Rule Extraction for Machine Translation

Andreas Maletti

Institute of Computer Science
Universität Leipzig, Germany

on leave from: Institute for Natural Language Processing
Universität Stuttgart, Germany

maletti@ims.uni-stuttgart.de

Darmstadt — September 10, 2014

Main notions

Machine translation (MT)

Automatic natural language *translation* (by a computer)

as opposed to:

- manual translation
- computer-aided translation (e.g., translation memory)

Main notions

Machine translation (MT)

Automatic natural language *translation* (by a computer)

as opposed to:

- manual translation
- computer-aided translation (e.g., translation memory)

Statistical machine translation (SMT)

MT using systems *automatically* obtained from *translations*

as opposed to:

- rule-based machine translation (old) SYSTRAN
- example-based machine translation translation by analogy

Short history

Timeline

1. Dark age (60s–90s)
   - rule-based systems (e.g., SYSTRAN)
   - CHOMSKYAN approach
   - perfect translation, poor coverage

Short history

Timeline

1. Dark age (60s–90s)
   - rule-based systems (e.g., SYSTRAN)
   - CHOMSKYAN approach
   - perfect translation, poor coverage
2. Reformation (1991–present)
   - phrase-based and syntax-based systems
   - statistical approach
   - cheap, automatically trained

Short history

Timeline

1. Dark age (60s–90s)
   - rule-based systems (e.g., SYSTRAN)
   - CHOMSKYAN approach
   - perfect translation, poor coverage
2. Reformation (1991–present)
   - phrase-based and syntax-based systems
   - statistical approach
   - cheap, automatically trained
3. Potential future
   - semantics-based systems (e.g., FRAMENET-based)
   - semi-supervised, statistical approach
   - basic understanding of translated text

Examples

Applications

- Technical manuals

Example (An mp3 player)

The synchronous manifestation of lyrics is a procedure for can broadcasting the music, waiting the mp3 file at the same time showing the lyrics.

Examples

Applications

- Technical manuals

Example (An mp3 player)

With the this kind method that the equipments that synchronous function of support up broadcast to make use of document create setup, you can pass the LCD window way the check at the document contents that broadcast.

Examples

Applications

- Technical manuals

Example (An mp3 player)

That procedure returns offerings to have to modify, and delete, and stick top , keep etc. edit function.

Examples

Applications

- Technical manuals
- 

Example (Hotel Uppsala, Sweden)

Wir hatten die Zimmer eingestuft wird als “Superior” weil sie renoviert wurde im letzten Jahr oder zwei. Unsere Zimmer hatten Parkettboden und waren sehr geräumig. Man musste allerdings nicht musste seitwärts bewegen.

Examples

Applications

- Technical manuals
- 

Example (Hotel Uppsala, Sweden)

Wir hatten die Zimmer eingestuft wird als “Superior” weil sie renoviert wurde im letzten Jahr oder zwei. Unsere Zimmer hatten Parkettboden und waren sehr geräumig. Man musste allerdings nicht musste seitwärts bewegen.

— *We stayed in rooms classified as “superior” because they had been renovated in the last year or two. Our rooms had wood floors and were roomy. You didn’t have to walk sideways to move around.*

Examples

Applications

- Technical manuals
- tripadvisor
- US military

Example (JONES, SHEN, HERZOG 2009)

Soldier:	Okay, what is your name?
Local:	Abdul.
Soldier:	And your last name?
Local:	Al Farran.

Examples

Applications

- Technical manuals
- tripadvisor
- US military

Example (JONES, SHEN, HERZOG 2009)

Soldier: Okay, what is your name?
Local: Abdul.
Soldier: And your last name?
Local: Al Farran.

Speech-to-text machine translation

Soldier: Okay, what's your name?
Local: milk a mechanic and I am here I mean yes

Examples

Applications

- Technical manuals
- tripadvisor
- US military

Example (JONES, SHEN, HERZOG 2009)

Soldier: Okay, what is your name?
Local: Abdul.
Soldier: And your last name?
Local: Al Farran.

Speech-to-text machine translation

Soldier: Okay, what's your name?
Local: milk a mechanic and I am here
I mean yes
Soldier: What is your last name?
Local: every two weeks
my son's name is ismail

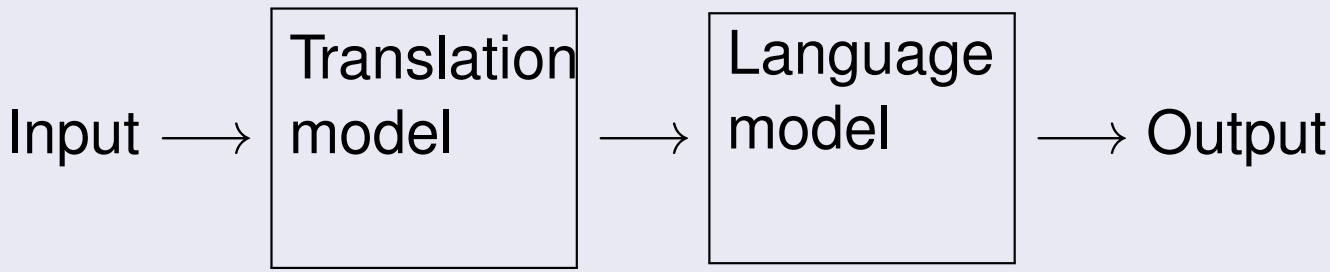
Examples

Applications

- Technical manuals
- 
- US military
- MSDN, Knowledge Base
- . . .

Standard pipeline

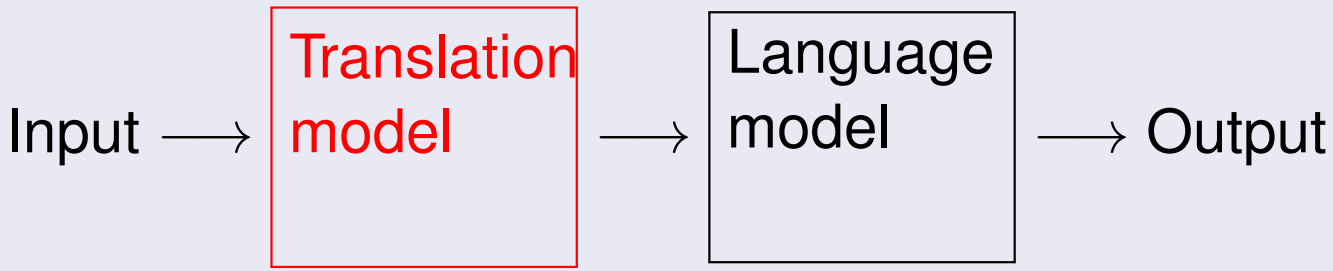
Schema



(the models are often integrated in practice)

Standard pipeline

Schema



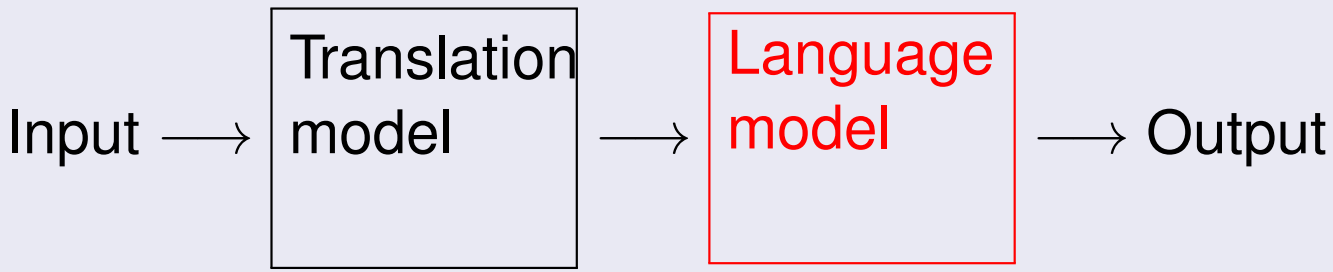
(the models are often integrated in practice)

Required resources

- bilingual text (sentences in both languages) 1.5M sent.

Standard pipeline

Schema



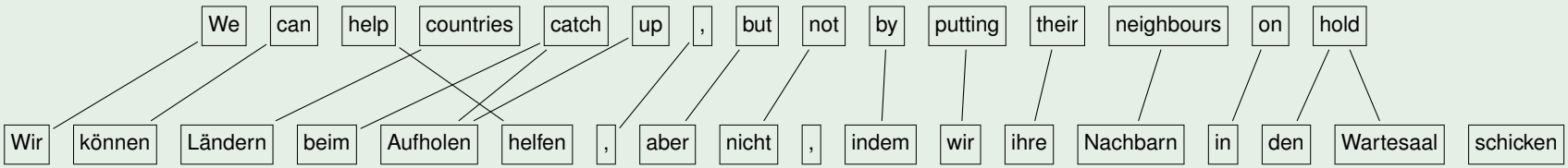
(the models are often integrated in practice)

Required resources

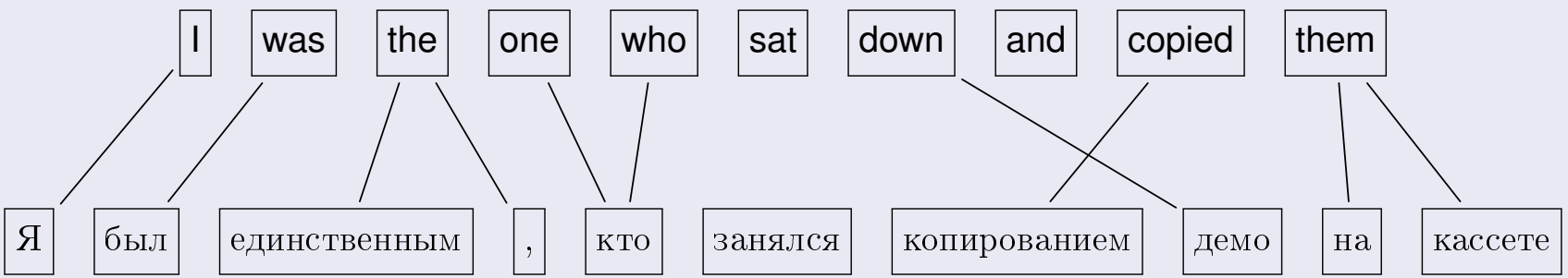
- bilingual text (sentences in both languages) 1.5M sent.
- monolingual text (in target language) 44M sent.

