Using Linguistic Syntax in Statistical Machine Translation

Greg Hanneman

11-731: Machine Translation
March 22, 2016

LTI

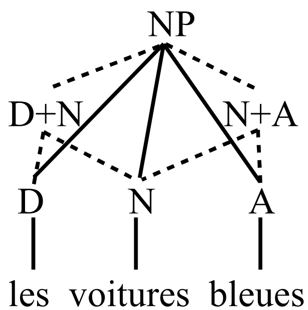
Carnegie Mellon

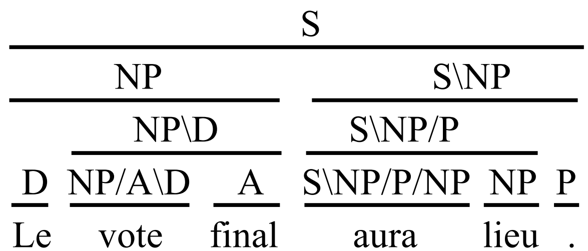
Syntax in Machine Translation

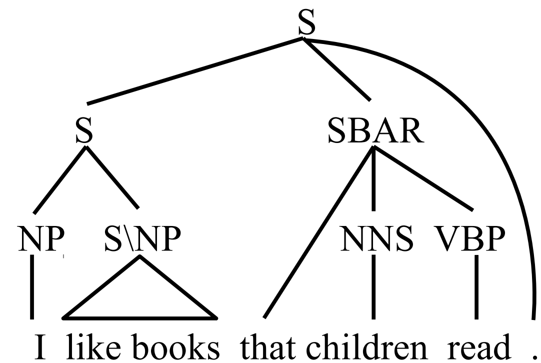
What does that even mean?

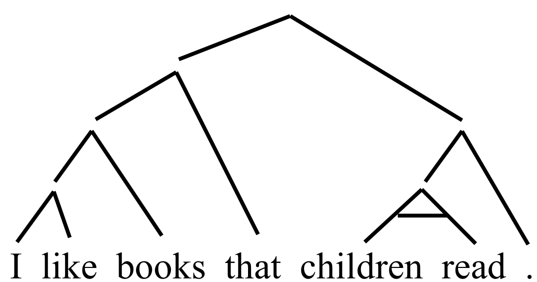
Syntax in Machine Translation

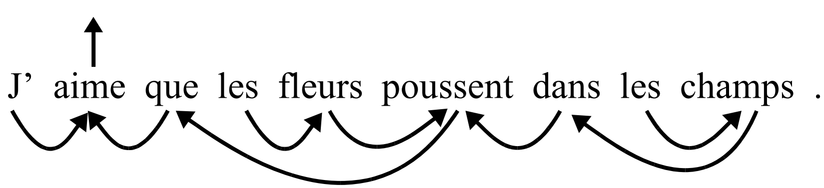
What does that even mean?

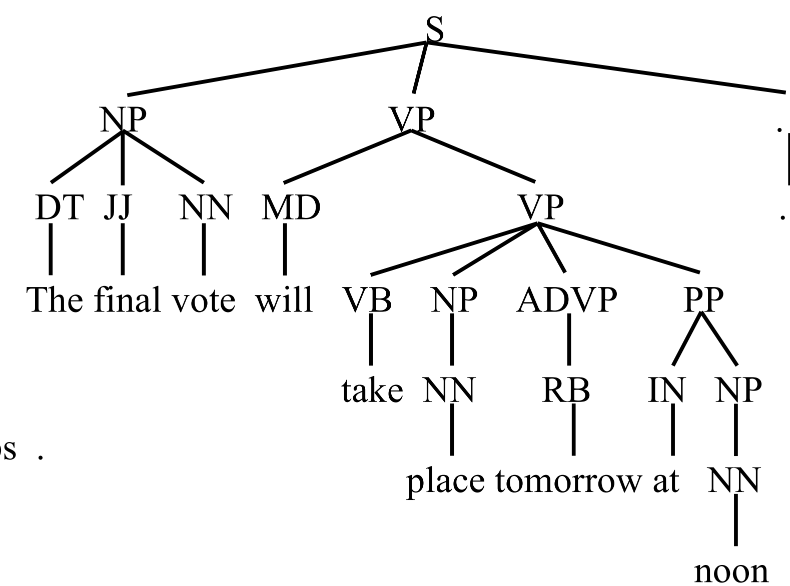












Syntax in Machine Translation

Why would we want that?

Syntax in Machine Translation

Why would we want that?

- Language is hierarchical, not finite-state!
- Generalize over different/unseen strings with same structure
- Languages have structural divergences in addition to lexical ones
- More accurate reordering at arbitrary distances

Syntax in Machine Translation

How would we make it happen?

