Models of Synchronous Grammar Induction for SMT

Workshop 2010

The Center for Speech and Language Processing
Johns Hopkins University

June 28, 2010

Statistical machine translation

Urdu → English

اس حملہ کے بعد بڑی تعداد میں مقامی باشندوں نے علاقوں کو خالی کردیا ہے ۔

- Statistical machine translation: Learn how to translate from parallel corpora.

Statistical machine translation:

Urdu → English

اس حملہ کے بعد بڑی تعداد میں مقامی باشندوں نے علاقوں کو خالی کردیا ہے .

After this incident, a large number of local residents fled from these areas.

- Statistical machine translation: Learn how to translate from parallel corpora

Statistical machine translation: state-of-the-art

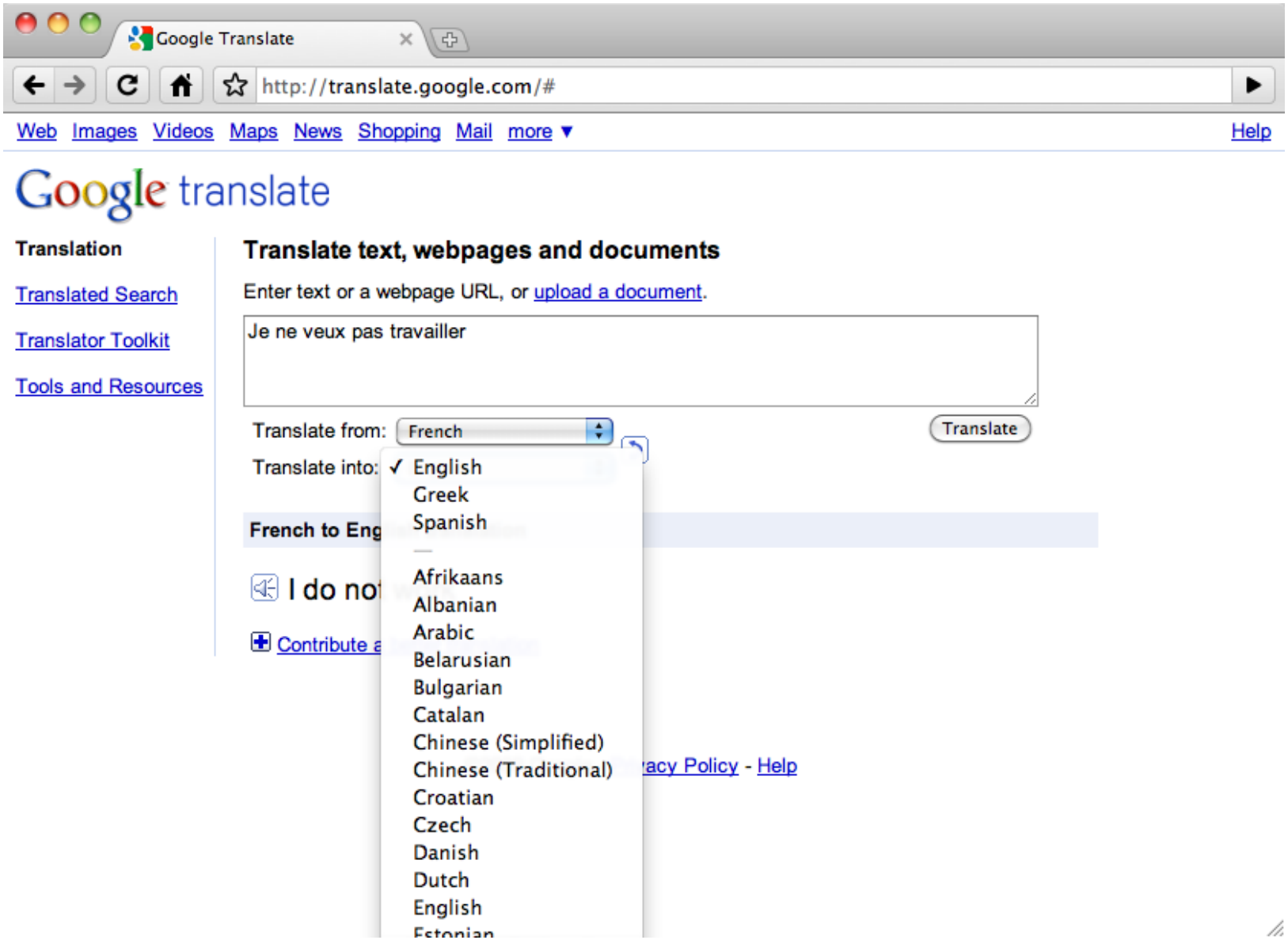
Urdu → English

اس حملہ کے بعد بڑی تعداد میں مقامی باشندوں نے علاقوں کو خالی کردیا ہے .

In this attack a large number of local residents has should vacate areas.

- Current state-of-the-art translation models struggle with language pairs which exhibit large differences in structure.

Statistical machine translation: successes



Statistical machine translation: limitations

Structural divergence between languages:

English	**Who wrote this letter?**
Arabic	من الذي كتب هذه الرسالة؟ (function-word) (who) (wrote) (this) (the-letter)
Chinese	这封 信 是 谁 写 的 ？ (this) (letter) (be) (who) (write) (come-from) (function-word)

Statistical machine translation: limitations

Structural divergence between languages:

English	**Who wrote this letter?**
Arabic	من الذي كتب هذه الرسالة؟ (function-word) (who) (wrote) (this) (the-letter)
Chinese	这封 信 是 谁 写 的 ？ (this) (letter) (be) (who) (write) (come-from) (function-word)

Statistical machine translation: limitations

Structural divergence between languages:

English	**Who wrote this letter?**
Arabic	من الذي كتب هذه الرسالة؟ (function-word) (who) (wrote) (this) (the-letter)
Chinese	这封 信 是 谁 写 的 ？ (this) (letter) (be) (who) (write) (come-from) (function-word)

Statistical machine translation: limitations

Structural divergence between languages:

English	**Who wrote this letter?**
Arabic	من الذي كتب هذه الرسالة؟ (function-word) (who) (wrote) (this) (the-letter)
Chinese	这封 信 是 谁 写 的 ？ (this) (letter) (be) (who) (write) (come-from) (function-word)

- Phrasal translation equivalences
- Constituent reordering
- Morphology

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

$S \rightarrow \langle X_{[1]},\ X_{[1]} \rangle$

$X \rightarrow \langle X_{[1]}\ X_{[2]},\ X_{[2]}\ X_{[1]} \rangle$

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

$X \rightarrow \langle X_{[1]}\ X_{[2]},\ X_{[1]}\ X_{[2]} \rangle$

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$S \Rightarrow \langle X_{[1]},\ X_{[1]} \rangle \Rightarrow \langle X_{[2]}\ X_{[3]},\ X_{[2]}\ X_{[3]} \rangle$

$\Rightarrow \langle Sie\ X_{[3]},\ She\ X_{[3]} \rangle \Rightarrow \langle Sie\ X_{[4]}\ X_{[5]},\ She\ X_{[4]}\ X_{[5]} \rangle$

$\Rightarrow \langle Sie\ will\ X_{[5]},\ She\ wants\ to\ X_{[5]} \rangle \Rightarrow \langle Sie\ will\ X_{[6]}X_{[7]},\ She\ wants\ to\ X_{[7]}X_{[6]} \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{[7]},\ She\ wants\ to\ X_{[7]}\ a\ cup\ of\ coffee \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

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

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

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

$X \rightarrow \langle \textit{Sie},\ \textit{She} \rangle$

$X \rightarrow \langle \textit{will},\ \textit{wants to} \rangle$

$X \rightarrow \langle \textit{eine Tasse Kaffee},\ \textit{a cup of coffee} \rangle$

$X \rightarrow \langle \textit{trinken},\ \textit{drink} \rangle$

Example Derivation

$$S \Rightarrow \langle X_{\boxed{1}},\ X_{\boxed{1}} \rangle \Rightarrow \langle X_{\boxed{2}}\ X_{\boxed{3}},\ X_{\boxed{2}}\ X_{\boxed{3}} \rangle$$

$$\Rightarrow \langle \textit{Sie}\ X_{\boxed{3}},\ \textit{She}\ X_{\boxed{3}} \rangle \Rightarrow \langle \textit{Sie}\ X_{\boxed{4}}\ X_{\boxed{5}},\ \textit{She}\ X_{\boxed{4}}\ X_{\boxed{5}} \rangle$$

$$\Rightarrow \langle \textit{Sie will}\ X_{\boxed{5}},\ \textit{She wants to}\ X_{\boxed{5}} \rangle \Rightarrow \langle \textit{Sie will}\ X_{\boxed{6}} X_{\boxed{7}},\ \textit{She wants to}\ X_{\boxed{7}} X_{\boxed{6}} \rangle$$

$$\Rightarrow \langle \textit{Sie will eine Tasse Kaffee}\ X_{\boxed{7}},\ \textit{She wants to}\ X_{\boxed{7}}\ \textit{a cup of coffee} \rangle$$

$$\Rightarrow \langle \textit{Sie will eine Tasse Kaffee trinken},\ \textit{She wants to drink a cup of coffee} \rangle$$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

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

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

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

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

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$S \Rightarrow \langle X_{\boxed{1}},\ X_{\boxed{1}} \rangle \Rightarrow \langle X_{\boxed{2}}\ X_{\boxed{3}},\ X_{\boxed{2}}\ X_{\boxed{3}} \rangle$