Word Alignment

English-German example



English-Russian example

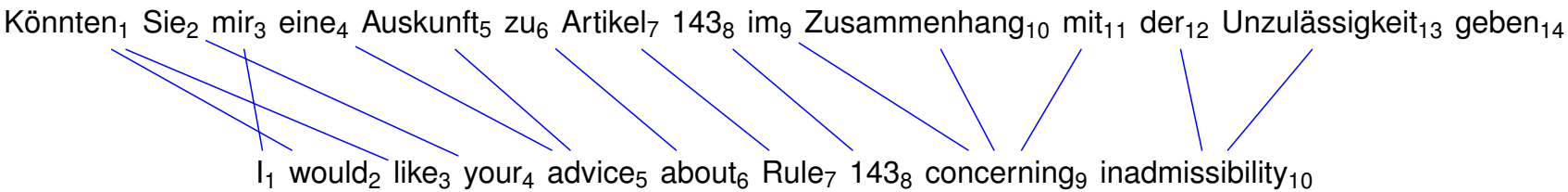


Parallel Corpus

EUROPARL German-English parallel corpus

- 1,920,209 parallel sentences
- 44,548,491 words in German
- 47,818,827 words in English
- sentence-aligned, but not word-aligned
- from parliament proceedings

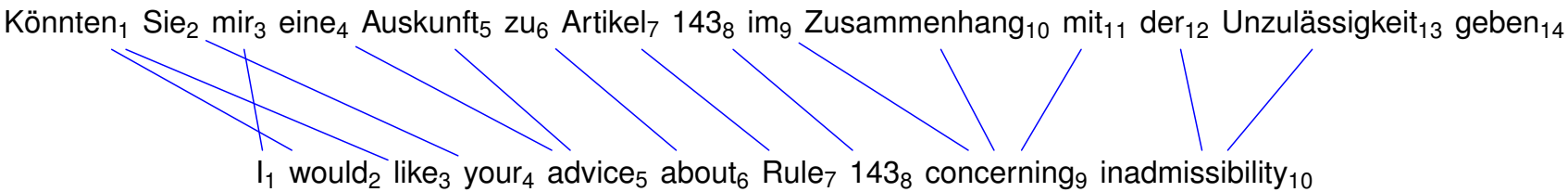
Phrase-based Models



Algorithm

1. phrase pair $([j, j'], [i, i'])$ consistently aligned if
   - $\ell' \in [i, i']$ for all $\ell \in [j, j']$ and $(\ell, \ell') \in A$
   - $\ell \in [j, j']$ for all $\ell' \in [i, i']$ and $(\ell, \ell') \in A$

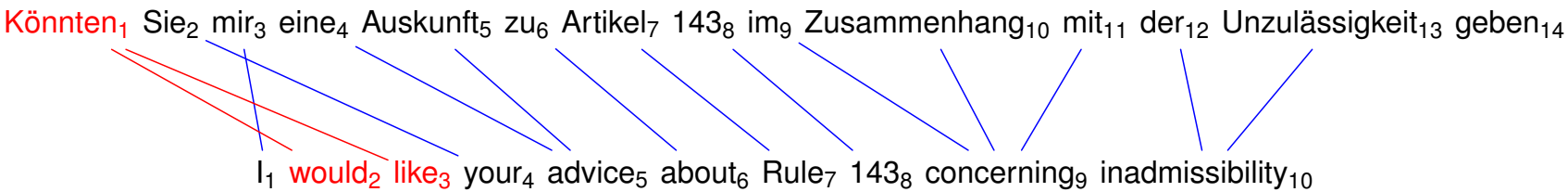
Phrase-based Models



Algorithm

1. phrase pair $([j, j'], [i, i'])$ consistently aligned if
   - $\ell' \in [i, i']$ for all $\ell \in [j, j']$ and $(\ell, \ell') \in A$
   - $\ell \in [j, j']$ for all $\ell' \in [i, i']$ and $(\ell, \ell') \in A$
2. extract all consistently aligned phrase pairs
3. (restrict length of phrases based on corpus size)

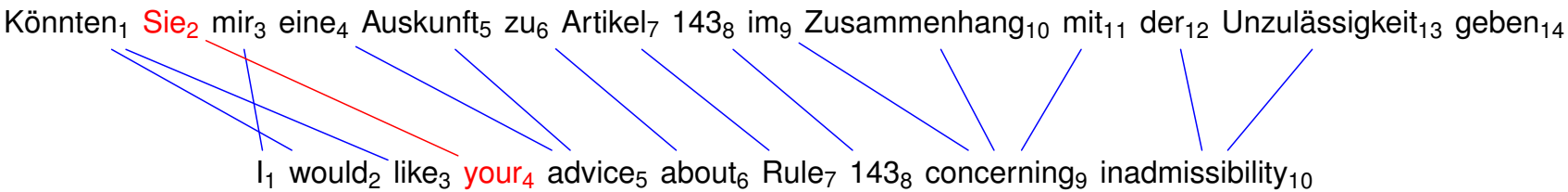
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

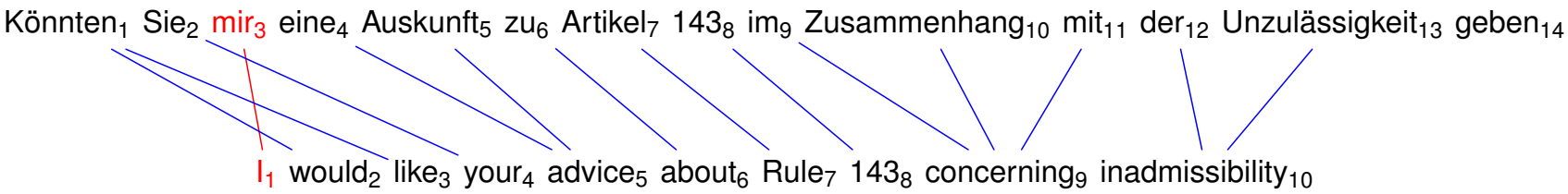
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

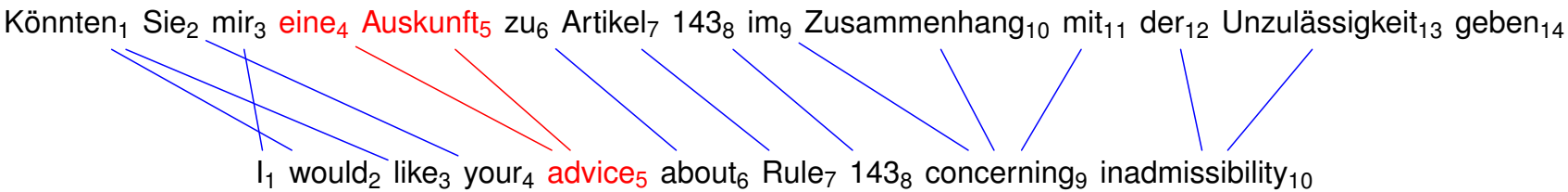
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

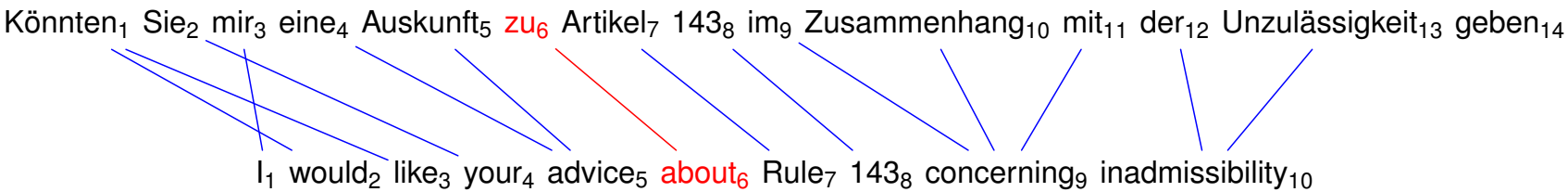
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

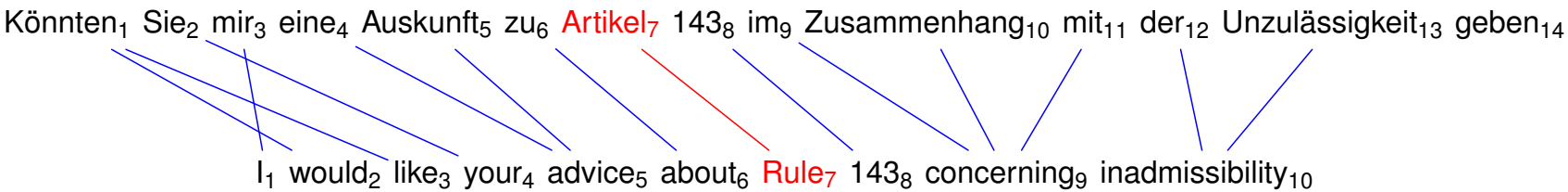
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

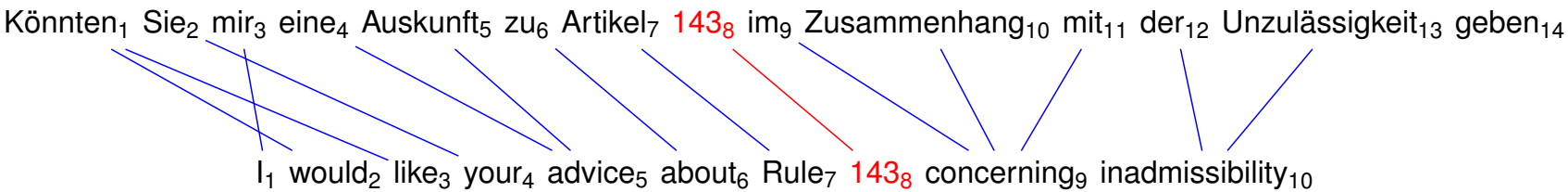
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

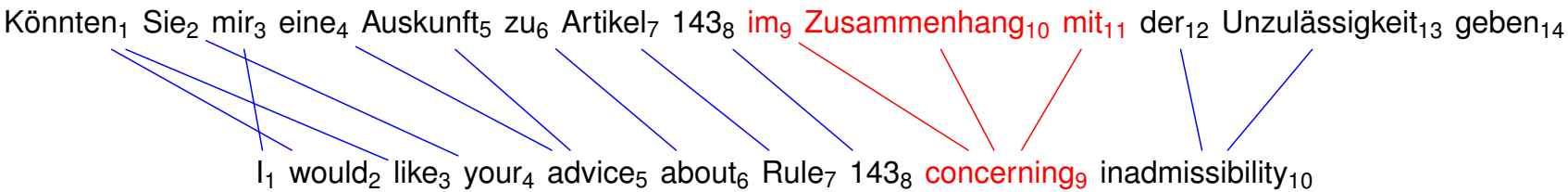
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

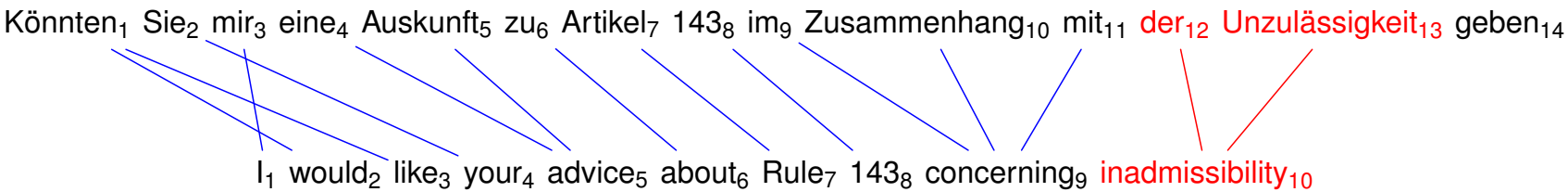
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

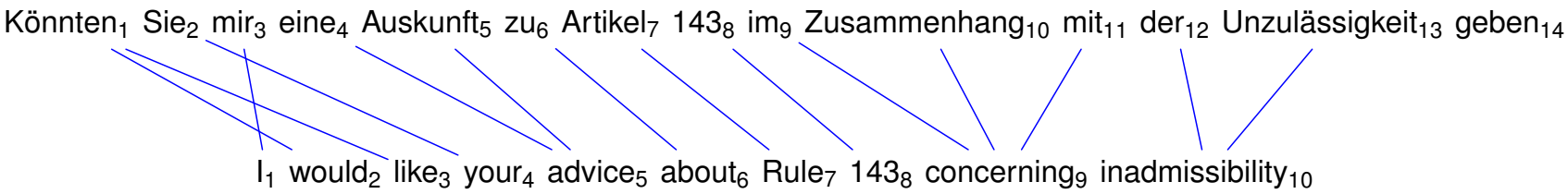
Phrase-based Models



Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

Phrase-based Models



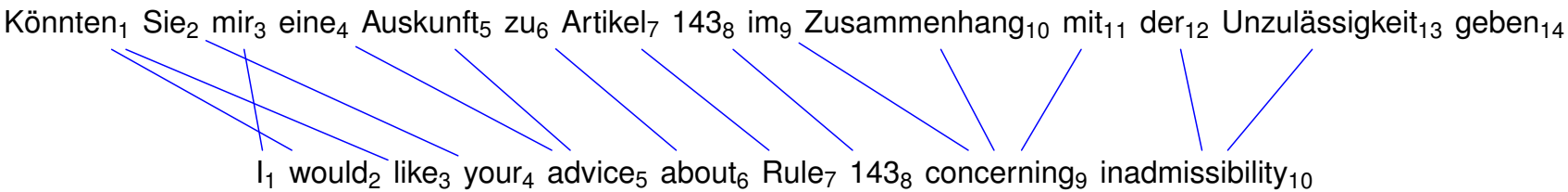
Formally:

([1,1], [2,3])	([2,2], [4,4])	([3,3], [1,1])
([4,5], [5,5])	([6,6], [6,6])	([7,7], [7,7])
([8,8], [8,8])	([9,11], [9,9])	([12,13], [10,10])

For better readability:

Könnten — would like	Sie — your	mir — I
eine Auskunft — advice	zu — about	Artikel — Rule
143 — 143	im Zusammenhang mit — concerning	der Unzulässigkeit — inadmissibility

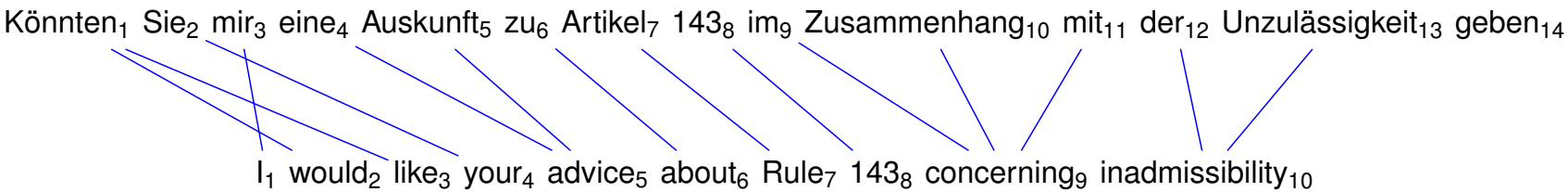
Phrase-based Models



Notes

- these were only minimal phrase pairs
- extract all (sensible) combinations of these
- e.g., $([1,1],[2,3])$ and $([2,2],[4,4])$ yield $([1,2],[2,4])$

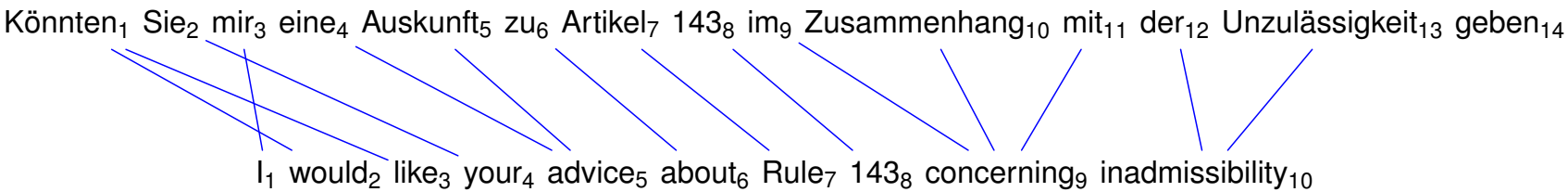
Phrase-based Models



Notes

- these were only minimal phrase pairs
- extract all (sensible) combinations of these
- e.g., $([1,1],[2,3])$ and $([2,2],[4,4])$ yield $([1,2],[2,4])$
- unaligned words can be added to neighboring phrases
- e.g., $([12,13],[10,10])$ extends to $([12,14],[10,10])$

Phrase-based Models

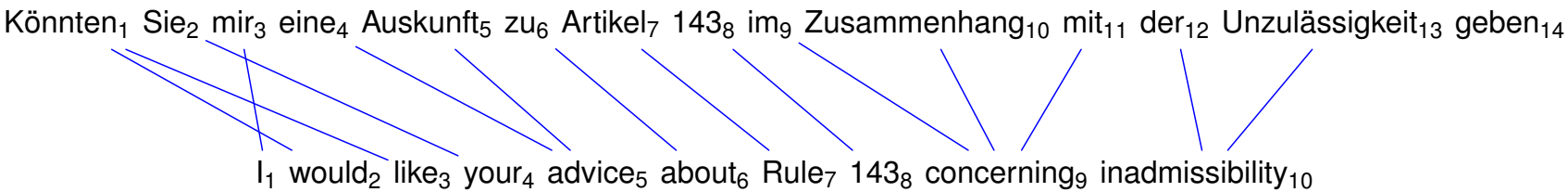


Notes

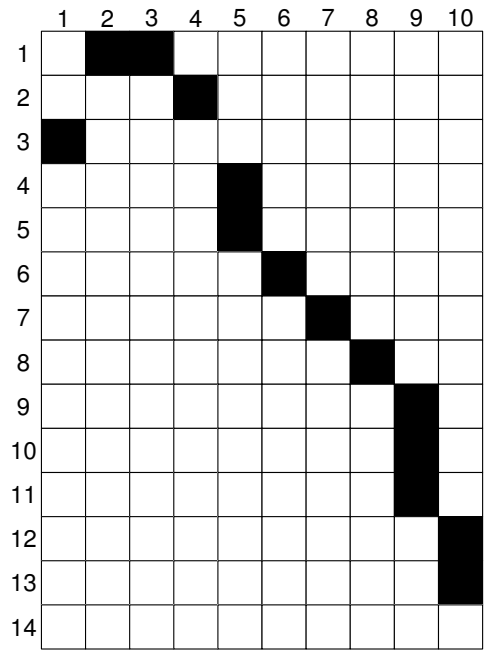
- these were only minimal phrase pairs
- extract all (sensible) combinations of these
- e.g., $([1,1],[2,3])$ and $([2,2],[4,4])$ yield $([1,2],[2,4])$
- unaligned words can be added to neighboring phrases
- e.g., $([12,13],[10,10])$ extends to $([12,14],[10,10])$

Könnten Sie — would like your der Unzulässigkeit geben — inadmissibility

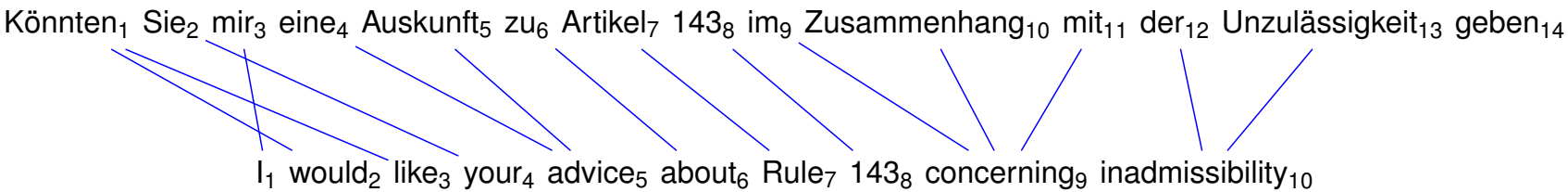
Phrase-based Models



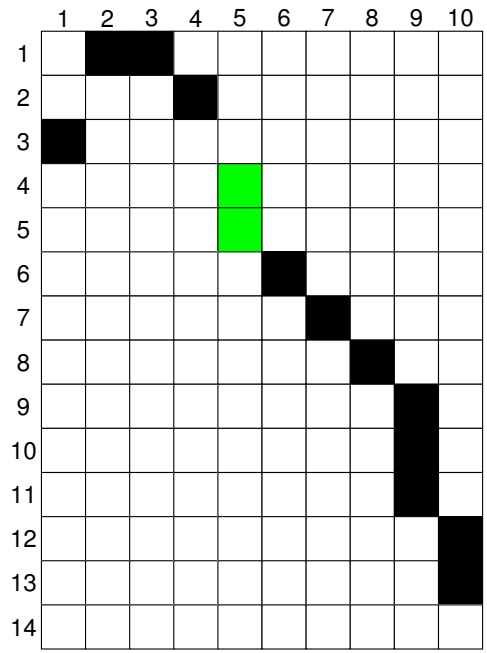
Alternative representation (rectangles):



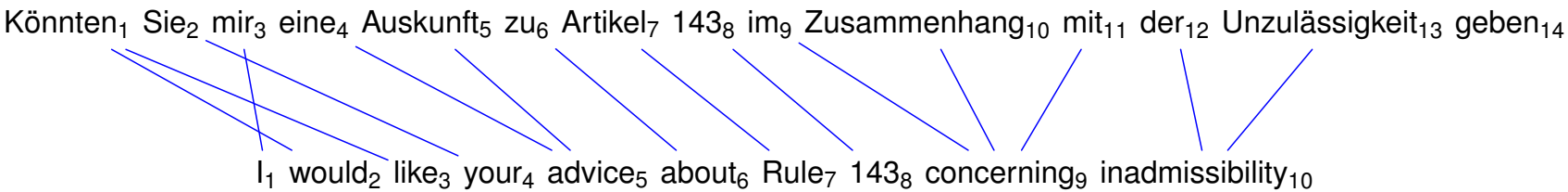
Phrase-based Models



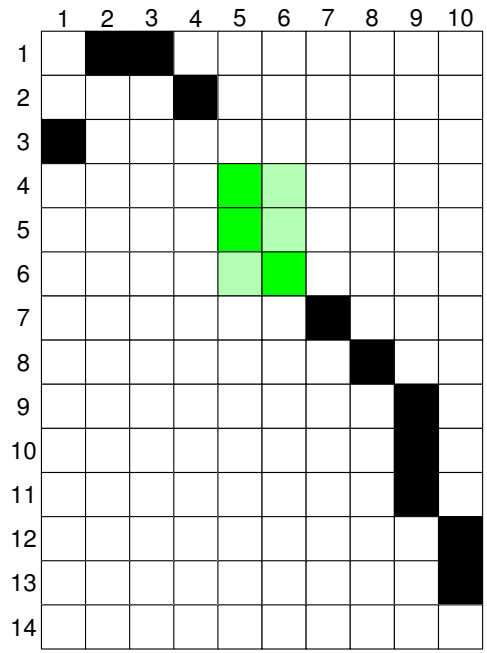
Alternative representation (rectangles):



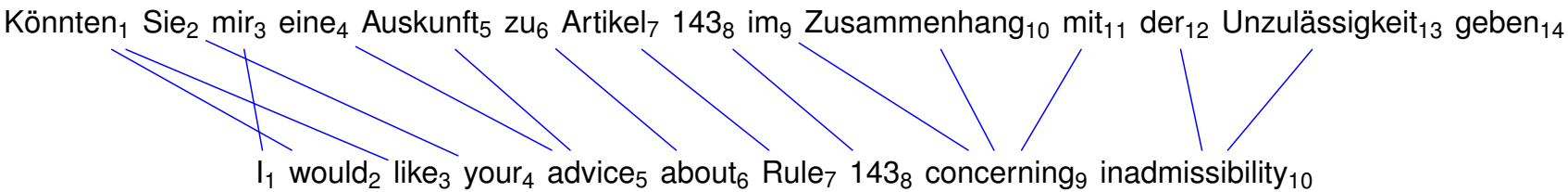
Phrase-based Models



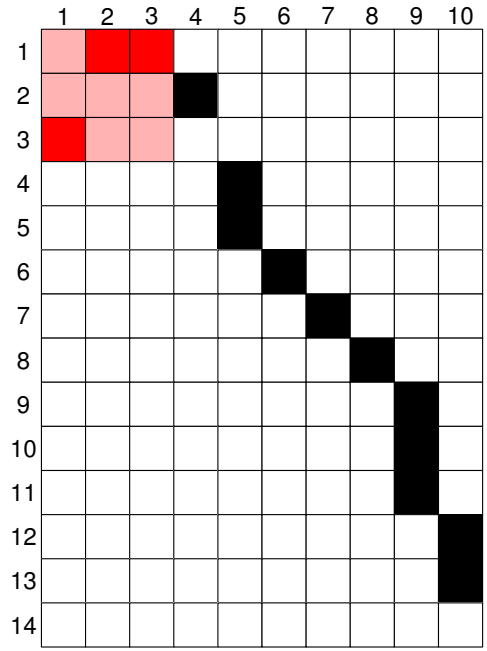
Alternative representation (rectangles):