Syntax in Machine Translation

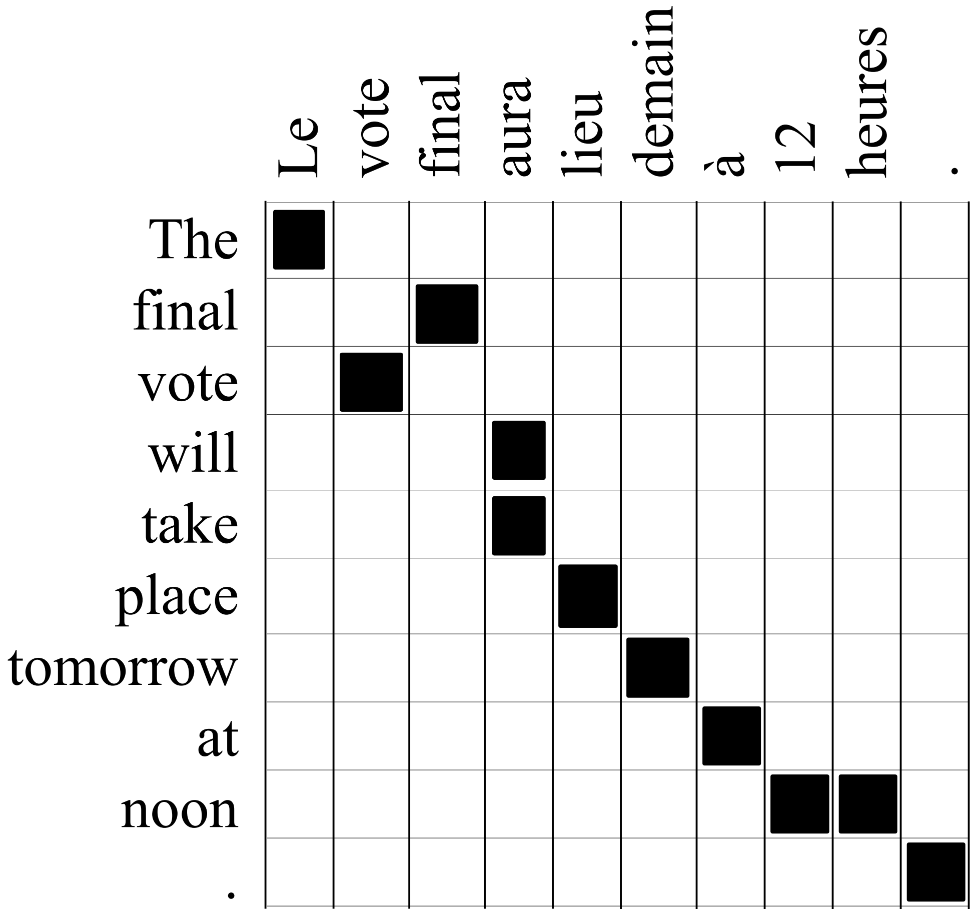
How would we make it happen?

- What notion of syntax?
- Syntax on source, target, or both?
- What do the rules look like?
- What do the labels look like?
- Input is string, tree, or forest?
- Decode bottom-up or top-down?

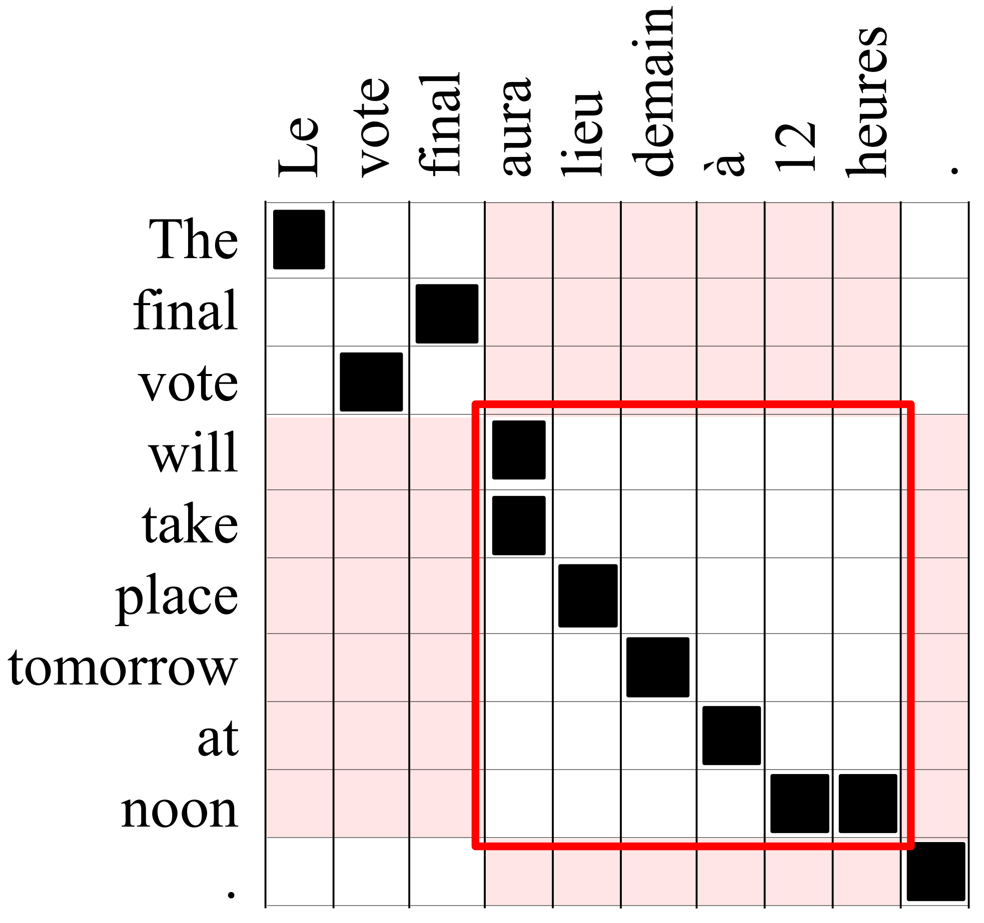
Syntax in Machine Translation

Side(s)	Rules	Labels	Input	Technique(s)
Source	SCFG	Treebank	Tree	Yamada and Knight [2001]
Source	STSG	Treebank	Tree	Huang et al. [2006], Liu et al. [2006]
Source	STSG	Treebank	Forest	Mi et al. [2008]
Target	SCFG	Treebank + combinations	String	Zollmann & Venugopal [2006]
Target	SCFG	"X" with soft preferences	String	Venugopal et al. [2009]
Target	STSG	Treebank	String	Galley et al. [2004]
Target	SCFG	Syntactic "X" with soft prefs	String	Huck et al. [2014]
Both	STSG	Bi-treebank	Tree	Xia and McCord [2004]
Both	SCFG	Bi-treebank	String	Lavie et al. [2008]
Both	STSG	Bi-treebank	Forest	Liu et al. [2009]
Both	SCFG	Bi-treebank + virtual nodes	String	Hanneman et al. [2011]
Both	SCFG	Bi-treebank + combinations, with soft matches	String	Chiang [2010]