$\Rightarrow \langle Sie\ X_{\boxed{3}},\ She\ X_{\boxed{3}} \rangle \Rightarrow \langle Sie\ X_{\boxed{4}}\ X_{\boxed{5}},\ She\ X_{\boxed{4}}\ X_{\boxed{5}} \rangle$

$\Rightarrow \langle Sie\ will\ X_{\boxed{5}},\ She\ wants\ to\ X_{\boxed{5}} \rangle \Rightarrow \langle Sie\ will\ X_{\boxed{6}} X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}} X_{\boxed{6}} \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}}\ a\ cup\ of\ coffee \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

$S \rightarrow \langle X_{[1]},\ X_{[1]} \rangle$

$X \rightarrow \langle X_{[1]}\ X_{[2]},\ X_{[2]}\ X_{[1]} \rangle$

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

$X \rightarrow \langle X_{[1]}\ X_{[2]},\ X_{[1]}\ X_{[2]} \rangle$

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$S \Rightarrow \langle X_{[1]},\ X_{[1]} \rangle \Rightarrow \langle X_{[2]}\ X_{[3]},\ X_{[2]}\ X_{[3]} \rangle$

$\Rightarrow \langle Sie\ X_{[3]},\ She\ X_{[3]} \rangle \Rightarrow \langle Sie\ X_{[4]}\ X_{[5]},\ She\ X_{[4]}\ X_{[5]} \rangle$

$\Rightarrow \langle Sie\ will\ X_{[5]},\ She\ wants\ to\ X_{[5]} \rangle \Rightarrow \langle Sie\ will\ X_{[6]}X_{[7]},\ She\ wants\ to\ X_{[7]}X_{[6]} \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{[7]},\ She\ wants\ to\ X_{[7]}\ a\ cup\ of\ coffee \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

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

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

$X \rightarrow \langle \textit{Sie},\ \textit{She} \rangle$

$X \rightarrow \langle \textit{eine Tasse Kaffee},\ \textit{a cup of coffee} \rangle$

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

$X \rightarrow \langle \textit{will},\ \textit{wants to} \rangle$

$X \rightarrow \langle \textit{trinken},\ \textit{drink} \rangle$

Example Derivation

$S \Rightarrow \langle X_{\boxed{1}},\ X_{\boxed{1}} \rangle \Rightarrow \langle X_{\boxed{2}}\ X_{\boxed{3}},\ X_{\boxed{2}}\ X_{\boxed{3}} \rangle$

$\Rightarrow \langle \textit{Sie}\ X_{\boxed{3}},\ \textit{She}\ X_{\boxed{3}} \rangle \Rightarrow \langle \textit{Sie}\ X_{\boxed{4}}\ X_{\boxed{5}},\ \textit{She}\ X_{\boxed{4}}\ X_{\boxed{5}} \rangle$

$\Rightarrow \langle \textit{Sie will}\ X_{\boxed{5}},\ \textit{She wants to}\ X_{\boxed{5}} \rangle \Rightarrow \langle \textit{Sie will}\ X_{\boxed{6}} X_{\boxed{7}},\ \textit{She wants to}\ X_{\boxed{7}} X_{\boxed{6}} \rangle$

$\Rightarrow \langle \textit{Sie will eine Tasse Kaffee}\ X_{\boxed{7}},\ \textit{She wants to}\ X_{\boxed{7}}\ \textit{a cup of coffee} \rangle$

$\Rightarrow \langle \textit{Sie will eine Tasse Kaffee trinken},\ \textit{She wants to drink a cup of coffee} \rangle$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

$S \rightarrow \langle X_{[1]},\ X_{[1]} \rangle$

$X \rightarrow \langle X_{[1]}\ X_{[2]},\ X_{[2]}\ X_{[1]} \rangle$

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

$X \rightarrow \langle X_{[1]}\ X_{[2]},\ X_{[1]}\ X_{[2]} \rangle$

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$S \Rightarrow \langle X_{[1]},\ X_{[1]} \rangle \Rightarrow \langle X_{[2]}\ X_{[3]},\ X_{[2]}\ X_{[3]} \rangle$

$\Rightarrow \langle Sie\ X_{[3]},\ She\ X_{[3]} \rangle \Rightarrow \langle Sie\ X_{[4]}\ X_{[5]},\ She\ X_{[4]}\ X_{[5]} \rangle$

$\Rightarrow \langle Sie\ will\ X_{[5]},\ She\ wants\ to\ X_{[5]} \rangle \Rightarrow \langle Sie\ will\ X_{[6]}X_{[7]},\ She\ wants\ to\ X_{[7]}X_{[6]} \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{[7]},\ She\ wants\ to\ X_{[7]}\ a\ cup\ of\ coffee \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

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

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

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

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$S \Rightarrow \langle X_{\boxed{1}},\ X_{\boxed{1}} \rangle \Rightarrow \langle X_{\boxed{2}}\ X_{\boxed{3}},\ X_{\boxed{2}}\ X_{\boxed{3}} \rangle$

$\Rightarrow \langle Sie\ X_{\boxed{3}},\ She\ X_{\boxed{3}} \rangle \Rightarrow \langle Sie\ X_{\boxed{4}}\ X_{\boxed{5}},\ She\ X_{\boxed{4}}\ X_{\boxed{5}} \rangle$

$\Rightarrow \langle Sie\ will\ X_{\boxed{5}},\ She\ wants\ to\ X_{\boxed{5}} \rangle \Rightarrow \langle Sie\ will\ X_{\boxed{6}}X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}}X_{\boxed{6}} \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}}\ a\ cup\ of\ coffee \rangle$

$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

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

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

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

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

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$$S \Rightarrow \langle X_{\boxed{1}},\ X_{\boxed{1}} \rangle \Rightarrow \langle X_{\boxed{2}}\ X_{\boxed{3}},\ X_{\boxed{2}}\ X_{\boxed{3}} \rangle$$

$$\Rightarrow \langle Sie\ X_{\boxed{3}},\ She\ X_{\boxed{3}} \rangle \Rightarrow \langle Sie\ X_{\boxed{4}}\ X_{\boxed{5}},\ She\ X_{\boxed{4}}\ X_{\boxed{5}} \rangle$$

$$\Rightarrow \langle Sie\ will\ X_{\boxed{5}},\ She\ wants\ to\ X_{\boxed{5}} \rangle \Rightarrow \langle Sie\ will\ X_{\boxed{6}} X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}} X_{\boxed{6}} \rangle$$

$$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}}\ a\ cup\ of\ coffee \rangle$$

$$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$$

Using syntax in Machine Translation:

Synchronous Context Free Grammar (SCFG)

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

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

$X \rightarrow \langle Sie,\ She \rangle$

$X \rightarrow \langle eine\ Tasse\ Kaffee,\ a\ cup\ of\ coffee \rangle$

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

$X \rightarrow \langle will,\ wants\ to \rangle$

$X \rightarrow \langle trinken,\ drink \rangle$

Example Derivation

$$S \Rightarrow \langle X_{\boxed{1}},\ X_{\boxed{1}} \rangle \Rightarrow \langle X_{\boxed{2}}\ X_{\boxed{3}},\ X_{\boxed{2}}\ X_{\boxed{3}} \rangle$$

$$\Rightarrow \langle Sie\ X_{\boxed{3}},\ She\ X_{\boxed{3}} \rangle \Rightarrow \langle Sie\ X_{\boxed{4}}\ X_{\boxed{5}},\ She\ X_{\boxed{4}}\ X_{\boxed{5}} \rangle$$

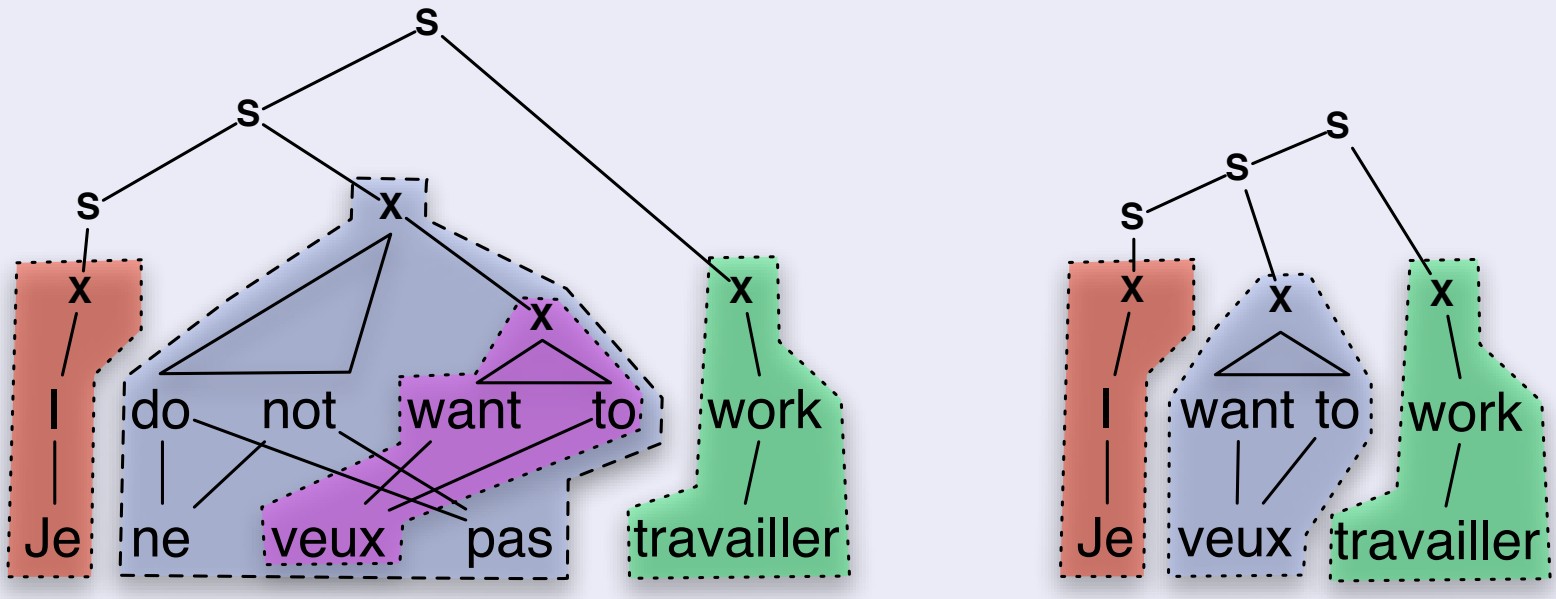
$$\Rightarrow \langle Sie\ will\ X_{\boxed{5}},\ She\ wants\ to\ X_{\boxed{5}} \rangle \Rightarrow \langle Sie\ will\ X_{\boxed{6}}X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}}X_{\boxed{6}} \rangle$$

