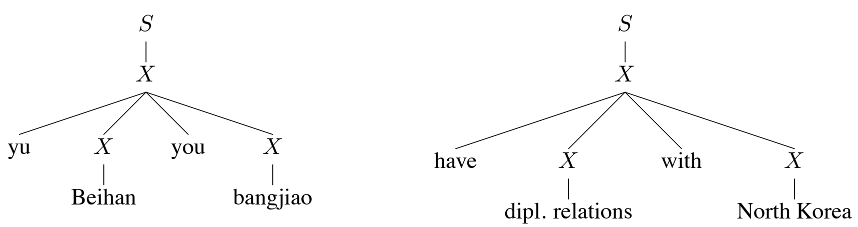
An Example Derivation



An Example Derivation



An Example Derivation



A Subtlety: Re-numbering Non-terminals

- Assume our current foreign/English string pair is as follows:

  $\langle$Aozhou shi $X_{\boxed{8}}$ de $X_{\boxed{1}}$ zhiyi, Australia is one of the $X_{\boxed{1}}$ that $X_{\boxed{8}}\rangle$

- We can apply the rule
  $X \rightarrow \langle$ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}} \rangle$
  to derive:

  $\langle$Aozhou shi yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$ de $X_{\boxed{1}}$ zhiyi,

  Australia is one of the $X_{\boxed{1}}$ that have $X_{\boxed{2}}$ with $X_{\boxed{1}}\rangle$

- This is a mistake: we now have two non-terminals of the form $X_{\boxed{1}}$.

A Subtlety: Re-numbering Non-terminals

- Assume our current foreign/English string pair is as follows:

  $\langle$Aozhou shi $X_{\boxed{8}}$ de $X_{\boxed{1}}$ zhiyi, Australia is one of the $X_{\boxed{1}}$ that $X_{\boxed{8}}\rangle$

- We can apply the rule
  $X \rightarrow \langle$ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}} \rangle$
  to derive:

  $\langle$Aozhou shi yu $X_{\boxed{2}}$ you $X_{\boxed{3}}$ de $X_{\boxed{1}}$ zhiyi,

  Australia is one of the $X_{\boxed{1}}$ that have $X_{\boxed{3}}$ with $X_{\boxed{2}}\rangle$

- **This is the correct way to do things:** we've renumbered the non-terminals in the applied rule to ensure there are no duplicated indices.

Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

Learning an s-CFG from Translation Examples

- Basic idea: leverage/generalize methods from phrase-based systems
- *First step*: use standard methods for phrase-based systems to learn a set of *initial* phrase pairs (see Machine Translation, Part 3, lecture).

A Motivating Example from Chiang 2007

A Chinese sentence:

Aozhou	shi	yu	Beihan	you	bangjiao	de	shaoshu	guojia	zhiyi
Australia	is	with	North Korea	have	dipl. rels.	that	few	countries	one of

The English translation:

Australia is one of the few countries that have diplomatic relations with North Korea

An Example of Initial Phrase Pairs

⟨ Aozhu, Australia ⟩
⟨ Beihan, North Korea ⟩
⟨ shi, is ⟩
⟨ bangjiao, diplomatic relations ⟩
⟨ shaoshu guojia, few countries ⟩
⟨ yu Beihan you bangjiao, have dipl. rels. with North Korea ⟩
⟨ yu Beihan you bangjiao de, that have dipl. rels. with North Korea ⟩
⟨ yu Beihan you bangjiao de shaoshu guojia,
the few countries that have dipl. rels. with North Korea ⟩
. . .

Learning an s-CFG from Translation Examples

The set of rules R is defined as follows:

1. If $\langle f_i^j, e_{i'}^{j'} \rangle$ is an initial phrase pair, then

$$X \rightarrow \langle f_i^j, e_{i'}^{j'} \rangle$$

   is a rule in R

2. If $X \rightarrow \langle \gamma, \alpha \rangle$ is a rule in R, and $\langle f_i^j, e_{i'}^{j'} \rangle$ is an initial phrase pair such that $\gamma = \gamma_1 f_i^j \gamma_2$ and $\alpha = \alpha_1 e_{i'}^{j'} \alpha_2$, then

$$X \rightarrow \langle \gamma_1 X_{\boxed{k}} \gamma_2, \alpha_1 X_{\boxed{k}} \alpha_2 \rangle$$

   where k is an index not used in γ and α, is a rule in R.

An Example

The first case generates rules such as the following:

$X \rightarrow$ ⟨ Aozhu, Australia ⟩
$X \rightarrow$ ⟨ Beihan, North Korea ⟩
$X \rightarrow$ ⟨ shi, is ⟩
$X \rightarrow$ ⟨ bangjiao, diplomatic relations ⟩
$X \rightarrow$ ⟨ shaoshu guojia, few countries ⟩
$X \rightarrow$ ⟨ yu Beihan you bangjiao, have dipl. rels. with North Korea ⟩
$X \rightarrow$ ⟨ yu Beihan you bangjiao de, that have dipl. rels. with North Korea ⟩
$X \rightarrow$ ⟨ yu Beihan you bangjiao de shaoshu guojia,
the few countries that have dipl. rels. with North Korea ⟩
...