Baseline: Hiero

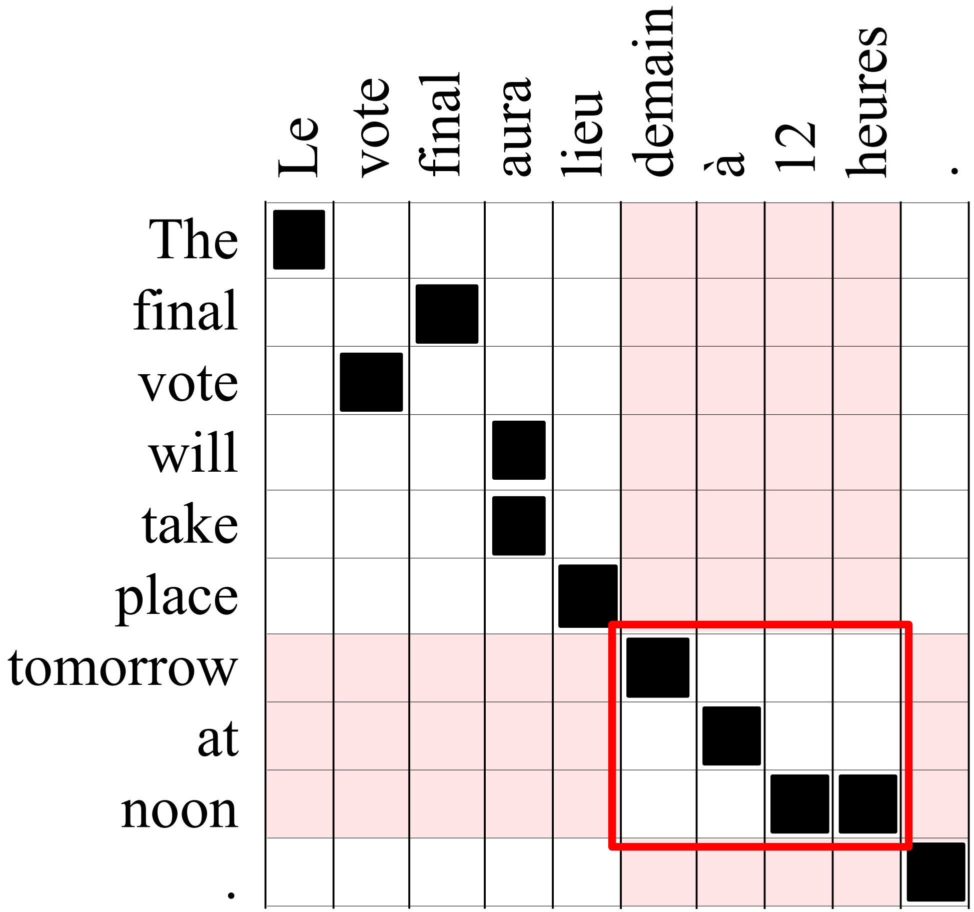


Baseline: Hiero



X → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

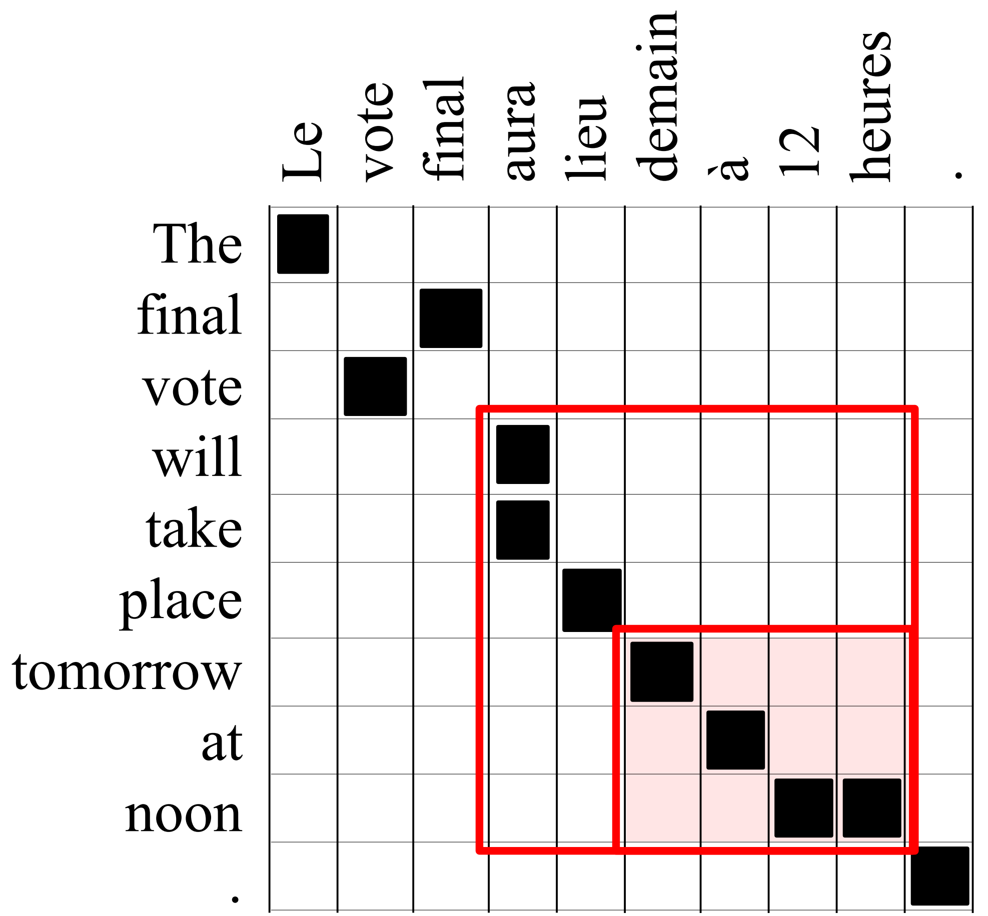
Baseline: Hiero



X → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

X → [demain à 12 heures]::[tomorrow at noon]

Baseline: Hiero



X → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

X → [demain à 12 heures]::[tomorrow at noon]

X → [aura lieu X_1]::[will take place X_1]

Hiero Pros and Cons

- Synchronous context-free grammar (SCFG)
- Easy to get given basic phrase extraction
- Sometimes works quite well

$$X \rightarrow [\text{que } X_1 \text{ les } X_2]::[\text{that } X_2\ X_1]$$

J' aime les livres que lisent les enfants .