$$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ X_{\boxed{7}},\ She\ wants\ to\ X_{\boxed{7}}\ a\ cup\ of\ coffee \rangle$$

$$\Rightarrow \langle Sie\ will\ eine\ Tasse\ Kaffee\ trinken,\ She\ wants\ to\ drink\ a\ cup\ of\ coffee \rangle$$

Models of translation

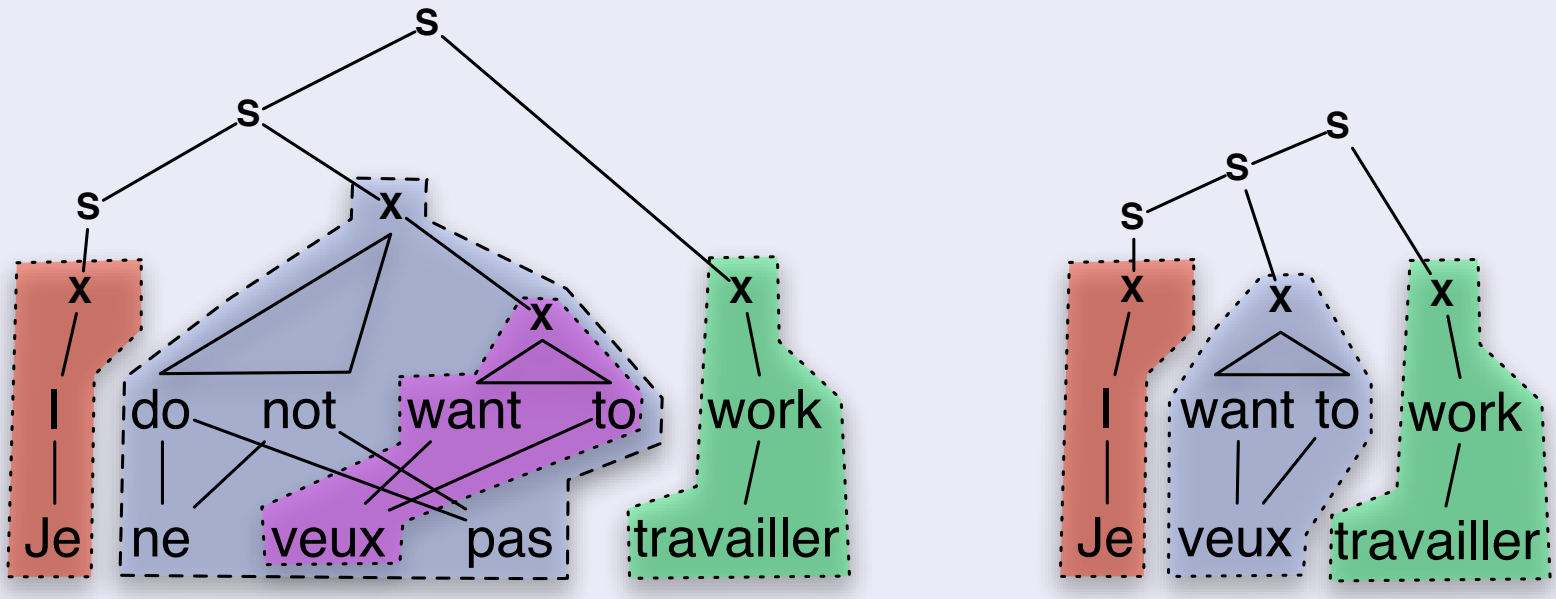
Unlabelled SCFG: Hiero



- S -> ⟨ $X_{[1]}$, $X_{[1]}$ ⟩,
  S -> ⟨ $S_{[1]}$ $X_{[2]}$, $S_{[1]}$ $X_{[2]}$ ⟩
- X -> ⟨ Je, I ⟩, X -> ⟨ ne $X_{[1]}$ pas, do not $X_{[1]}$ ⟩,
  X -> ⟨ veux, want to⟩, X -> ⟨ travailler, work ⟩

Models of translation

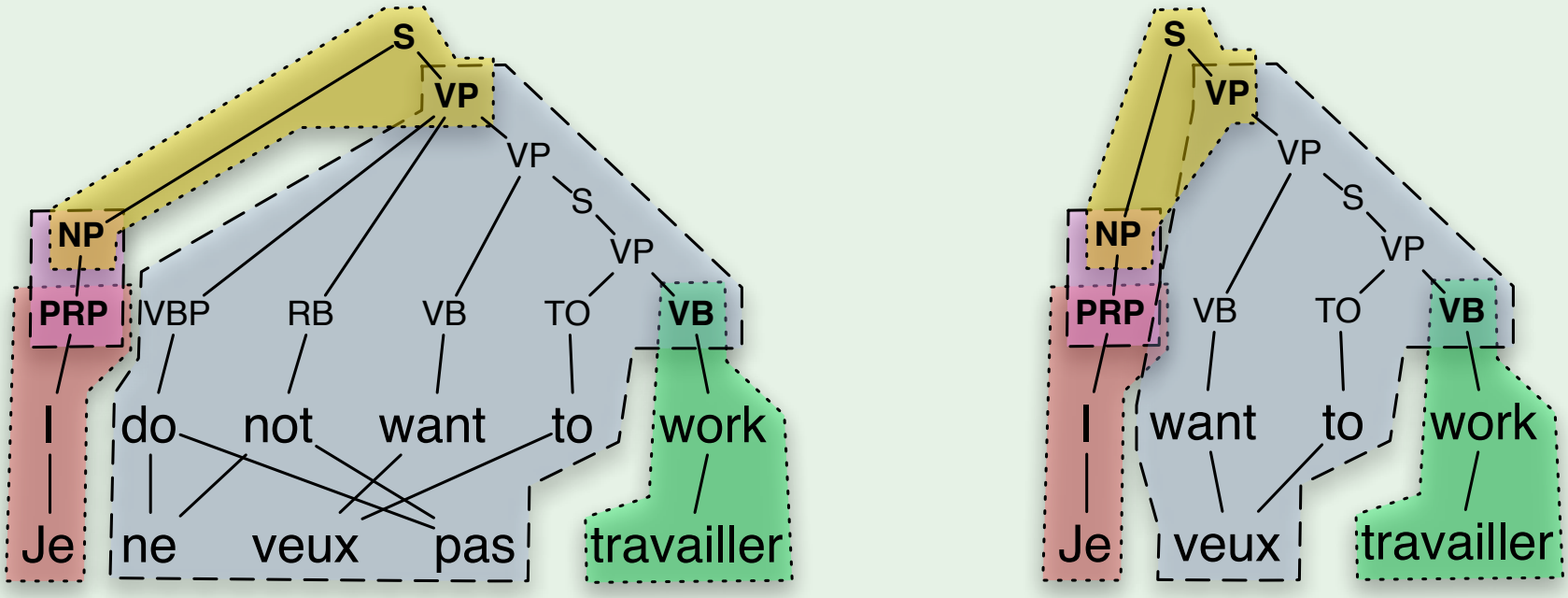
Unlabelled SCFG: Hiero



- Only requires the parallel corpus.
- But weak model of sentence structure.