An Example of Case 2

- We have the rules

  $X \rightarrow$ ⟨ yu Beihan you bangjiao,
  have dipl. rels. with North Korea ⟩

  $X \rightarrow$ ⟨ Beihan, North Korea ⟩

- We can generate a new rule

  $X \rightarrow$ ⟨ yu $X_{\boxed{1}}$ you bangjiao, have dipl. rels. with $X_{\boxed{1}}$ ⟩

Another Example of Case 2

- We have the rules

  $X \rightarrow$ ⟨ yu $X_{\boxed{1}}$ you bangjiao, have dipl. rels. with $X_{\boxed{1}}$ ⟩

  $X \rightarrow$ ⟨ bangjiao, diplomatic relations ⟩

- We can generate a new rule

  $X \rightarrow$ ⟨ yu $X_{\boxed{1}}$ you $X_{\boxed{2}}$, have $X_{\boxed{2}}$ with $X_{\boxed{1}}$ ⟩

Additional Constraints on Rules

The 2 cases generate a very large number of rules. Examples of additional criteria that Chiang (2007) uses, in an effort to reduce the number of rules:

1. Initial phrases are limited to a length of 10 words on either side.
2. Rules are limited to 5 non-terminals plus terminals on the foreign side.
3. Rules can have at most 2 non-terminals (this simplifies the translation algorithm).
4. It is prohibited for non-terminals to be adjacent on the foreign side.
5. A rule must have at least one pair of aligned words, so that translation decisions are always based on some lexical evidence.

Two Final Rules

The following rules are also always included by Chiang:

$S \rightarrow \langle\, S_{\boxed{1}}\, X_{\boxed{2}}\, ,\, S_{\boxed{1}}\, X_{\boxed{2}}\, \rangle$

$S \rightarrow \langle\, X_{\boxed{1}}, X_{\boxed{1}}\, \rangle$

Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

Adding Probabilities

- Each rule of the form

$$X \rightarrow \langle \gamma, \alpha \rangle$$

has an associated probability

$$P(\gamma|\alpha) = \frac{Count(\gamma, \alpha)}{Count(\alpha)}$$

Scoring Translations

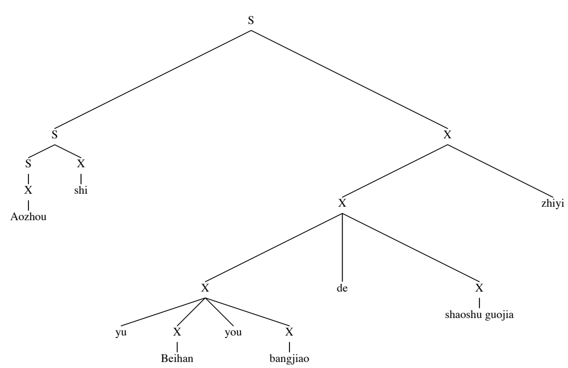
- Define c to be a Chinese sentence, e to be an English sentence, T to be a parse tree under an s-CFG, which generates the pair of strings (c, e)
- A simple model defines

$$Score(c, e, T) = \prod_{X \to \langle \gamma, \alpha \rangle \in T} P(\gamma | \alpha)$$

- The best translation for c is

$$\arg\max_{e,T} Score(c, e, T)$$

Scoring Translations



- The parse tree has a score that is a product of rule probabilities
- We can search for the highest scoring parse tree for a given Chinese sentence using dynamic programming (basically the same algorithm as parsing with a PCFG). Can then transform the Chinese parse tree to its English translation (as we saw earlier).

Scoring Translations: Adding a Language Model

- Define c to be a Chinese sentence, e to be an English sentence, T to be a parse tree under an s-CFG, which generates the pair of strings (c, e)
- A model with a language model defines

$$Score(c, e, T) = P_l(e) \prod_{X \to \langle \gamma, \alpha \rangle \in T} P(\gamma | \alpha)$$

  where $P_l(e)$ is the score under a language model (typically a trigram language model).

- The best translation for c is again

$$\arg\max_{e,T} Score(c, e, T)$$

  Dynamic programming can still be used to find the $\arg\max$*, but it's much more complicated—we won't get into this.*

Overview

The synchronous CFG approach to translation from Chiang (2005, 2007):

- A motivating example
- Synchronous context-free grammars (s-CFGs)
- Translation with s-CFGs
- Derivations in s-CFGs
- Learning a synchronous CFG from translation examples
- Adding probabilities
- Experiments/results from Chiang 2007

Results from Chiang 2007

	MT03	MT04	MT05
ATA	30.84	31.74	30.50
Hiero	33.72	34.57	31.79

- Results are for translation from Chinese to English. MT03, MT04, and MT05 are 3 different test sets. All scores are Bleu scores.

- ATS is a (state-of-the-art) phrase-based system

- Hiero is the synchronous grammar