I like books that children read .

✓

Hiero Pros and Cons

- Synchronous context-free grammar (SCFG)
- Easy to get given basic phrase extraction
- But sometimes doesn't…

$$X \longrightarrow [\text{que } X_1 \text{ les } X_2]::[\text{that } X_2\ X_1]$$

J' aime les livres que lisent les enfants .

I like books that children read .

J' aime que les fleurs poussent dans les champs .

I like that fields flowers grow in .

The Trouble with "X"s

- Nonterminal boundaries can be anywhere
- Can extract rule in one context and apply it in another – all "X"s look the same
- "Formally but not linguistically syntactic"

Part 1

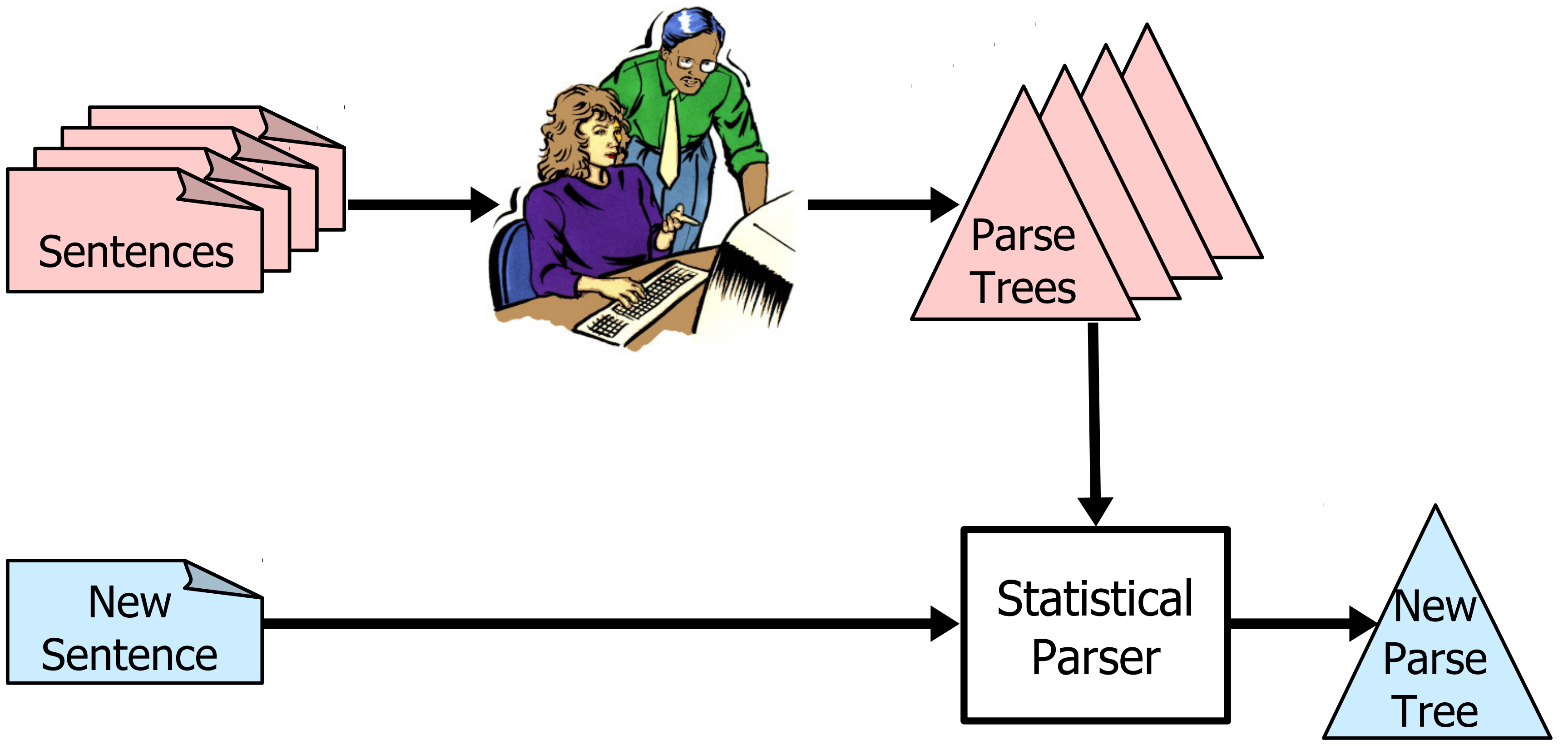
String-to-Tree Methods

“Syntax-Augmented” MT

[Zollmann and Venugopal, 2006; Zollmann, 2011]