Phrase-based Models



Alternative representation (rectangles):



Phrase-based Models

Rule weights

- simple relative frequencies during extraction
- normally different normalizations as features

Phrase-based Models

Rule weights

- simple relative frequencies during extraction
- normally different normalizations as features

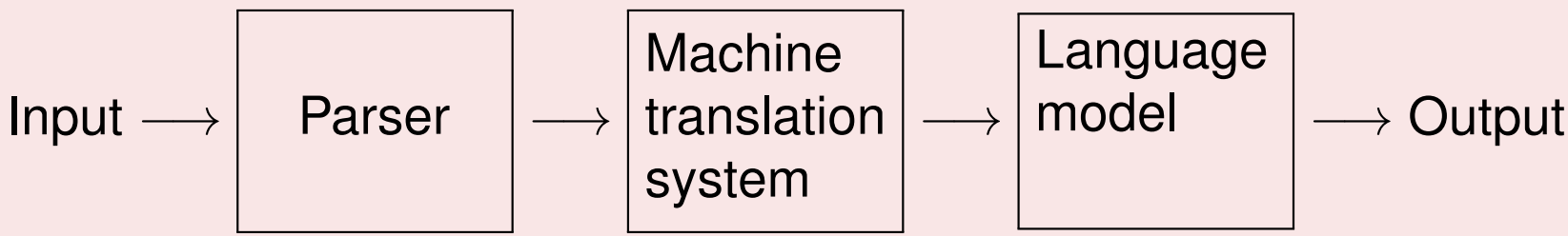
Ideally

- weights should be set to utility of the rule for explaining the training data
- would require reprocessing of training data
- EM or similar algorithms available

→ impractical, EM not used

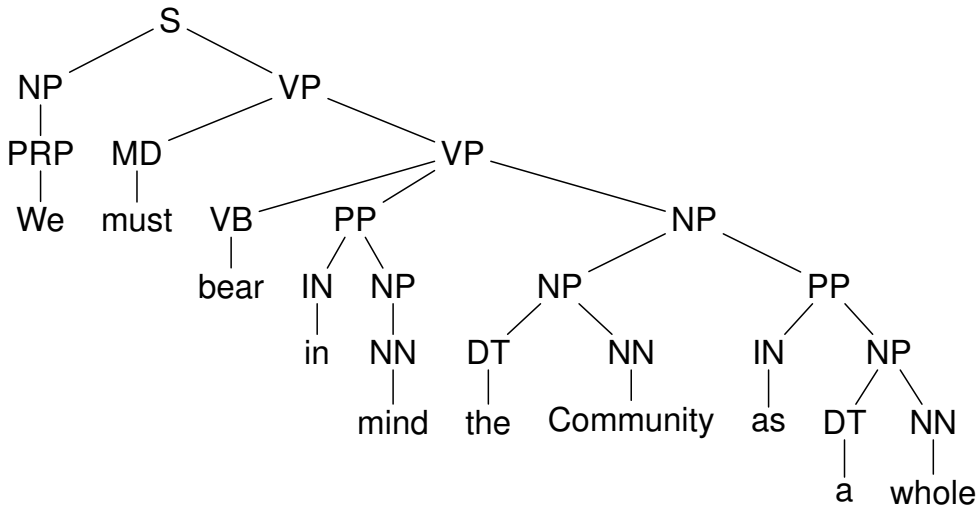
Syntax-based Machine Translation

Syntax-based systems

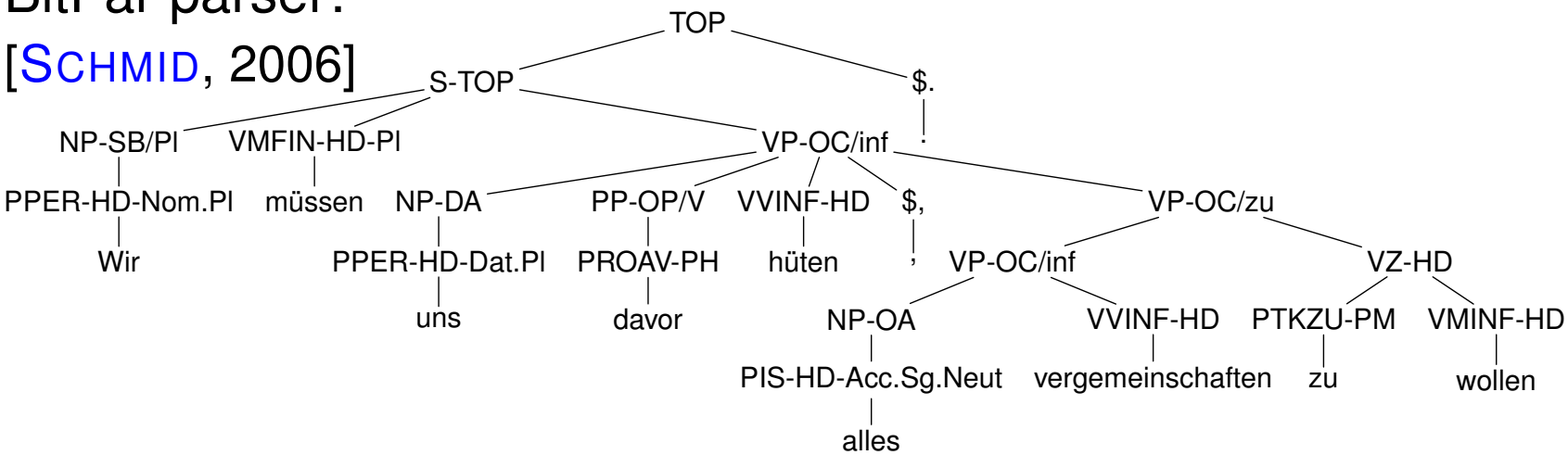


Syntax-based Machine Translation

CHARNIAK parser: [CHARNIAK, JOHNSON, 2005]



BitPar parser:
[SCHMID, 2006]



Syntax-based Machine Translation

Arabic-English

Yugoslav President Voislav signed for Serbia.

و تولى التوقيع عن صربيا الرئيس اليوغوسلافي فويسلاف

Translit.: w twlY AltwqyE En SrbyA Alr}ys Alywgwslafy fwyslAf.

And then the matter was decided, and everything was put in place.

ف كان ان تم الحسم و وضعت الأمور في نصاب ها

Translit.: f kAn An tm AlHsm w wDEt Al>mwr fy nSAb hA.

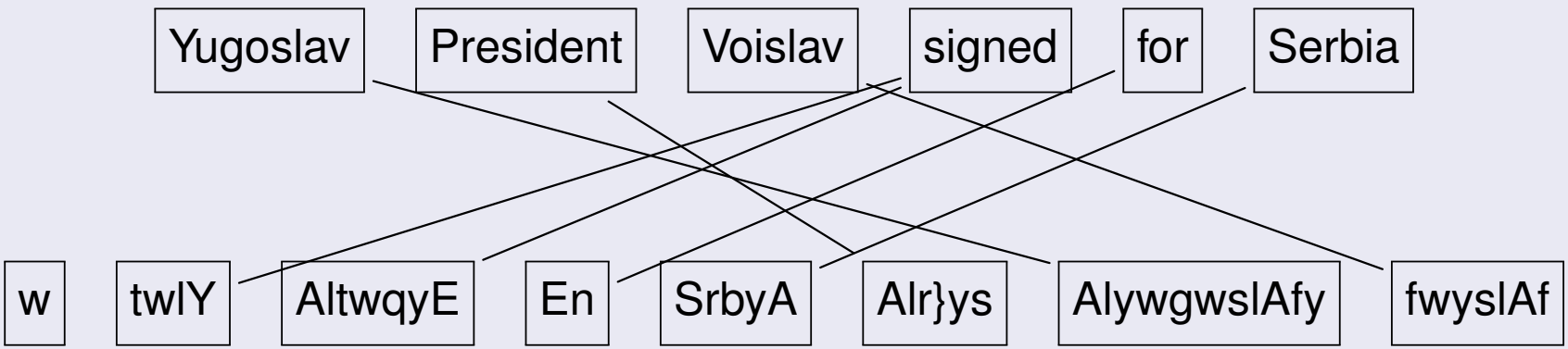
Below are the male and female winners in the different categories.

و هنا الأوائل و الأوليات في مختلف الفئات

Translit.: w hnA Al>wA}l w Al>wlyAt fy mxtlf Alf}At.