Models of translation

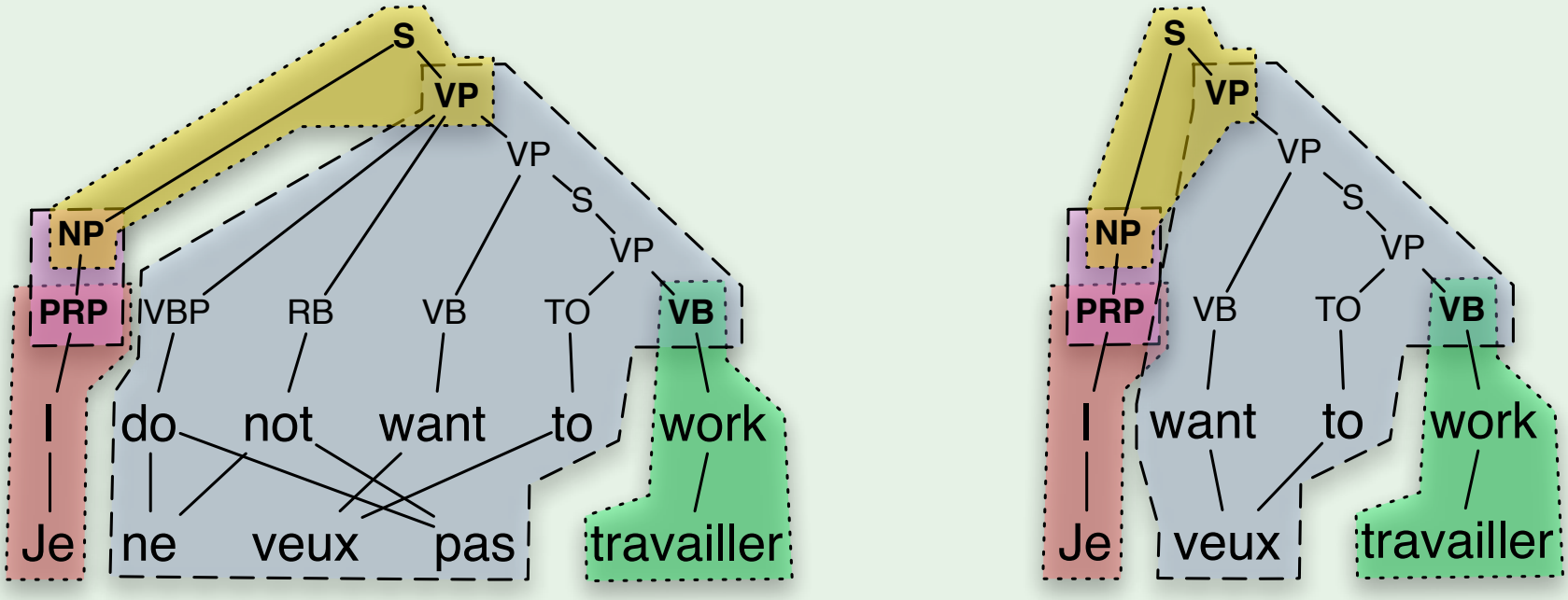
Supervised SCFG: Syntactic Tree-to-String



- S -> ⟨ $NP_{[1]}$ $VP_{[2]}$, $NP_{[1]}$ $VP_{[2]}$ ⟩,
  NP -> ⟨ $PRP_{[1]}$, $PRP_{[1]}$ ⟩
- PRP -> ⟨ Je, I ⟩, VB -> ⟨ travailler, work ⟩
  VP -> ⟨ ne veux pas $VB_{[1]}$, do not want to $VB_{[1]}$ ⟩

Models of translation

Supervised SCFG: Syntactic Tree-to-String



- Strong model of sentence structure.
- Reliant on a treebank to train the parser.

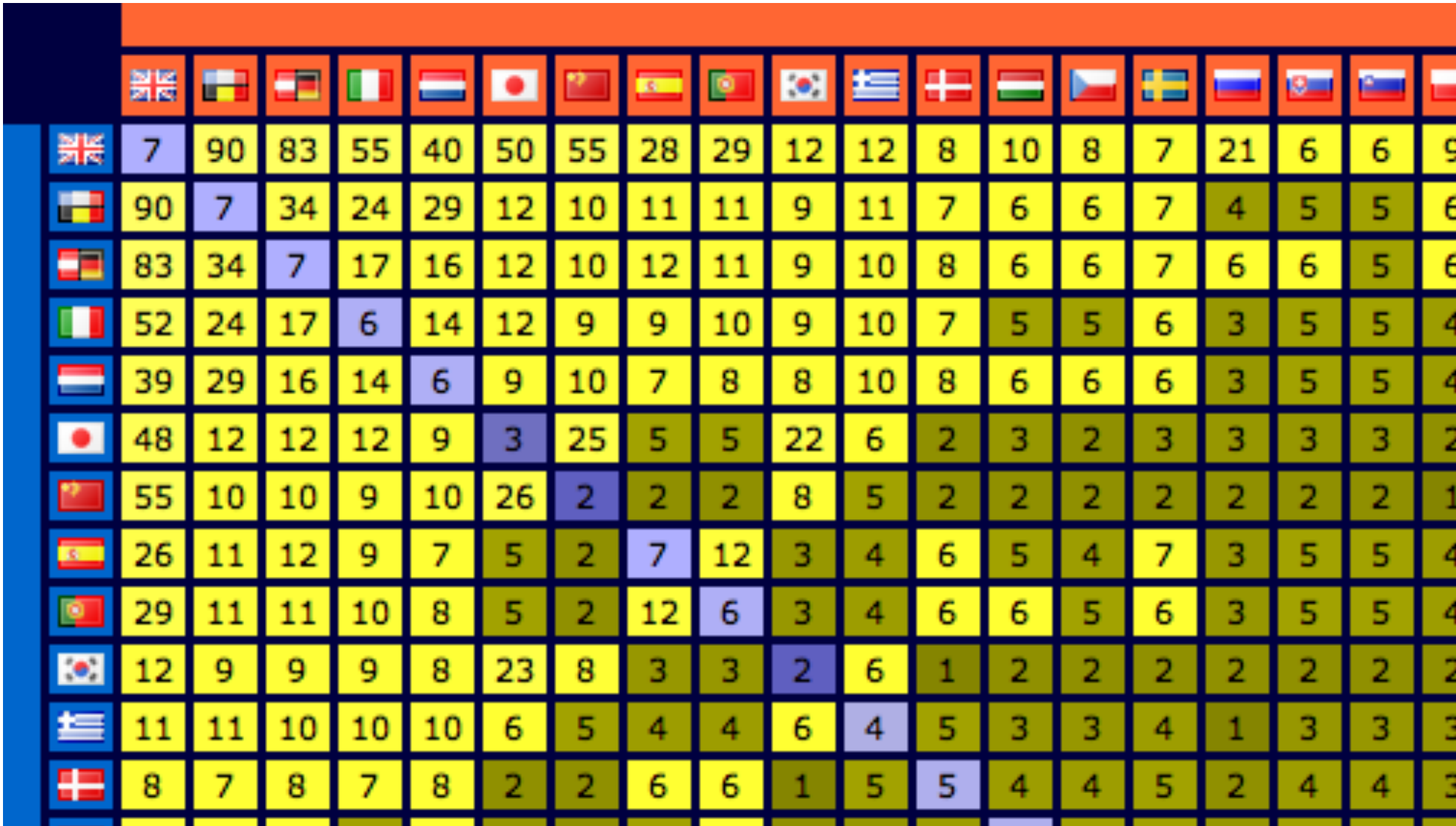
Impact

Language	Words	Domain
English	4.5M	Financial news
Chinese	0.5M	Broadcasting news
Arabic	300K (1M planned)	News
Korean	54K	Military

Table: Major treebanks: data size and domain

Impact

Parallel corpora far exceed treebanks (millions of words):



	7	90	83	55	40	50	55	28	29	12	12	8	10	8	7	21	6	6
	90	7	34	24	29	12	10	11	11	9	11	7	6	6	7	4	5	5
	83	34	7	17	16	12	10	12	11	9	10	8	6	6	7	6	6	5
	52	24	17	6	14	12	9	9	10	9	10	7	5	5	6	3	5	5
	39	29	16	14	6	9	10	7	8	8	10	8	6	6	6	3	5	5
	48	12	12	12	9	3	25	5	5	22	6	2	3	2	3	3	3	3
	55	10	10	9	10	26	2	2	2	8	5	2	2	2	2	2	2	2
	26	11	12	9	7	5	2	7	12	3	4	6	5	4	7	3	5	5
	29	11	11	10	8	5	2	12	6	3	4	6	6	5	6	3	5	5
	12	9	9	9	8	23	8	3	3	2	6	1	2	2	2	2	2	2
	11	11	10	10	10	6	5	4	4	6	4	5	3	3	4	1	3	3
	8	7	8	7	8	2	2	6	6	1	5	5	4	4	5	2	4	4

Models of translation

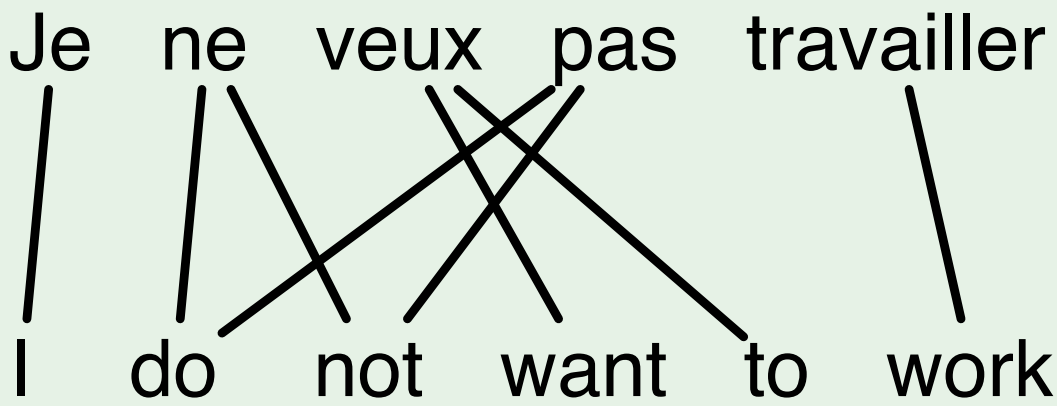
Phrase extraction:

Je ne veux pas travailler

I do not want to work

Models of translation

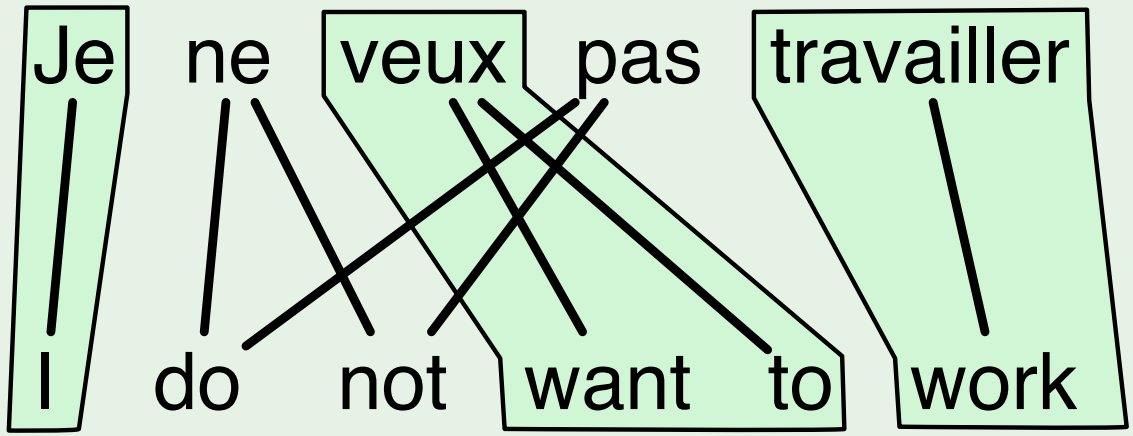
Phrase extraction:



- Use a word-based translation model to annotate the parallel corpus with word-alignments

Models of translation

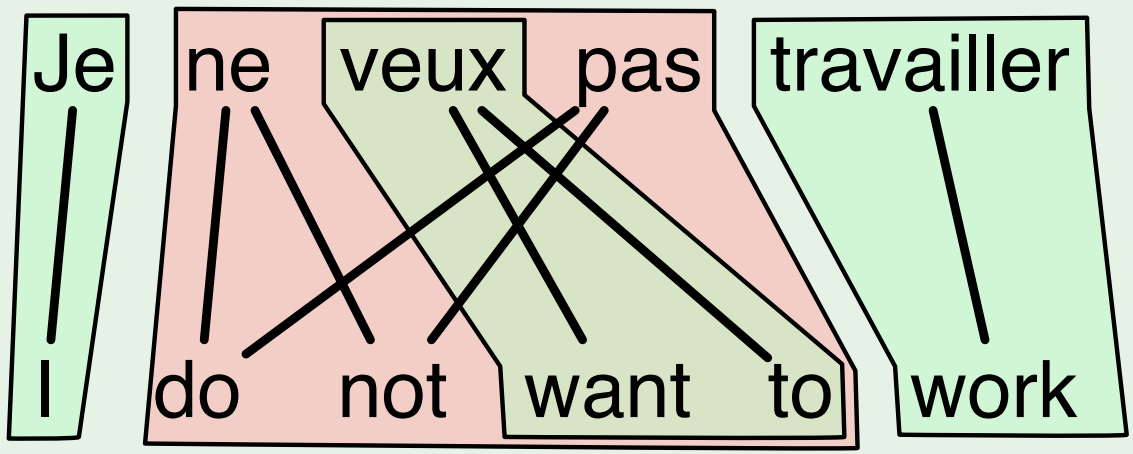
Phrase extraction:



- $\langle$ Je, I $\rangle$, $\langle$ veux, want to $\rangle$, $\langle$ travailler, work $\rangle$

Models of translation

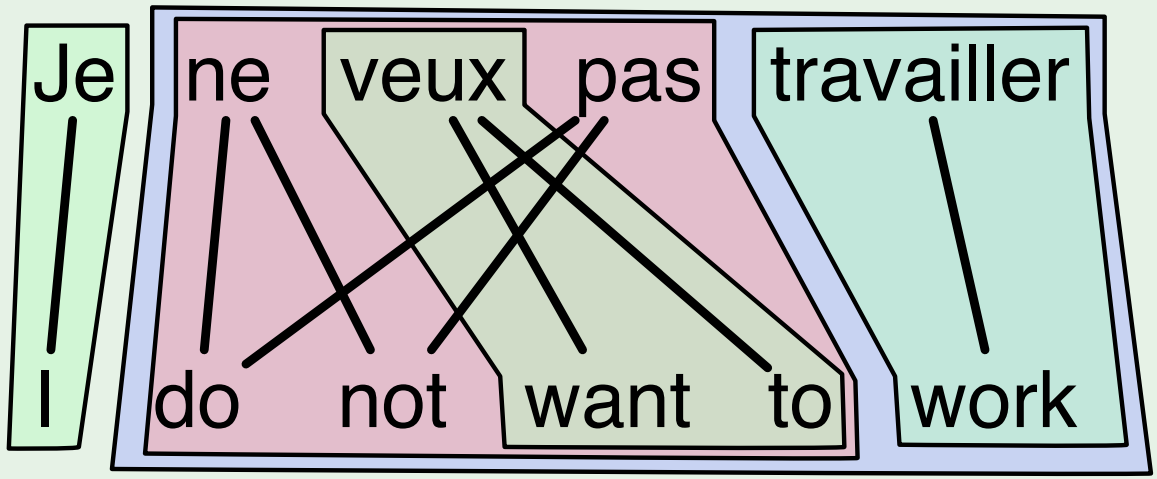
Phrase extraction:



- $\langle$ Je, I $\rangle$, $\langle$ veux, want to $\rangle$, $\langle$ travailler, work $\rangle$, $\langle$ ne veux pas, do not want to $\rangle$

Models of translation

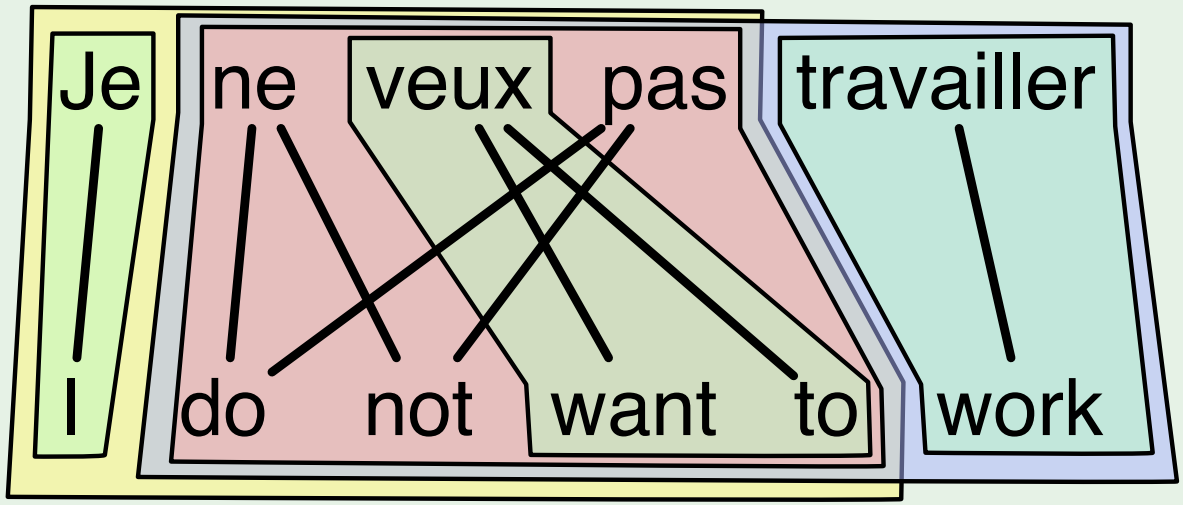
Phrase extraction:



- $\langle$ Je, I $\rangle$, $\langle$ veux, want to $\rangle$, $\langle$ travailler, work $\rangle$, $\langle$ ne veux pas, do not want to $\rangle$, $\langle$ ne veux pas travailler, do not want to work $\rangle$

Models of translation

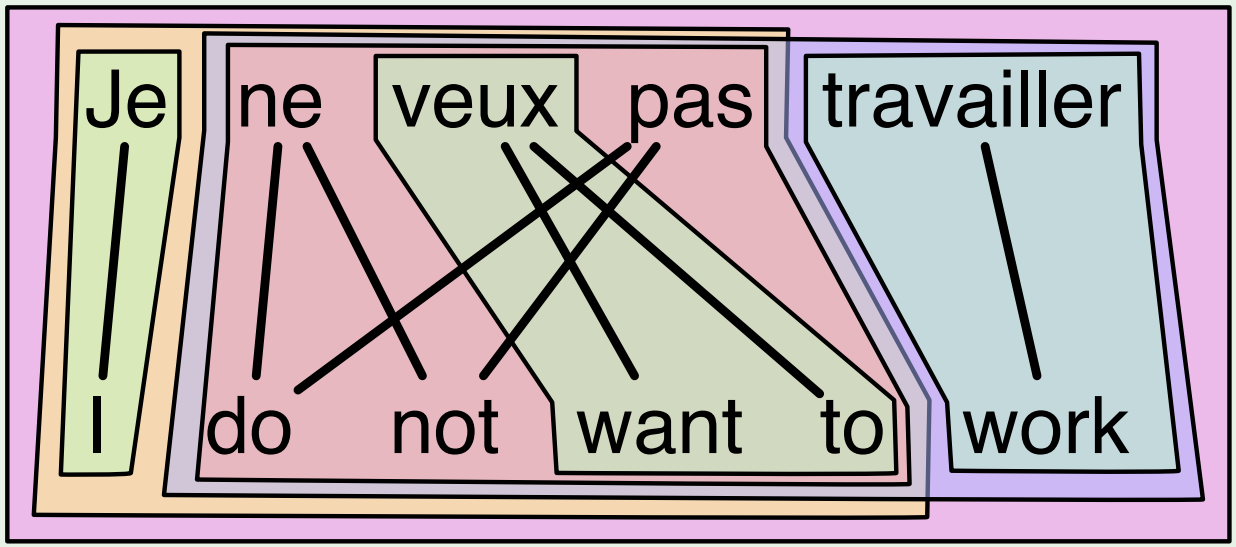
Phrase extraction:



- $\langle$ Je, I $\rangle$, $\langle$ veux, want to $\rangle$, $\langle$ travailler, work $\rangle$, $\langle$ ne veux pas, do not want to $\rangle$, $\langle$ ne veux pas travailler, do not want to work $\rangle$, $\langle$ Je ne veux pas, I do not want to $\rangle$

Models of translation

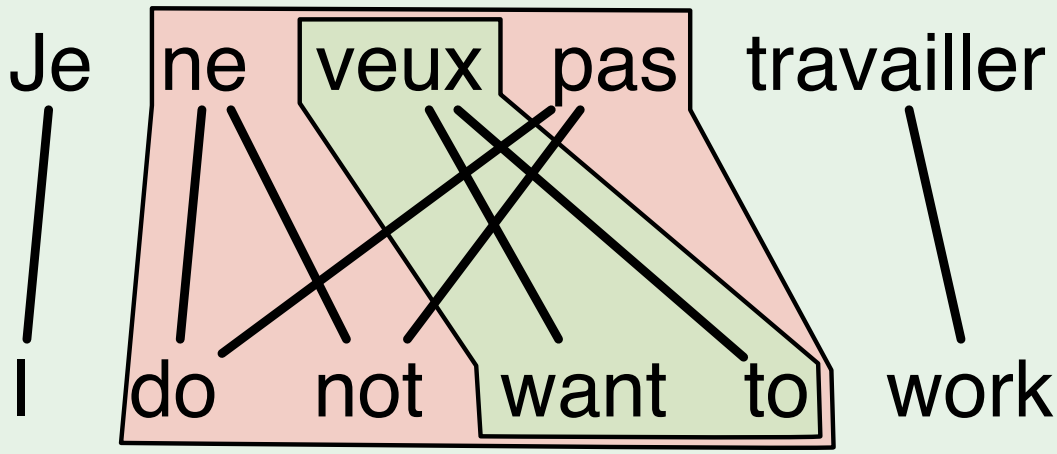
Phrase extraction:



- ⟨ Je, I ⟩, ⟨ veux, want to ⟩, ⟨ travailler, work ⟩, ⟨ ne veux pas, do not want to ⟩, ⟨ ne veux pas travailler, do not want to work ⟩, ⟨ Je ne veux pas, I do not want to ⟩, ⟨ Je ne veux pas travailler, I do not want to work ⟩

Models of translation

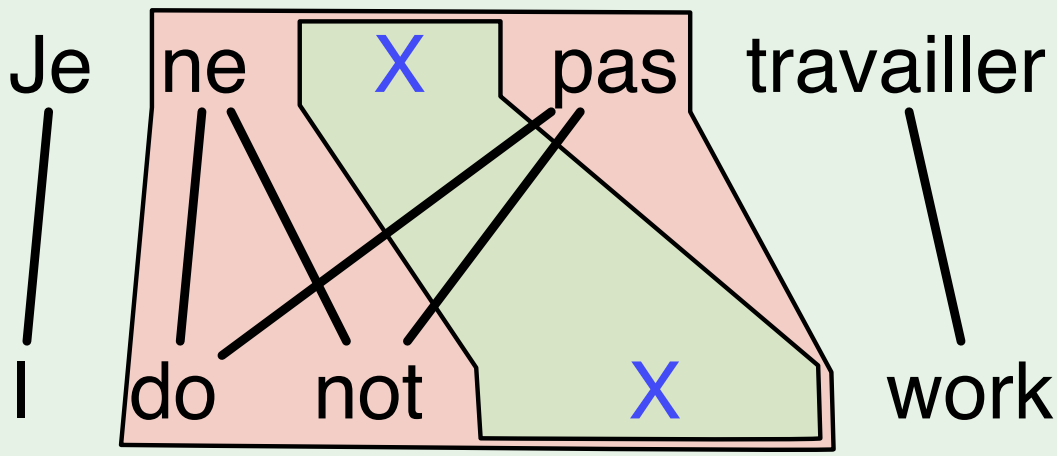
SCFG Rule extraction:



- X -> ⟨ ne veux pas, do not want to ⟩

Models of translation

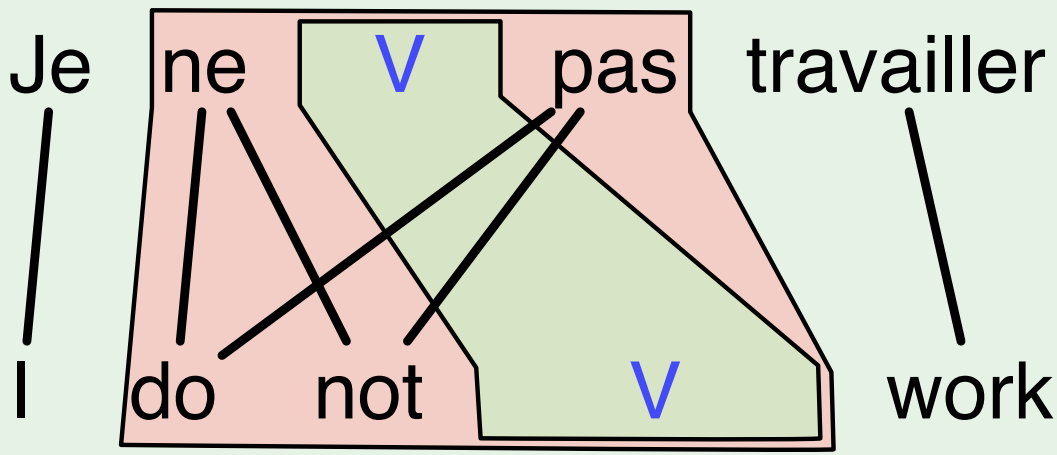
SCFG Rule extraction:



- X -> ⟨ ne veux pas, do not want to ⟩,
- X -> ⟨ ne $X_{[1]}$ pas, do not $X_{[1]}$ ⟩

Models of translation

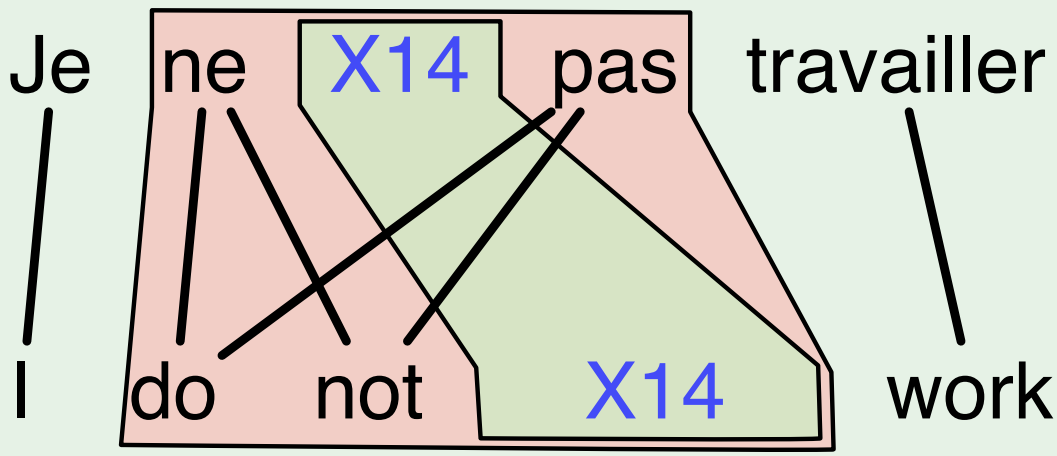
SCFG Rule extraction:



- VP/NN -> ⟨ ne veux pas, do not want to ⟩,
- VP/NN -> ⟨ ne $V_{\boxed{1}}$ pas, do not $V_{\boxed{1}}$ ⟩

Models of translation

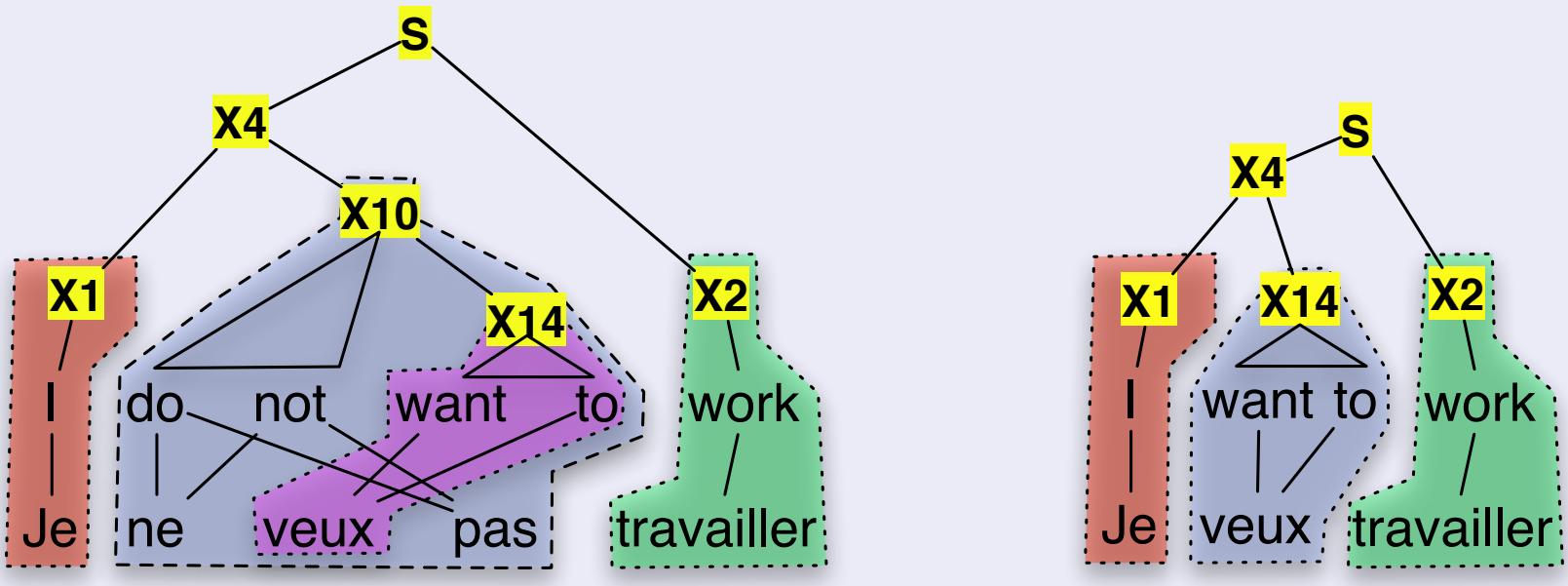
SCFG Rule extraction:



- X10 -> ⟨ ne veux pas, do not want to ⟩,
- X10 -> ⟨ ne $X14_{\boxed{1}}$ pas, do not $X14_{\boxed{1}}$ ⟩

Models of translation

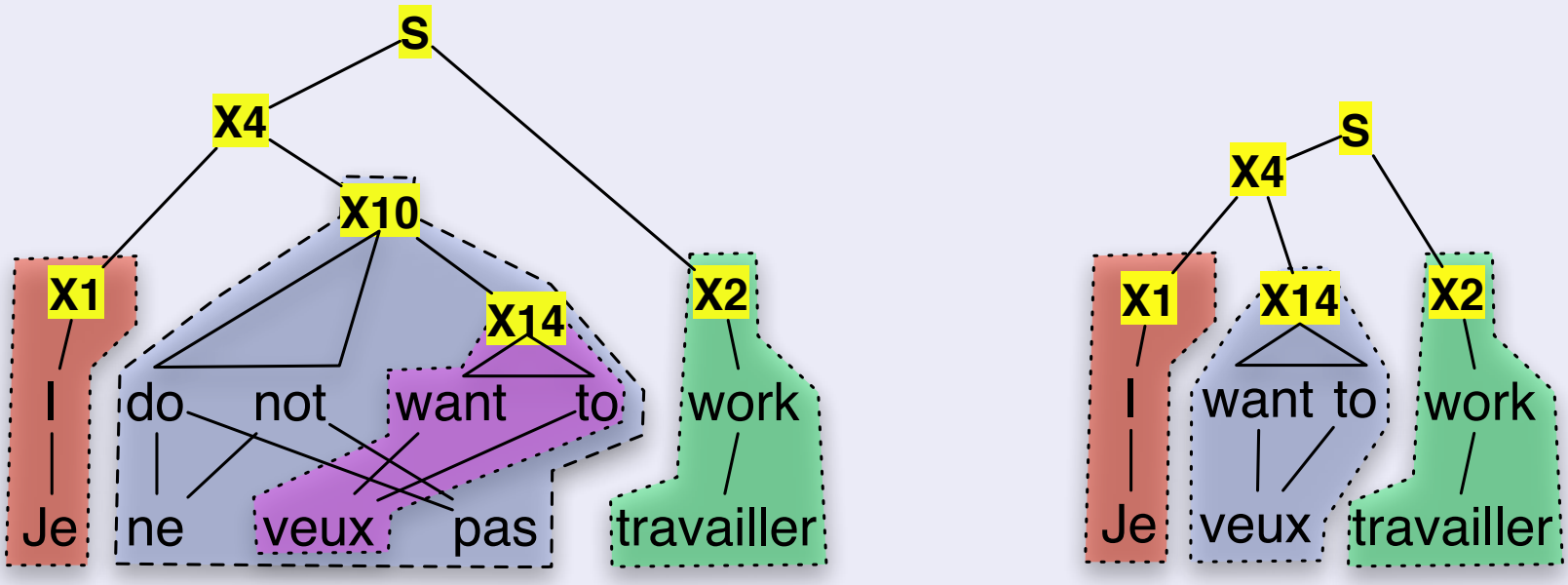
This workshop



- S -> ⟨ $X4_{[1]}$ $X2_{[2]}$, $X4_{[1]}$ $X2_{[2]}$ ⟩, X4 -> ⟨ $X1_{[1]}$ $X10_{[2]}$, $X1_{[1]}$ $X10_{[2]}$ ⟩
- X1 -> ⟨ Je, I ⟩, X10 -> ⟨ ne $X14_{[1]}$ pas, do not $X14_{[1]}$ ⟩, X14 -> ⟨ veux, want to⟩, X10 -> ⟨ travailler, work ⟩

Models of translation

This workshop



- Only requires the parallel corpus.
- But also gives a strong model of sentence structure.

Workshop overview

Input:

- Existing procedures for unlabelled synchronous grammar extraction

Output:

- New unsupervised models for large scale synchronous grammar extraction,
- A comparison and analysis of the existing and proposed models,
- Extended decoders (cdec/Joshua) capable of working efficiently with these models.

Workshop Streams

1. Implement scalable labelled SCFG grammar induction algorithms:
   - by clustering translation phrases which occur in the same context we can learn which phrases are substituteable,
   - we have implemented both parametric and non-parametric Bayesian clustering algorithms.
2. Improve SCFG decoders to efficiently handle the grammars produced:
   - translation complexity scales quadratically as we add more categories,
   - in order to decode efficiently with the grammars we've induced we have created faster search algorithms tuned for syntactic grammars.
3. Investigate discriminative training regimes to leverage features extracted from these grammars:
   - to make the most of our induced grammars we need discriminative training algorithms that learn from more than a handful of features,
   - we've implemented two large scale discriminative algorithms for training our models.

Extrinsic evaluation: Bleu

Ngram overlap metrics:

Source: 欧盟 办事处 与 澳洲 大使馆 在 同 一 建筑 内

Candidate: the chinese embassy in australia and the eu representative office in the same building

Reference Translations:

1. the eu office and the australian embassy are housed in the same building
2. the european union office is in the same building as the australian embassy
3. the european union 's office and the australian embassy are both located in the same building
4. the eu 's mission is in the same building with the australian embassy

Extrinsic evaluation: Bleu

Ngram overlap metrics: 1-gram precision $p_1 = \frac{11}{14}$

Source: 欧盟 办事处 与 澳洲 大使馆 在 同 一 建筑 内

Candidate: the chinese embassy in australia and the eu representative office in the same building

Reference Translations:

1. the eu office and the australian embassy are housed in the same building
2. the european union office is in the same building as the australian embassy
3. the european union 's office and the australian embassy are both located in the same building
4. the eu 's mission is in the same building with the australian embassy

Extrinsic evaluation: Bleu

Ngram overlap metrics: 2-gram precision $p_2 = \frac{5}{13}$

Source: 欧盟 办事处 与 澳洲 大使馆 在 同 一 建筑 内

Candidate: the chinese embassy in australia and the eu representative office in the same building

Reference Translations:

1. the eu office and the australian embassy are housed in the same building
2. the european union office is in the same building as the australian embassy
3. the european union 's office and the australian embassy are both located in the same building
4. the eu 's mission is in the same building with the australian embassy

Extrinsic evaluation: Bleu

Ngram overlap metrics: 3-gram precision $p_3 = \frac{2}{12}$

Source: 欧盟 办事处 与 澳洲 大使馆 在 同 一 建筑 内

Candidate: the chinese embassy in australia and the eu representative office in the same building

Reference Translations:

1. the eu office and the australian embassy are housed in the same building
2. the european union office is in the same building as the australian embassy
3. the european union 's office and the australian embassy are both located in the same building
4. the eu 's mission is in the same building with the australian embassy

Extrinsic evaluation: Bleu

Ngram overlap metrics: 4-gram precision $p_4 = \frac{1}{11}$

Source: 欧盟 办事处 与 澳洲 大使馆 在 同 一 建筑 内

Candidate: the chinese embassy in australia and the eu representative office in the same building

Reference Translations:

1. the eu office and the australian embassy are housed in the same building
2. the european union office is in the same building as the australian embassy
3. the european union 's office and the australian embassy are both located in the same building
4. the eu 's mission is in the same building with the australian embassy

Extrinsic evaluation: Bleu

BLEU

$$BLEU_n = BP \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

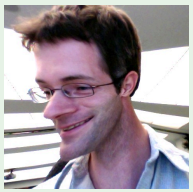
$$BP = \begin{cases} 1 & \text{if } c > r \\ \exp\left(1 - \frac{R'}{C'}\right) & \text{if } c <= r \end{cases}$$

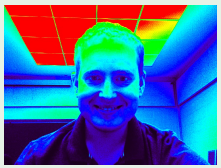
- BP is the *Brevity Penalty*, w_n is the ngram length weights (usually $\frac{1}{n}$), p_n is precision of ngram predictions, R' is the total length of all references and C' is the sum of the best matching candidates.
- statistics are calculate over the whole *document*, i.e. all the sentences.