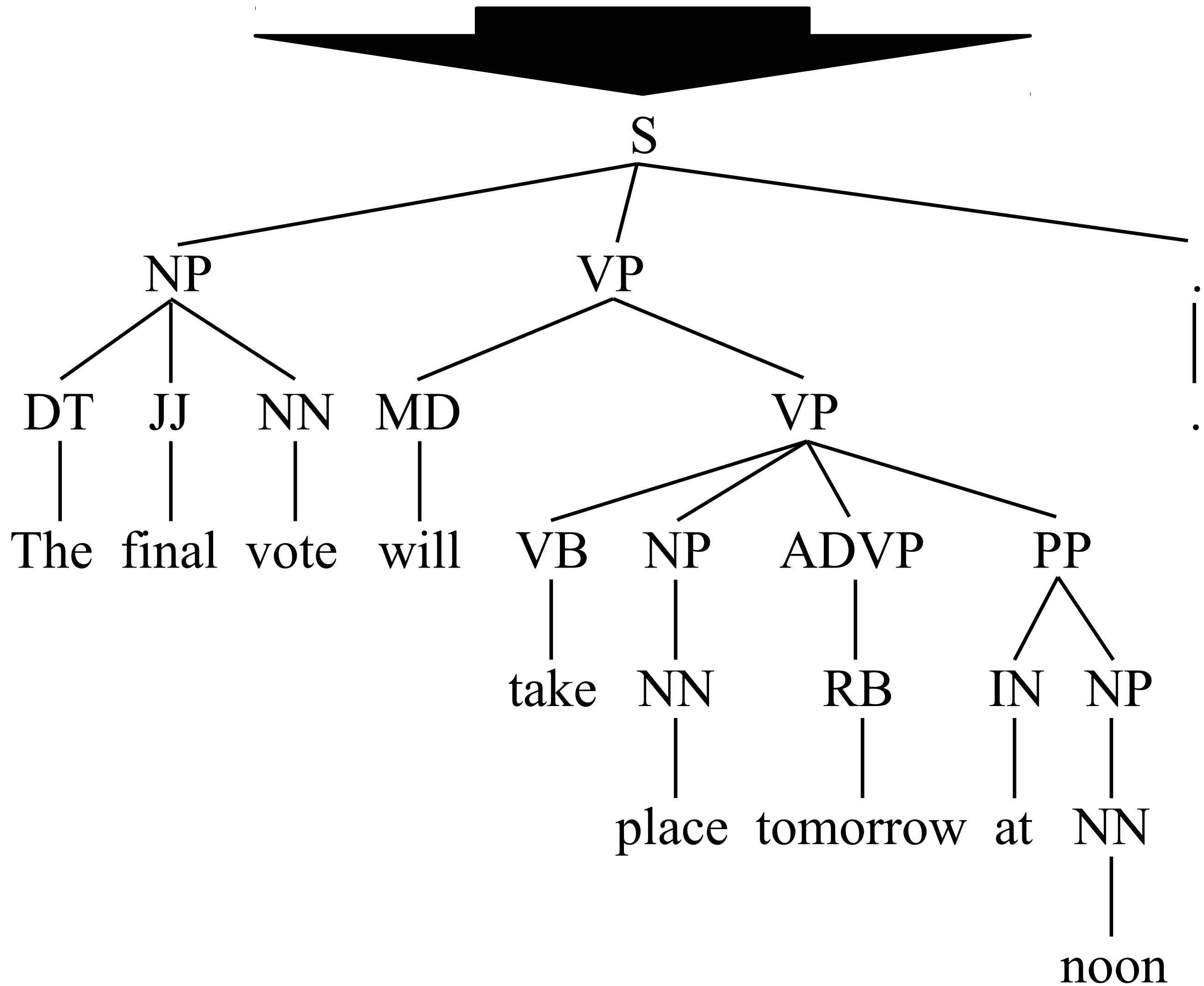
- Let’s label these Hiero rules!
- Replace “X” with labels derived from a constituency parse of the target side
- Why target?
- What about phrases that aren’t constituents? Do we lose coverage?