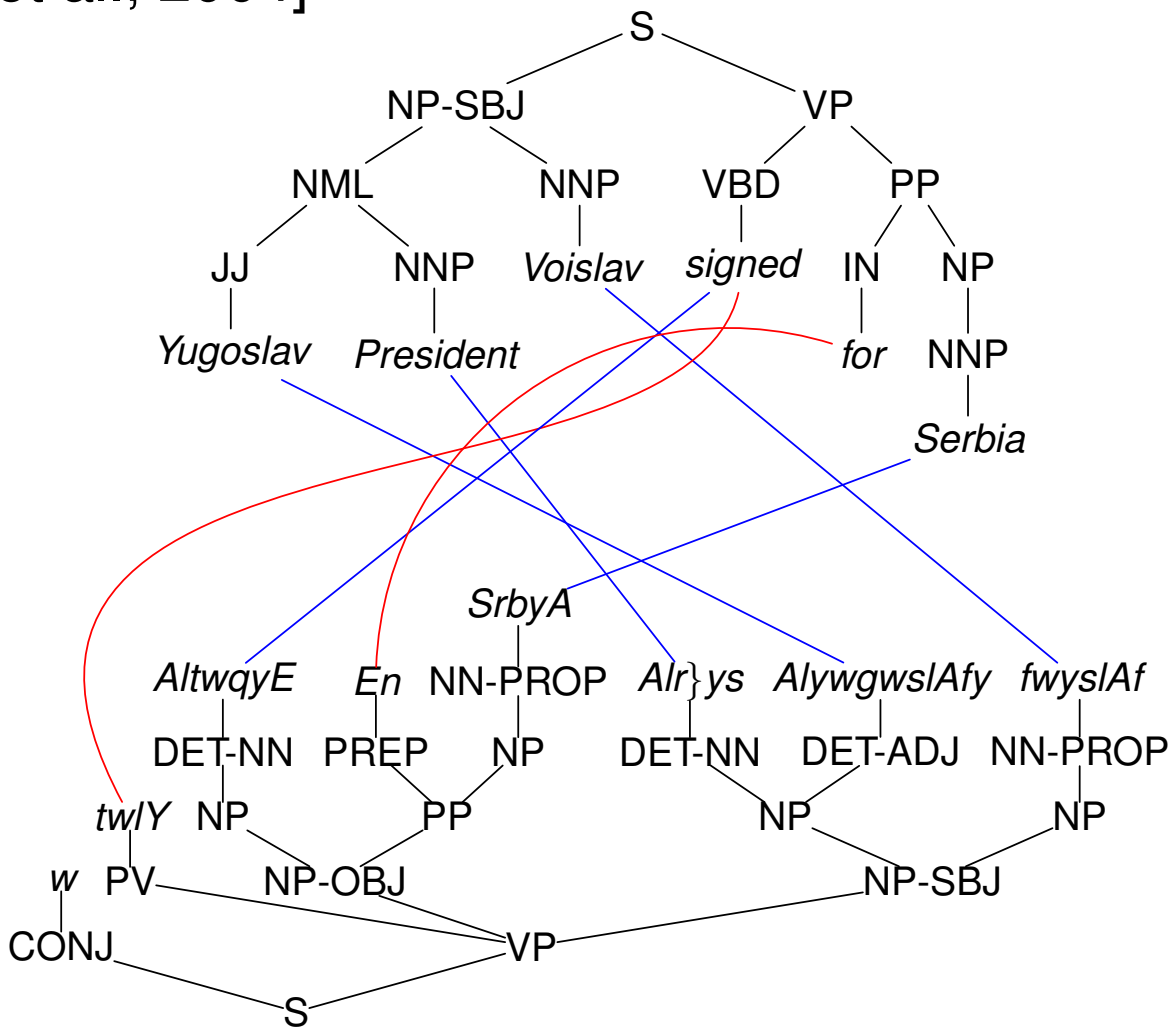
Syntax-based Machine Translation

Alignment

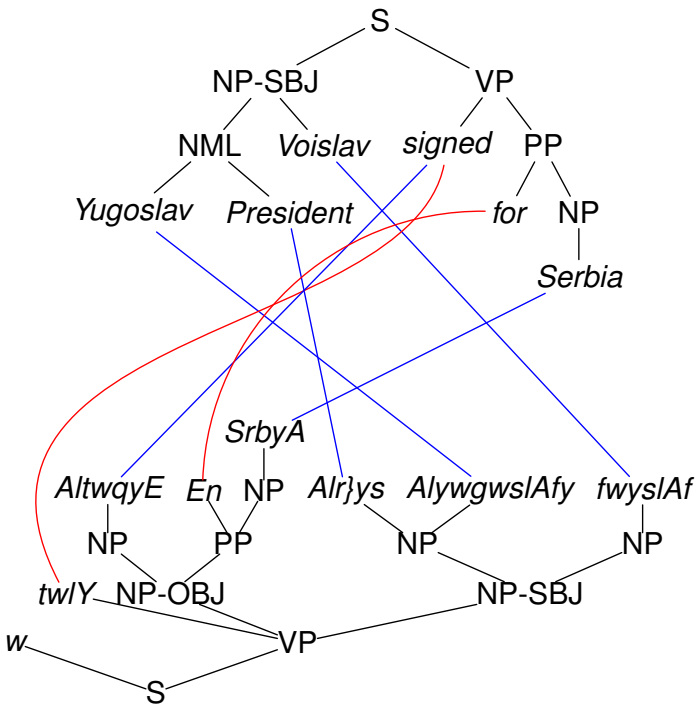


Syntax-based Machine Translation

[GALLEY et al., 2004]

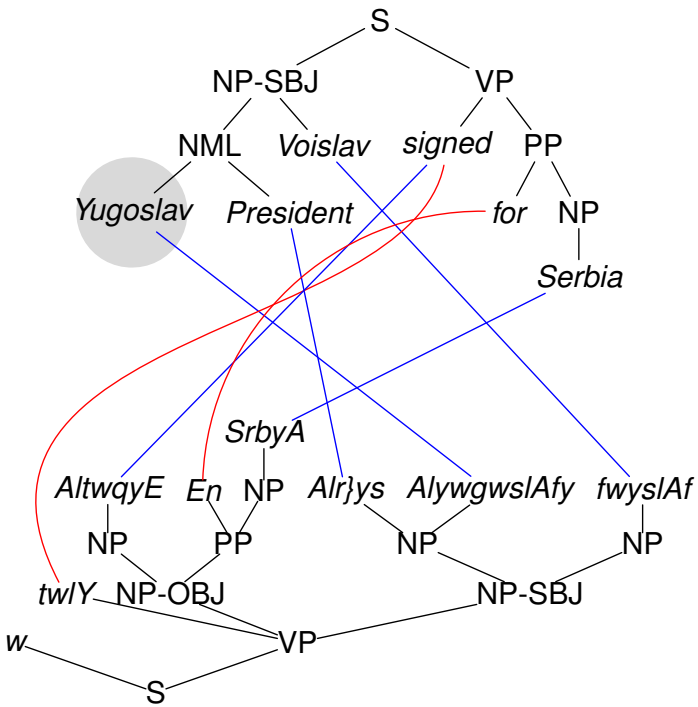


Syntax-based Machine Translation



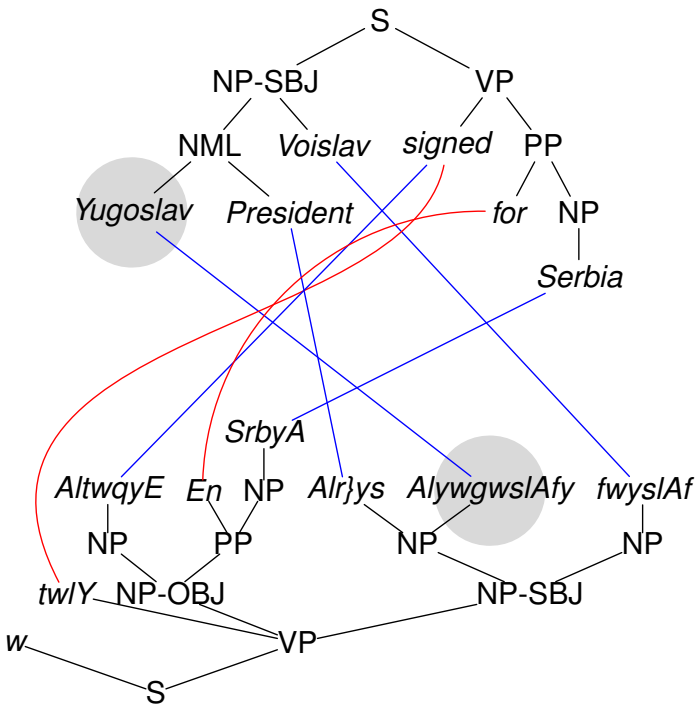
- Select next node bottom-up

Syntax-based Machine Translation



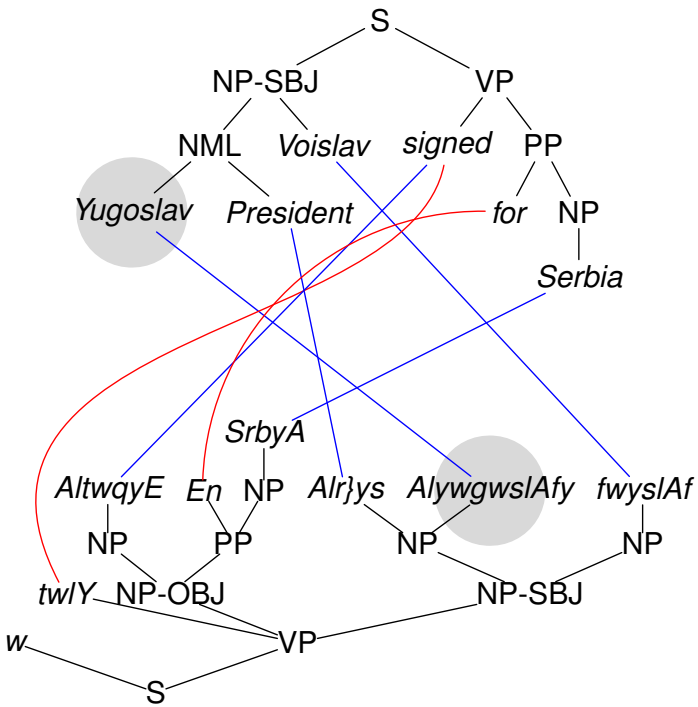
- Select next node bottom-up
- Identify maximal subtree of aligned nodes

Syntax-based Machine Translation



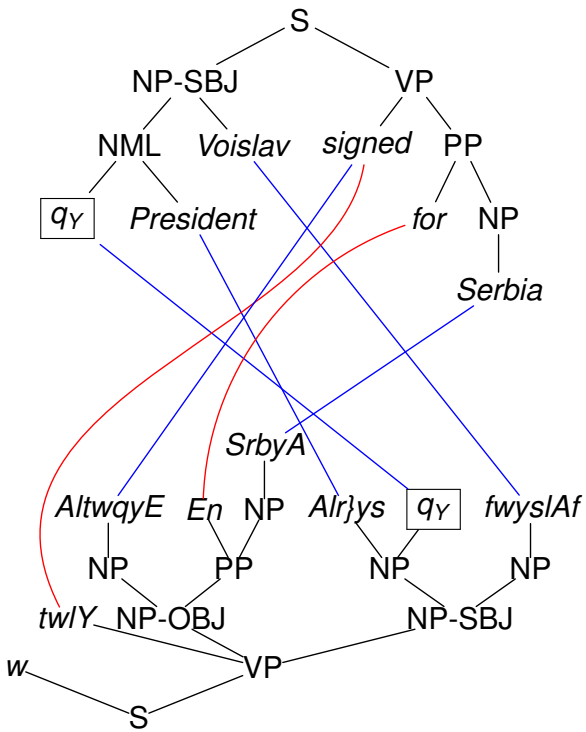
- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state

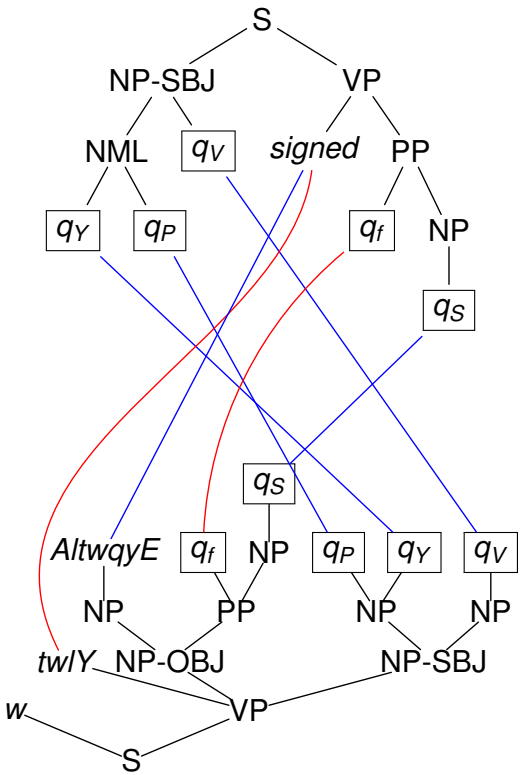
Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

  Yugoslav $\frac{q_Y}{}$ *AlywgwslAfy*

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\textit{Yugoslav} \xrightarrow{q_Y} \textit{AlywgwslAfy}$$

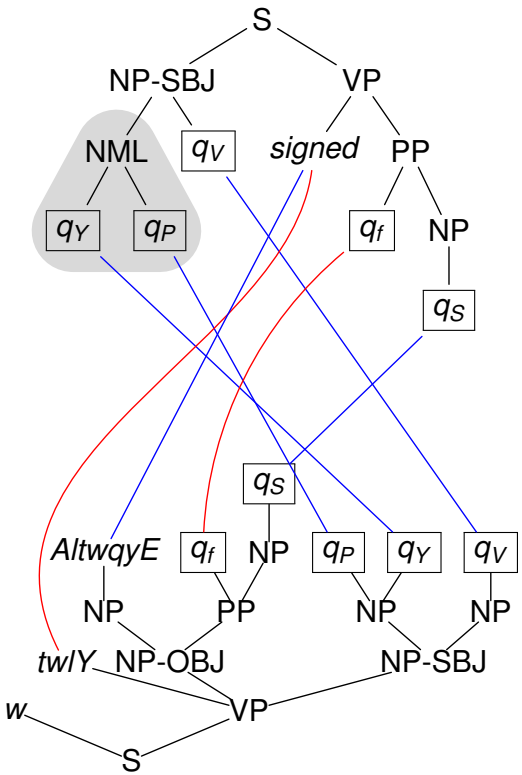
$$\textit{President} \xrightarrow{q_P} \textit{Alr}\}\textit{ys}$$

$$\textit{Voislav} \xrightarrow{q_V} \textit{fwyslAf}$$

$$\textit{for} \xrightarrow{q_f} \textit{En}$$

$$\textit{Serbia} \xrightarrow{q_S} \textit{SrbyA}$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\textit{Yugoslav} \xrightarrow{q_Y} \textit{AlywgwslAfy}$$