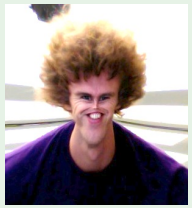
Language pairs

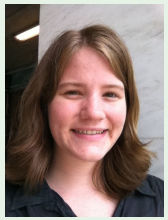
- BTEC Chinese-English:
  - 44k sentence pairs, short sentences
  - Widely reported 'prototyping' corpus
  - Hiero baseline score: 57.0 (16 references)
- NIST Urdu-English:
  - 50k sentence pairs
  - Hiero baseline score: 21.1 (4 references)
  - Major challenges: major long-range reordering, SOV word order
- Europarl Dutch-French:
  - 100k sentence pairs, standard Europarl test sets
  - Hiero baseline score: Europarl 2008 - 15.75 (1 reference)
  - Major challenges: V2 / V-final word order, morphology

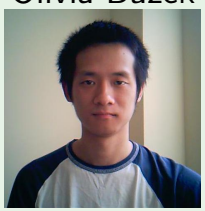
Outline


Trevor Cohn


Chris Dyer


Jan Botha


Olivia Buzek


Desai Chen

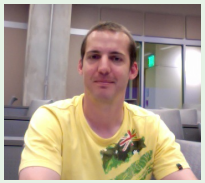
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

Outline

Vlad Eidelman

Ziyuan Wang

Adam Lopez


Jon Graehl

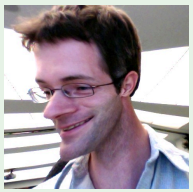
ThuyLinh Nguyen

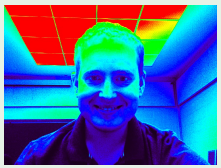
- 3:15pm Training models with rich features spaces. Vlad
- 3:30pm Decoding with complex grammars. Adam
- 4:00pm Closing remarks. Phil
- 4:05pm Finish.

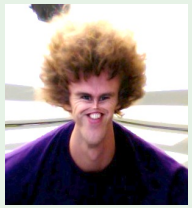
Remember:

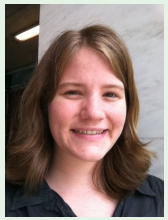
- Idea: Learn synchronous grammar labels which encode substituteability; phrases which occur in the same context should receive the same label.
- Result: Better models of translation structure, morphology and improved decoding algorithms.

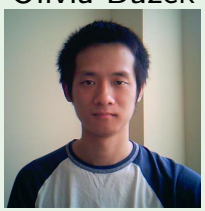
Outline


Trevor Cohn


Chris Dyer

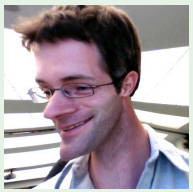

Jan Botha

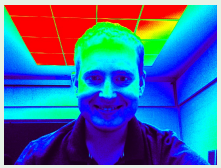

Olivia Buzek

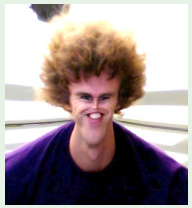

Desai Chen

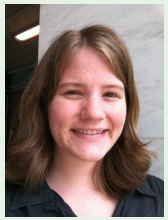
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

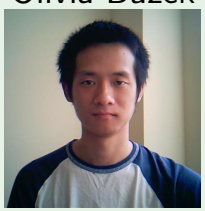
Outline


Trevor Cohn


Chris Dyer

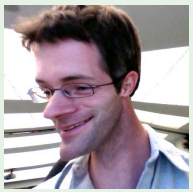

Jan Botha

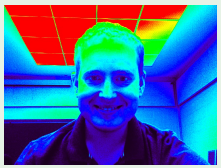

Olivia Buzek

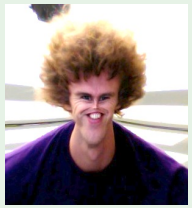

Desai Chen

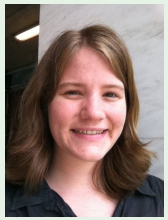
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

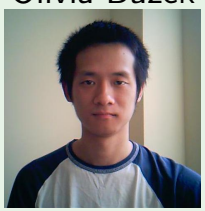
Outline


Trevor Cohn


Chris Dyer

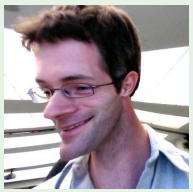

Jan Botha

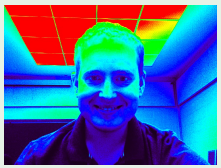

Olivia Buzek

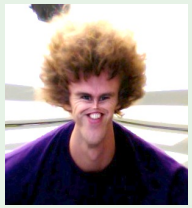

Desai Chen

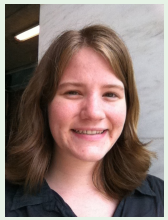
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

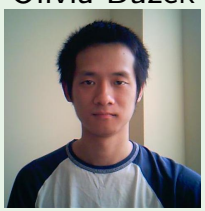
Outline


Trevor Cohn


Chris Dyer

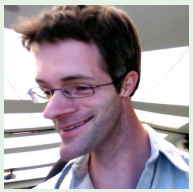

Jan Botha

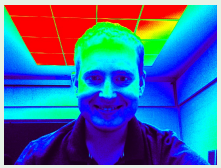

Olivia Buzek

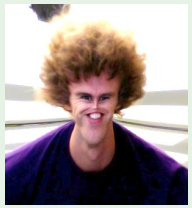

Desai Chen

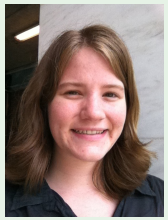
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

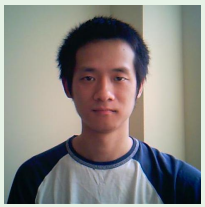
Outline


Trevor Cohn


Chris Dyer

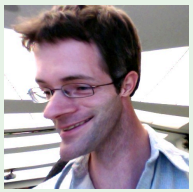

Jan Botha

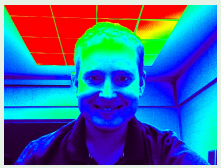

Olivia Buzek

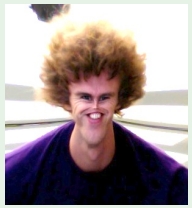

Desai Chen

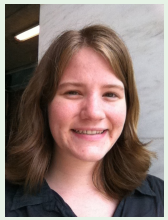
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

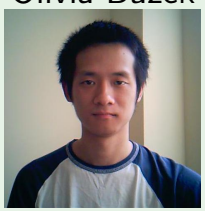
Outline


Trevor Cohn


Chris Dyer


Jan Botha


Olivia Buzek

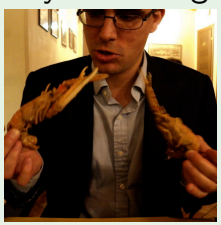

Desai Chen

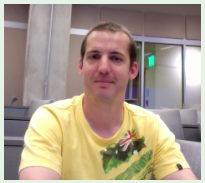
- 1:55pm Grammar induction and evaluation. Trevor
- 2:10pm Non-parametric models of category induction. Chris
- 2:25pm Inducing categories for morphology. Jan
- 2:35pm Smoothing, backoff and hierarchical grammars. Olivia
- 2:45pm Parametric models: posterior regularisation. Desai
- 3:00pm Break.

Outline

Vlad Eidelman

Ziyuan Wang


Adam Lopez


Jon Graehl

ThuyLinh Nguyen

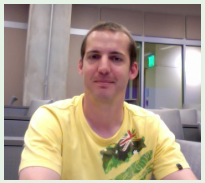
- 3:20pm Parametric models: posterior regularisation. Desai
- 3:35pm Training models with rich features spaces. Vlad
- 3:50pm Decoding with complex grammars. Adam
- 4:20pm Closing remarks. Phil
- 4:25pm Finish.

Outline

Vlad Eidelman

Ziyuan Wang

Adam Lopez


Jon Graehl

ThuyLinh Nguyen

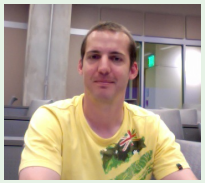
- 3:20pm Parametric models: posterior regularisation. Desai
- 3:35pm Training models with rich features spaces. Vlad
- 3:50pm Decoding with complex grammars. Adam
- 4:20pm Closing remarks. Phil
- 4:25pm Finish.

Outline

Vlad Eidelman

Ziyuan Wang

Adam Lopez


Jon Graehl

ThuyLinh Nguyen

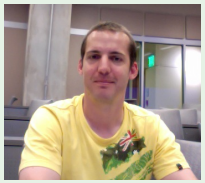
- 3:20pm Parametric models: posterior regularisation. Desai
- 3:35pm Training models with rich features spaces. Vlad
- 3:50pm Decoding with complex grammars. Adam
- 4:20pm Closing remarks. Phil
- 4:25pm Finish.

Outline

Vlad Eidelman

Ziyuan Wang

Adam Lopez


Jon Graehl

ThuyLinh Nguyen

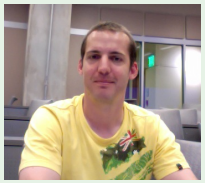
- 3:20pm Parametric models: posterior regularisation. Desai
- 3:35pm Training models with rich features spaces. Vlad
- 3:50pm Decoding with complex grammars. Adam
- 4:20pm Closing remarks. Phil
- 4:25pm Finish.

Outline

Vlad Eidelman

Ziyuan Wang

Adam Lopez


Jon Graehl

ThuyLinh Nguyen

- 3:20pm Parametric models: posterior regularisation. Desai
- 3:35pm Training models with rich features spaces. Vlad
- 3:50pm Decoding with complex grammars. Adam
- 4:20pm Closing remarks. Phil
- 4:25pm Finish.