Aside: Statistical Parsers

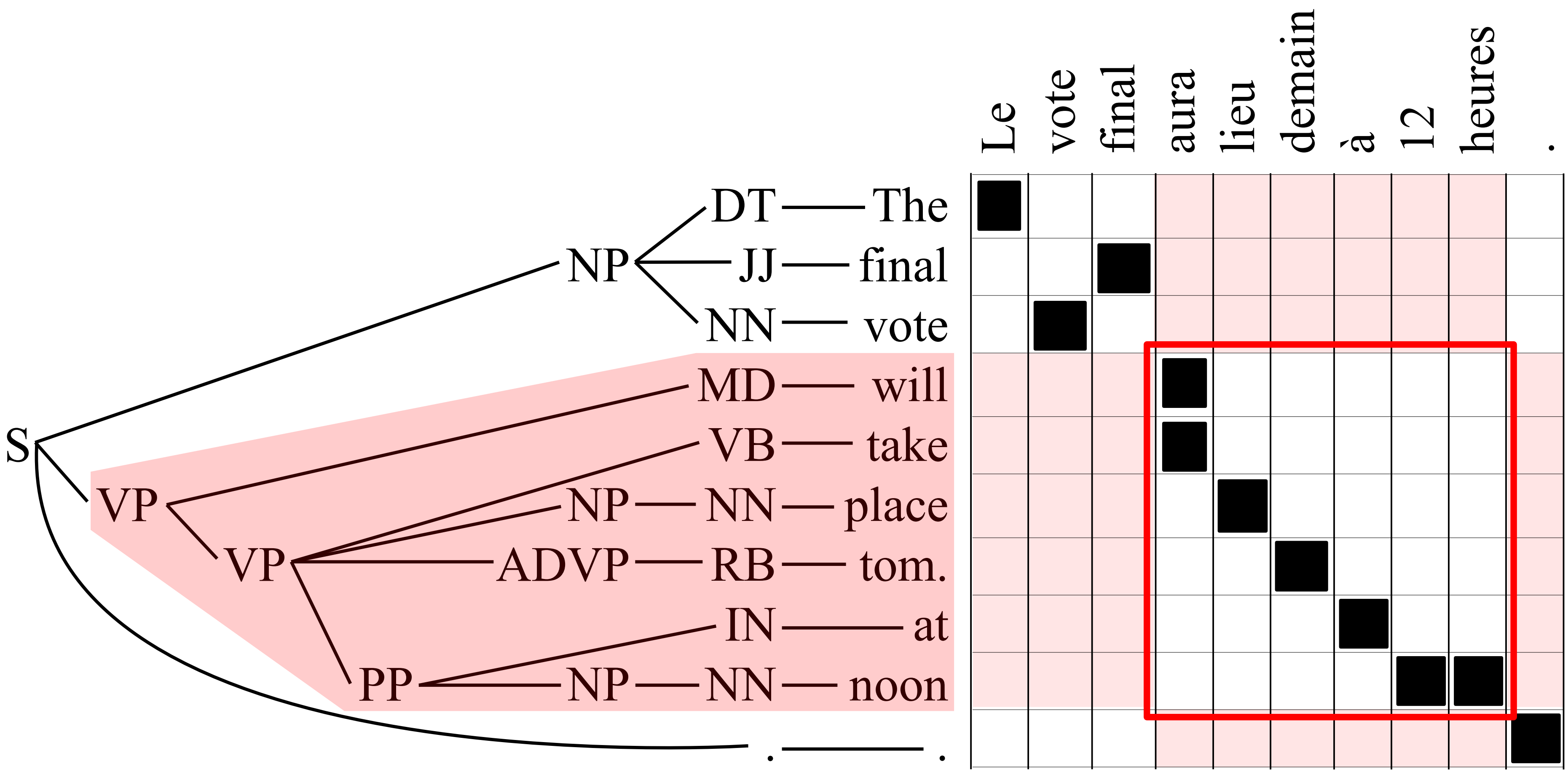


Aside: Statistical Parsers

The final vote will take place tomorrow at noon .