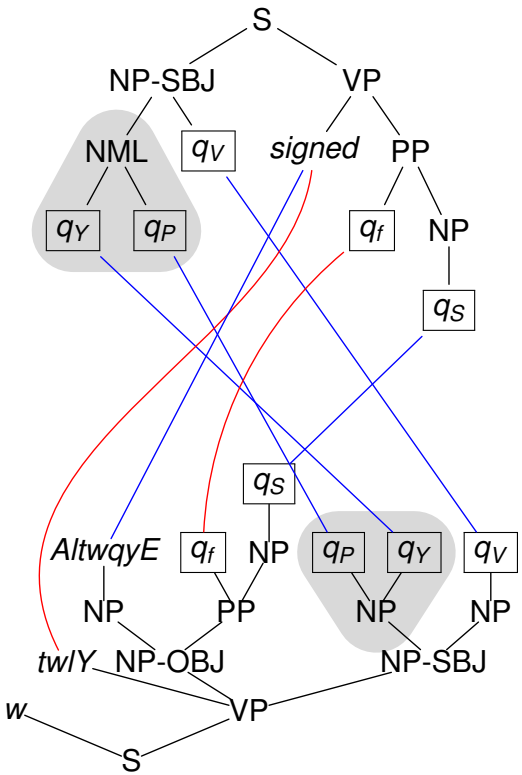
$$\textit{President} \xrightarrow{q_P} \textit{Alr}\}\textit{ys}$$

$$\textit{Voislav} \xrightarrow{q_V} \textit{fwyslAf}$$

$$\textit{for} \xrightarrow{q_f} \textit{En}$$

$$\textit{Serbia} \xrightarrow{q_S} \textit{SrbyA}$$

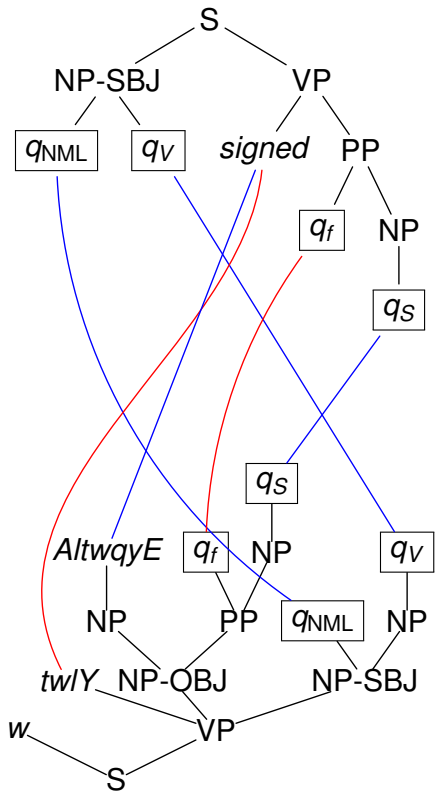
Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\mathrm{NML}(q_Y, q_P) \xrightarrow{q_{\mathrm{NML}}} \mathrm{NP}(q_P, q_Y)$$

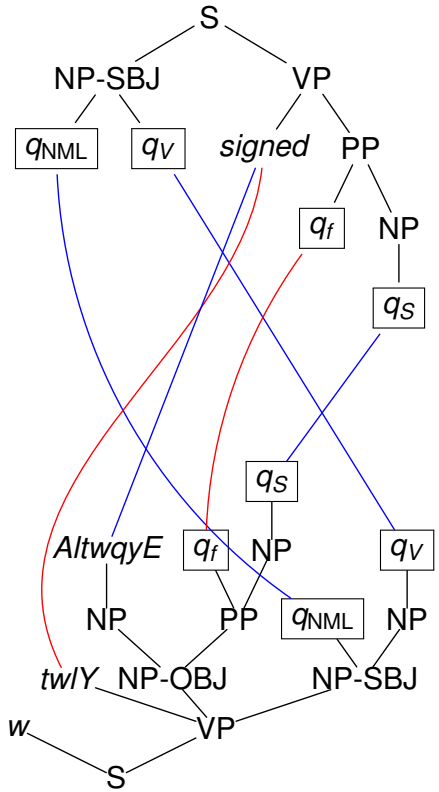
Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\quad}} \text{NP}(q_P, q_Y)$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

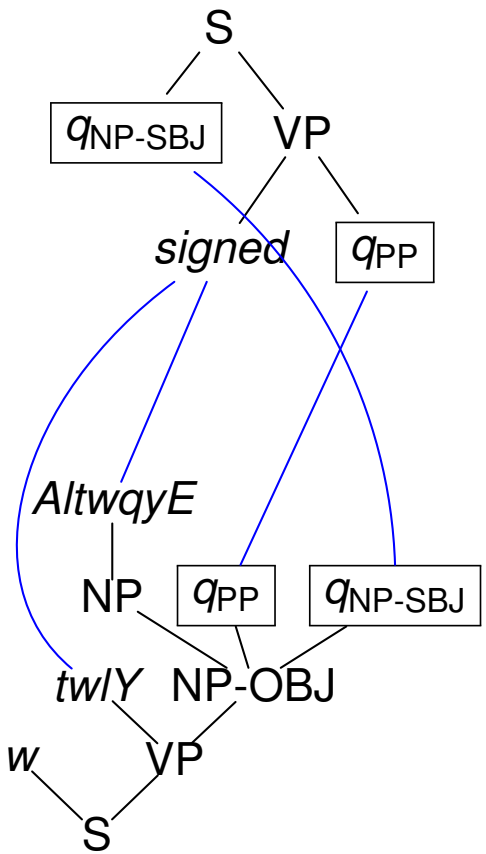
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\quad}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\quad}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\quad}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\quad}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

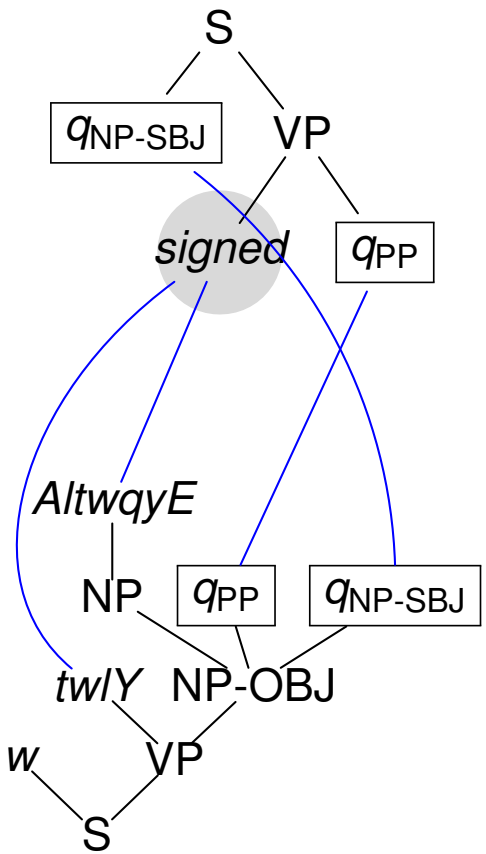
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

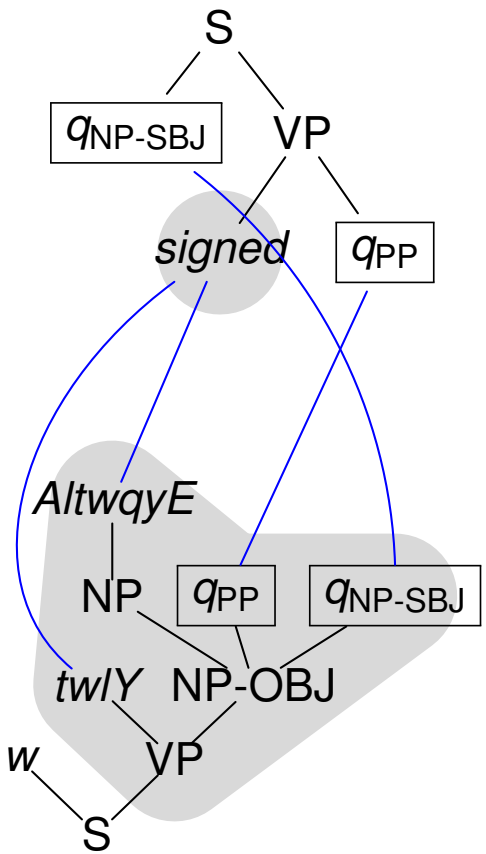
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

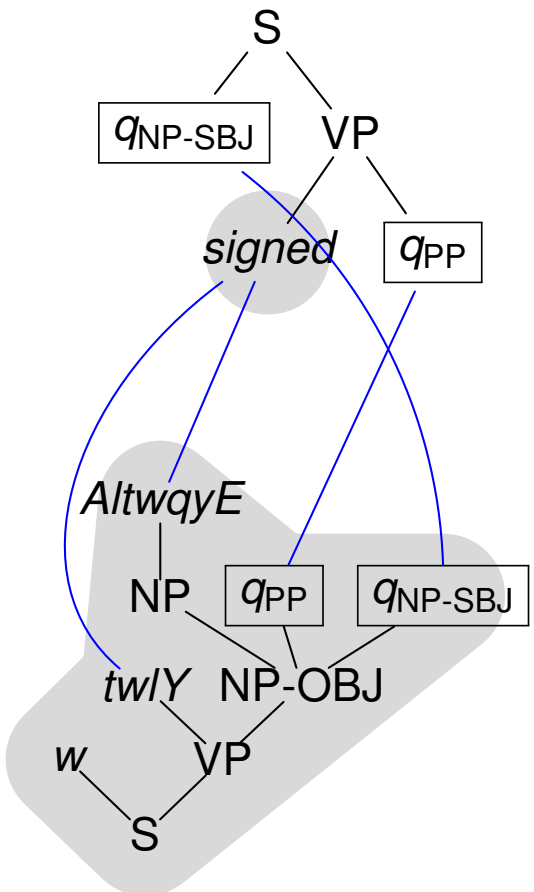
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

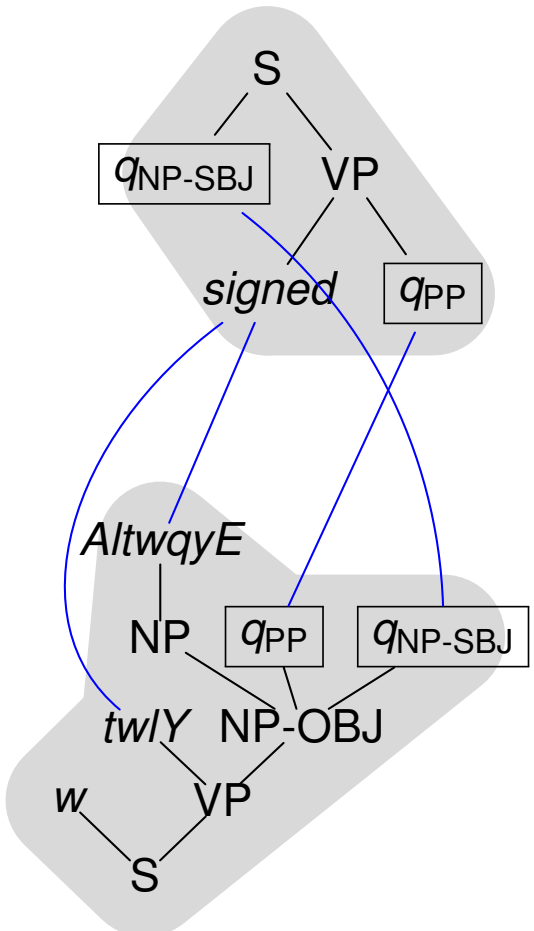
$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation



- Select next node bottom-up
- Identify maximal subtree of aligned nodes
- Identify subtree of nodes aligned to aligned nodes, etc.
- Extract rule and leave state
- Repeat

$$\text{NML}(q_Y, q_P) \overset{q_{\text{NML}}}{\underline{\hspace{1em}}} \text{NP}(q_P, q_Y)$$

$$\text{NP}(q_S) \overset{q_{\text{NP}}}{\underline{\hspace{1em}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \overset{q_{\text{PP}}}{\underline{\hspace{1em}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \overset{q_{\text{NP-SBJ}}}{\underline{\hspace{1em}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation

Rules

$$\textit{Yugoslav} \xrightarrow{q_Y} \textit{AlywgwslAfy} \qquad \textit{President} \xrightarrow{q_P} \textit{Alr\}ys}$$

$$\textit{Voislav} \xrightarrow{q_V} \textit{fwyslAf} \qquad \textit{for} \xrightarrow{q_f} \textit{En}$$

$$\textit{Serbia} \xrightarrow{q_S} \textit{SrbyA}$$

$$\text{NML}(q_Y, q_P) \xrightarrow{q_{\text{NML}}} \text{NP}(q_P, q_Y) \qquad \text{NP}(q_S) \xrightarrow{q_{\text{NP}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \xrightarrow{q_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \xrightarrow{q_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

Syntax-based Machine Translation

Rules

$$\textit{Yugoslav} \xrightarrow{q_Y} \textit{AlywgwslAfy} \qquad \textit{President} \xrightarrow{q_P} \textit{Alr\}ys}$$

$$\textit{Voislav} \xrightarrow{q_V} \textit{fwyslAf} \qquad \textit{for} \xrightarrow{q_f} \textit{En}$$

$$\textit{Serbia} \xrightarrow{q_S} \textit{SrbyA}$$

$$\text{NML}(q_Y, q_P) \xrightarrow{q_{\text{NML}}} \text{NP}(q_P, q_Y) \qquad \text{NP}(q_S) \xrightarrow{q_{\text{NP}}} \text{NP}(q_S)$$

$$\text{PP}(q_f, q_{\text{NP}}) \xrightarrow{q_{\text{PP}}} \text{PP}(q_f, q_{\text{NP}})$$

$$\text{NP-SBJ}(q_{\text{NML}}, q_V) \xrightarrow{q_{\text{NP-SBJ}}} \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

$\rightarrow$ Rules of an Extended Top-down Tree Transducer

Extended Top-down Tree Transducer

Advantages

- ✓ simple and natural model
- ✓ easy to train (from linguistic resources)
  [GRAEHL et al., 2008]
- ✓ symmetric

Extended Top-down Tree Transducer

Advantages

- ✓ simple and natural model
- ✓ easy to train (from linguistic resources)
  [GRAEHL et al., 2008]
- ✓ symmetric

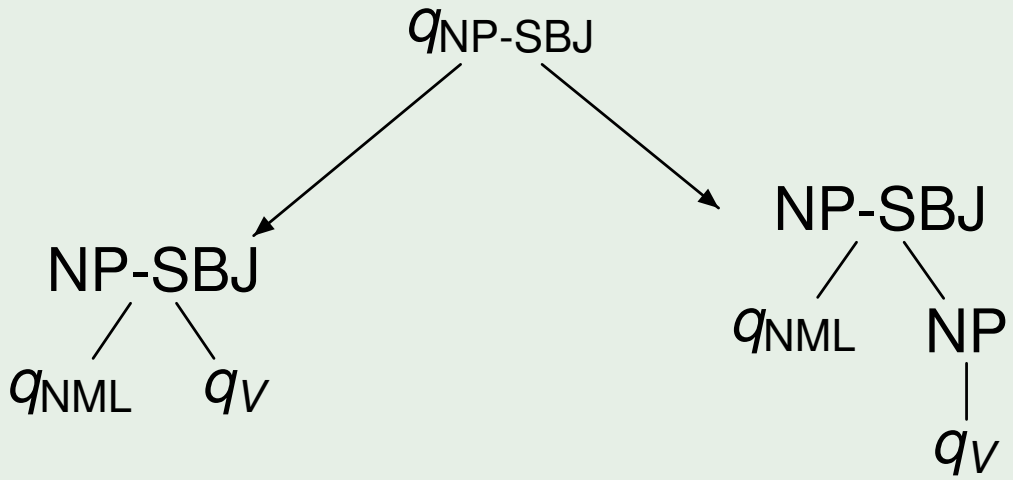
Disadvantages

- weights set in the same manner
- does not capture utility, EM available

→ EM not used in practice

From Automata to Transducers

Graphical representation



corresponds to two productions

$$q_{\text{NP-SBJ}} \longrightarrow \text{NP-SBJ}(q_{\text{NML}}, q_V)$$
$$q_{\text{NP-SBJ}} \longrightarrow \text{NP-SBJ}(q_{\text{NML}}, \text{NP}(q_V))$$

From Automata to Transducers

General idea

Synchronous grammars are essentially two grammars
over the same nonterminals whose productions are paired

Convention

same nonterminals are synchronized (or linked)
and develop at the same time

From Automata to Transducers

Approach

- join two productions $q_1 \rightarrow r_1$ and $q_2 \rightarrow r_2$ to $(q_1, q_2) \rightarrow (r_1, r_2)$

From Automata to Transducers

Approach

- join two productions $q_1 \to r_1$ and $q_2 \to r_2$ to $(q_1, q_2) \to (r_1, r_2)$
- demand $q_1 = q = q_2$ for simplicity and write $r_1 \xrightarrow{q} r_2$

From Automata to Transducers

Approach

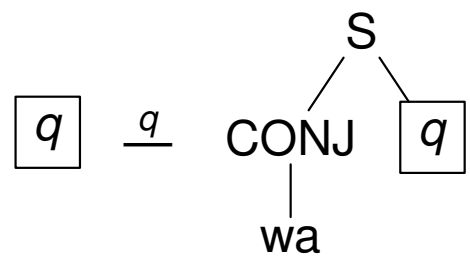
- join two productions $q_1 \rightarrow r_1$ and $q_2 \rightarrow r_2$ to $(q_1, q_2) \rightarrow (r_1, r_2)$
- demand $q_1 = q = q_2$ for simplicity and write $r_1 \overset{q}{\text{---}} r_2$
- paired productions develop input and output tree at the same time

From Automata to Transducers

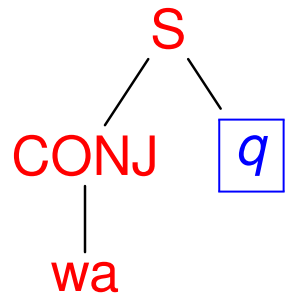
$\boxed{q}$ $\boxed{q}$

Used rule:

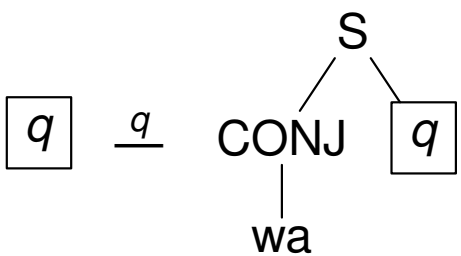
Next rule:



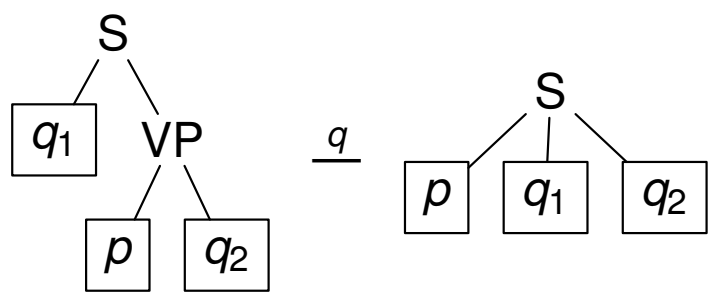
From Automata to Transducers



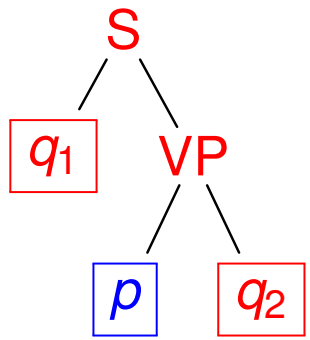
Used rule:

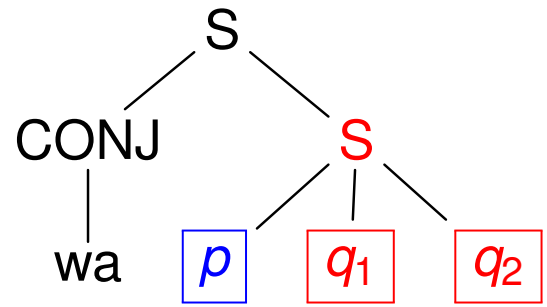


Next rule:

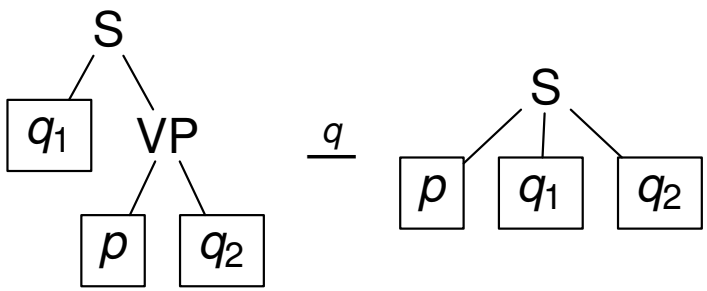


From Automata to Transducers

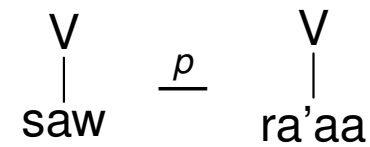




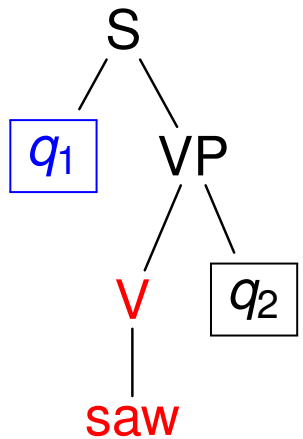
Used rule:

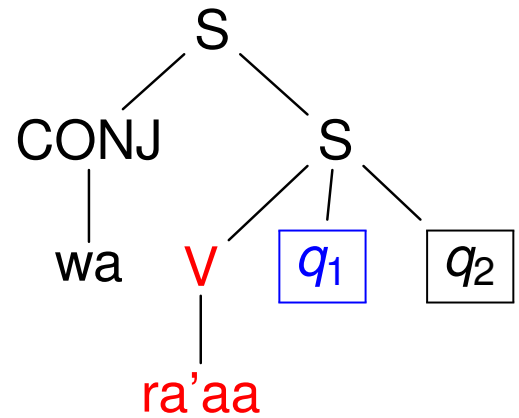


Next rule:

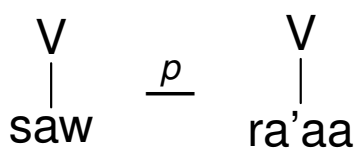


From Automata to Transducers

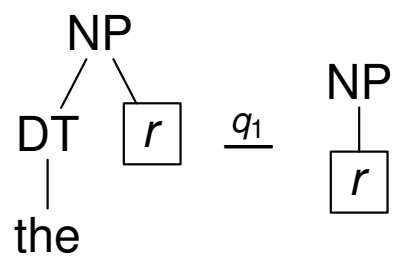




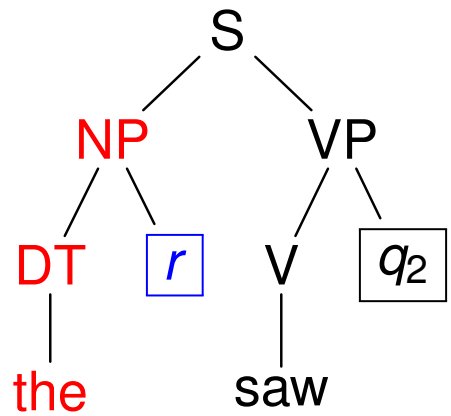
Used rule:

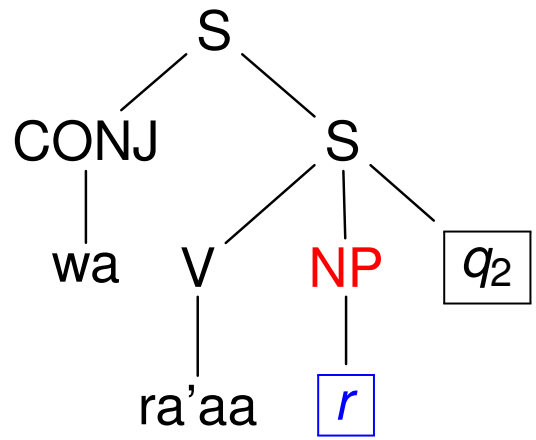


Next rule:

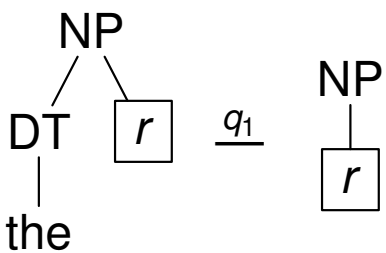


From Automata to Transducers

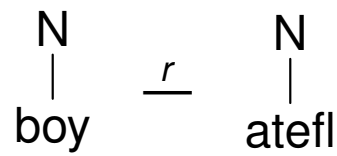




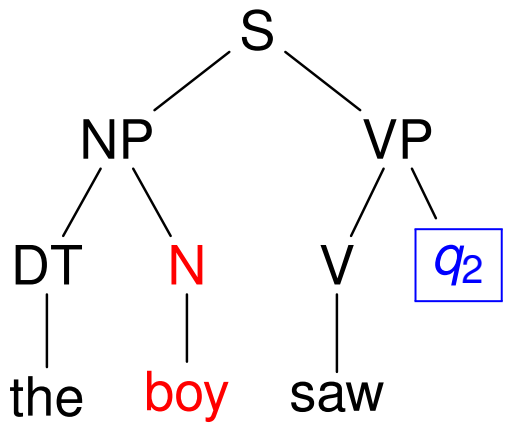
Used rule:

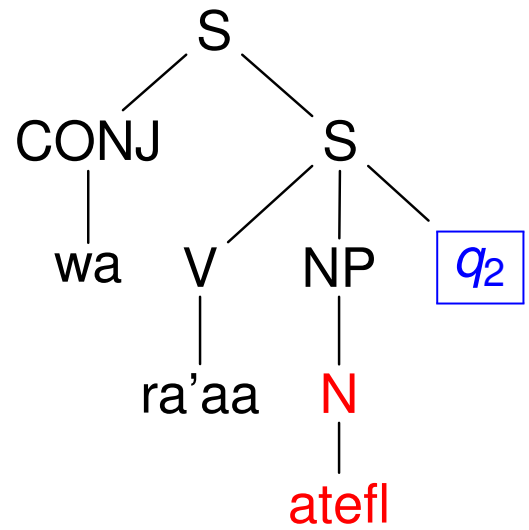


Next rule:

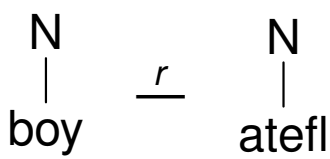


From Automata to Transducers

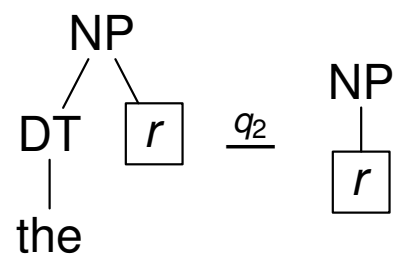




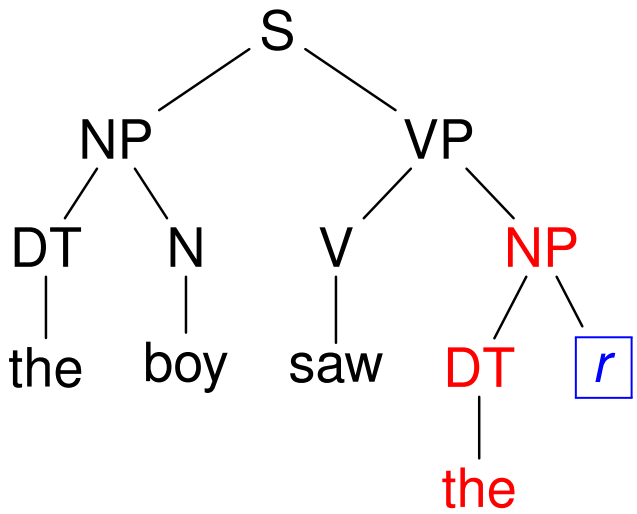
Used rule:

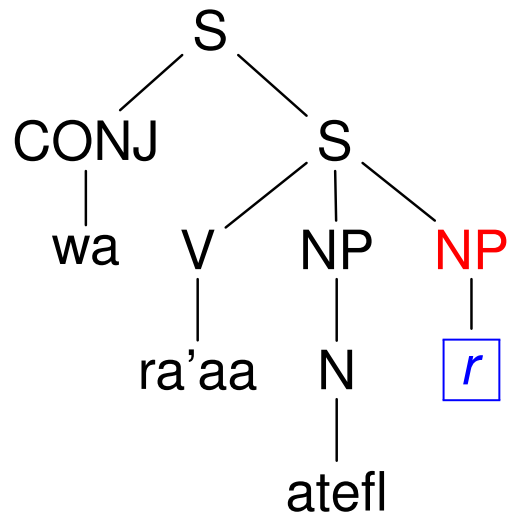


Next rule:

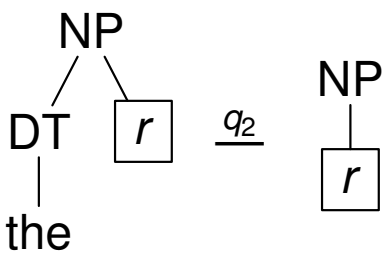


From Automata to Transducers

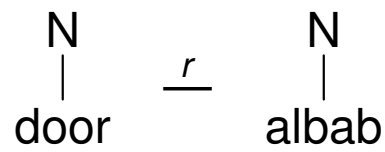




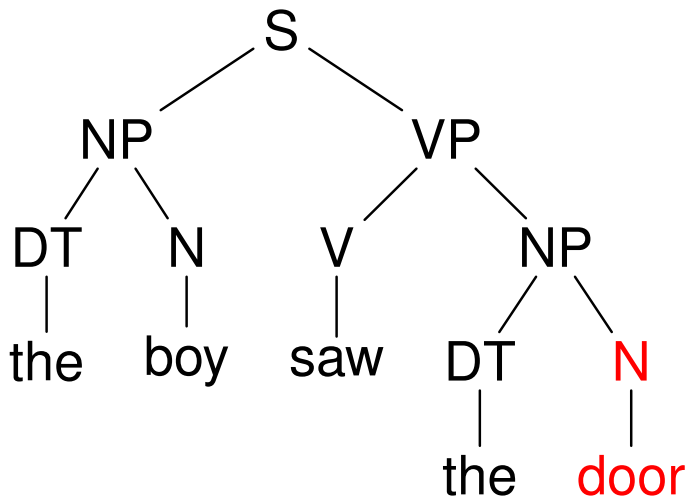
Used rule:

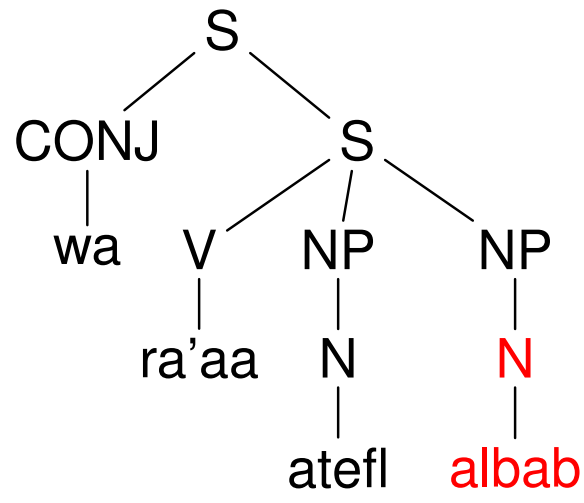


Next rule:

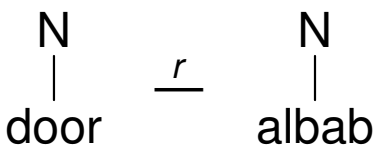


From Automata to Transducers





Used rule:



Next rule:

From Automata to Transducers

Remarks

- synchronization breaks almost all existing constructions (e.g., the normalization construction)

$\rightarrow$ the basic grammar model very important

Other Syntax-based Models

Important relations

- SCFG = synchronous context-free grammar LTG-LTG
  [CHIANG, 2007] (synchronous local tree grammar)
  $\subseteq$ ln-TOP special top-down tree transducer
- STSG = synchronous tree substitution grammar TSG-TSG
  [EISNER, 2003]
  $\subseteq$ ln-XTOP special extended top-down tree transducer
- STAG = synchronous tree adjunction grammar TAG-TAG
  [SHIEBER, SCHABES, 1990]
- SCFTG = synchronous context-free tree grammar
  [NEDERHOF, VOGLER, 2012] CFTG-CFTG

Other Syntax-based Models

Towards asymetric relations

- STSSG = synch. tree-sequence substitution grammar [ZHANG et al., 2008] TSSG-TSSG
- ℓMBOT = local shallow multi bottom-up tree transducer [BRAUNE et al., 2013] LTG-TSSG

In-XMBOT corresponds rougly to RTG-TSSG

Where is Machine Learning?

Examples

- unsupervised learning of translation models for STSG [BLUNSOM et al., 2008]
- uses GIBBS sampling, hierarchical DIRICHLET process, ...

Where is Machine Learning?

Examples

- unsupervised learning of translation models for STSG [BLUNSOM et al., 2008]
- uses GIBBS sampling, hierarchical DIRICHLET process, ...

But ...

- can only be used on small data
- does not scale well

→ currently not used in practice

Machine Learning & Grammatical Inference

Quo vadis?

Literature

Selected references


ARNOLD, DAUCHET: *Morphismes et Bimorphismes d'Arbres*
Theoret. Comput. Sci. 20, 1982


BRAUNE, SEEMANN, QUERNHEIM, ∼: *Shallow local multi bottom-up tree transducers in statistical machine translation.* Proc. ACL, 2013


GALLEY, HOPKINS, KNIGHT, MARCU
What's in a Translation Rule? Proc. NAACL, 2004


KOEHN: *Statistical Machine Translation*
Cambridge University Press, 2010


∼, GRAEHL, HOPKINS, KNIGHT: *The Power of Extended Top-down Tree Transducers.* SIAM J. Comput. 39, 2009


OCH, NEY: *The Alignment Template Approach to Statistical Machine Translation.* Comput. Linguist. 30, 2004