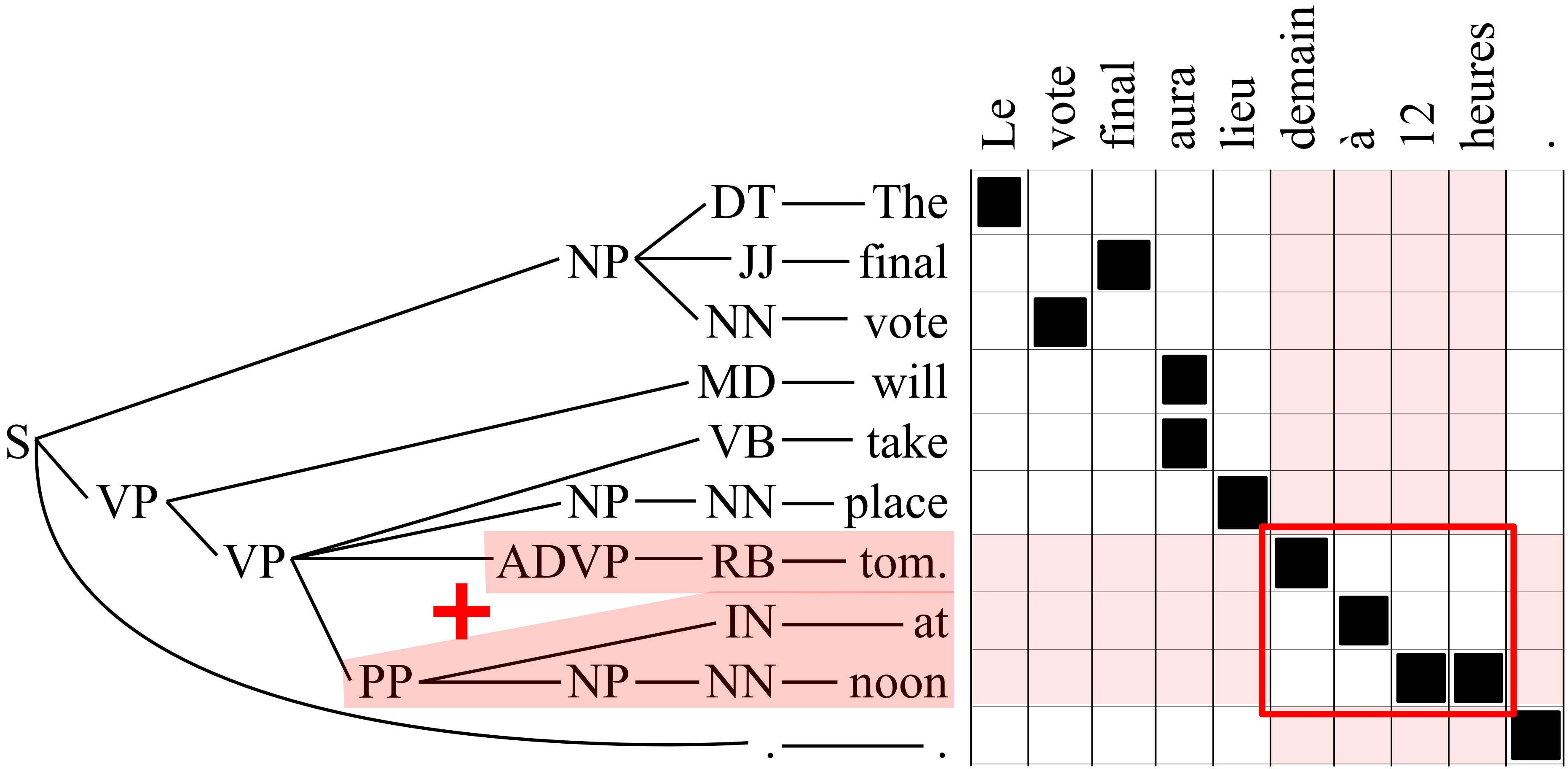


SAMT Rule Extraction



VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

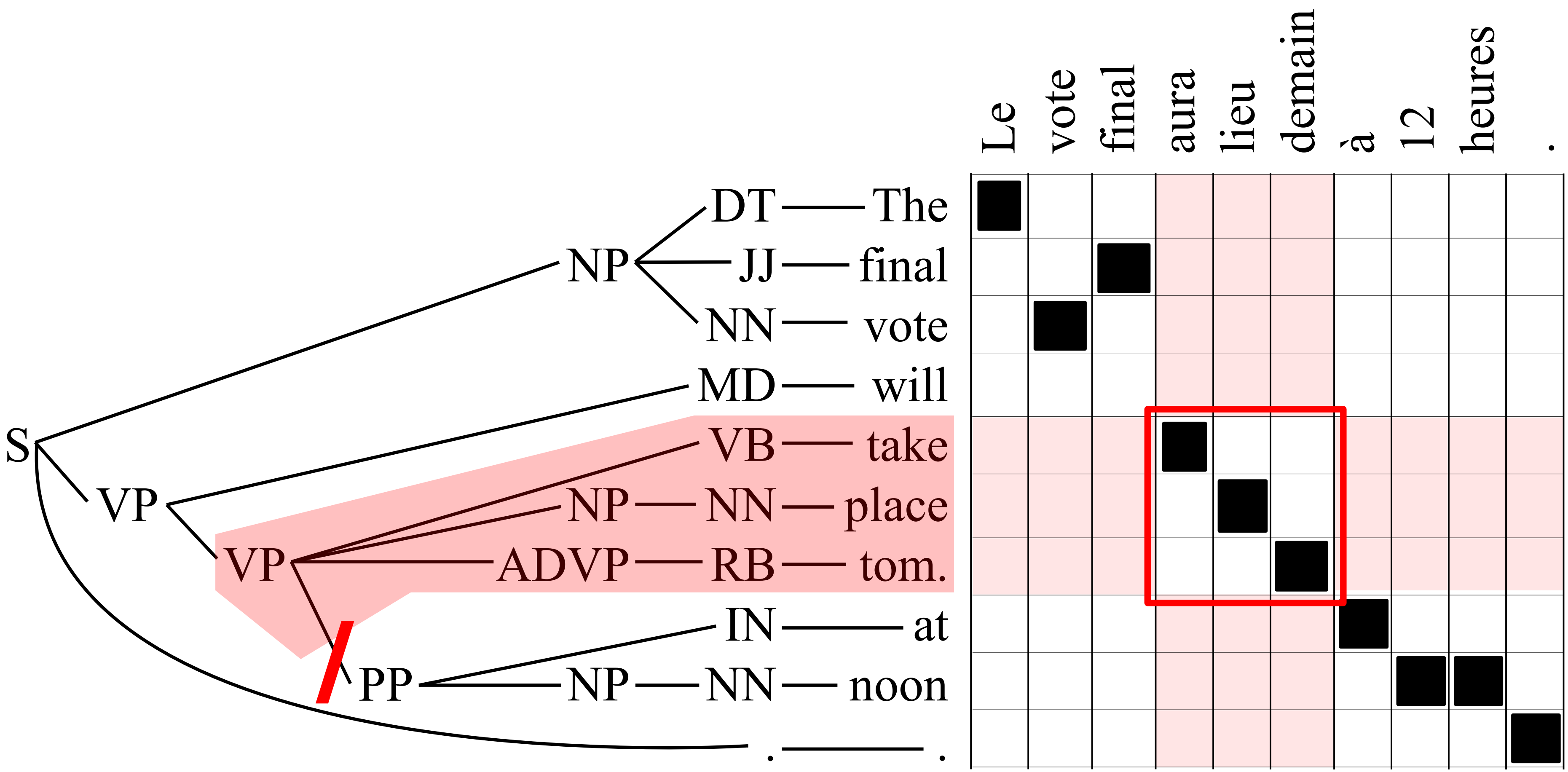
SAMT Rule Extraction



VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

ADVP+PP → [demain à 12 heures]::[tomorrow at noon]

SAMT Rule Extraction

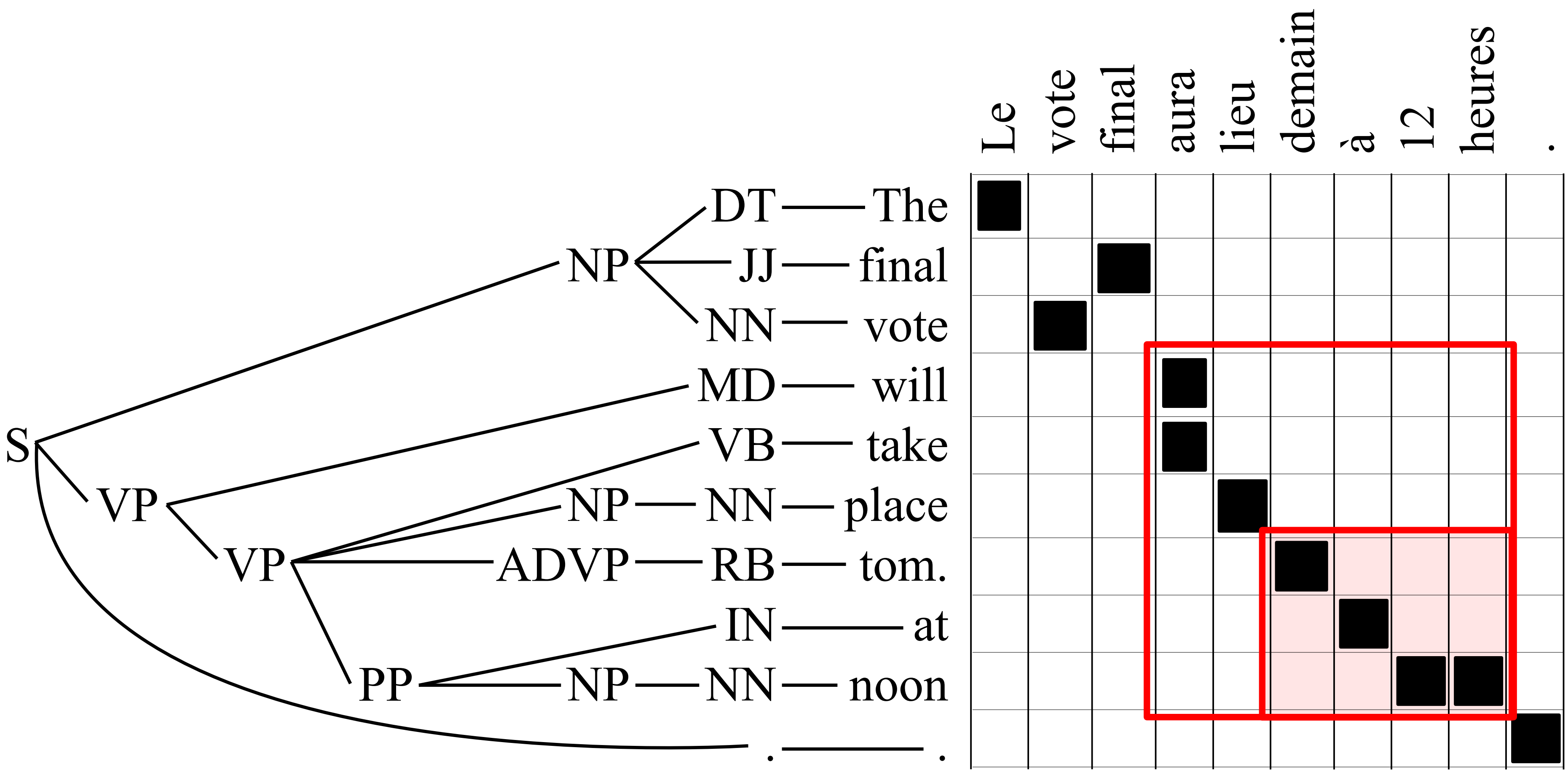


VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

ADVP+PP → [demain à 12 heures]::[tomorrow at noon]

VP/PP → [aura lieu demain]::[take place tomorrow]

SAMT Rule Extraction



VP → [aura lieu demain à 12 heures]::[will take place tomorrow at noon]

ADVP+PP → [demain à 12 heures]::[tomorrow at noon]

VP → [aura lieu ADVP+PP$_1$]::[will take place ADVP+PP$_1$]

SAMT Labeling

- Try, in order:
  - A exact constituent match
  - A+B adjacent constituents (not nec. siblings)
  - A/B partial A "missing a B to its right"
  - A\B partial A "missing a B to its left"
  - X if all else fails
- Slash labels are from combinatory categorial grammar
- About 4000 unique labels for English!

Translating with SAMT

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow [\text{que } VBP_1 \text{ les } NNS_2]::[\text{that } NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que } S/NP_1 \text{ les } NNS_2]::[\text{that } S/NP_1 \text{ the } NNS_2]$

$VBP \rightarrow$ [lisent]::[read]

Translating with SAMT

NNS → [enfants]::[children]

NP → [J']::[I]

S → [NP_1 $S\backslash NP_2$]::[NP_1 $S\backslash NP_2$]

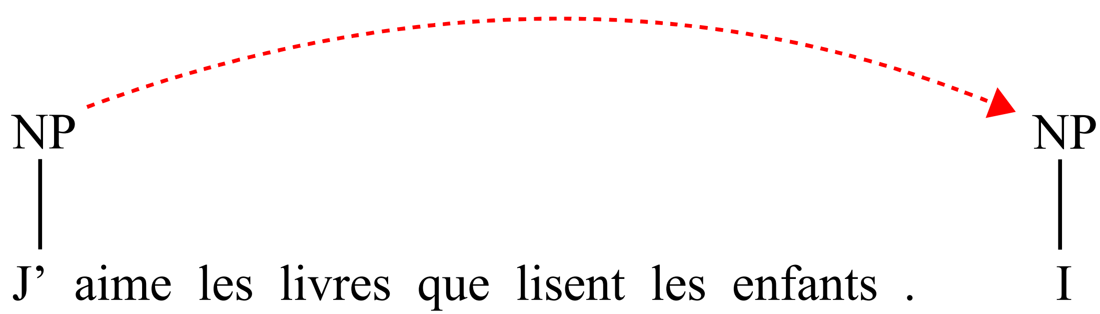
S → [S_1 $SBAR_2$.]::[S_1 $SBAR_2$.]

S\NP → [aime les livres]::[like books]

SBAR → [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

SBAR → [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

VBP → [lisent]::[read]



Translating with SAMT

NNS → [enfants]::[children]

NP → [J’]::[I]

S → [NP_1 $S\backslash NP_2$]::[NP_1 $S\backslash NP_2$]

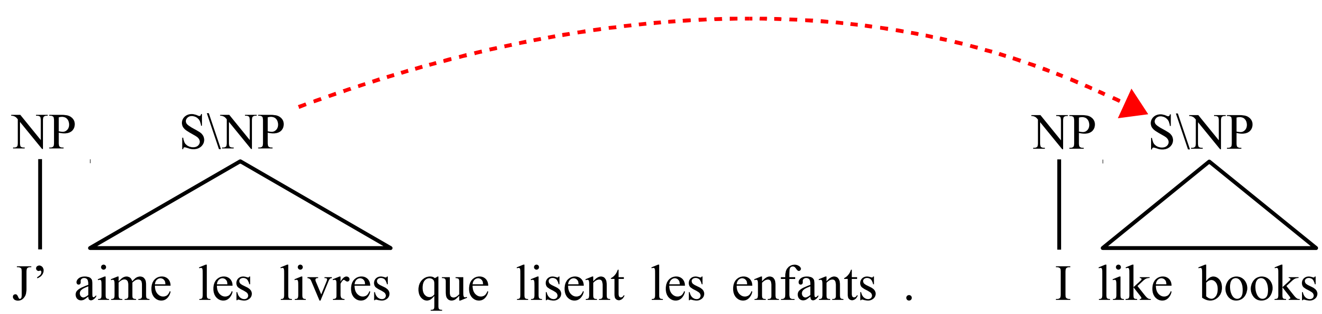
S → [S_1 $SBAR_2$.]::[S_1 $SBAR_2$.]

S\NP → [aime les livres]::[like books]

SBAR → [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

SBAR → [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

VBP → [lisent]::[read]



Translating with SAMT

NNS → [enfants]::[children]

NP → [J’]::[I]

S → [NP_1 $S\backslash NP_2$]::[NP_1 $S\backslash NP_2$]

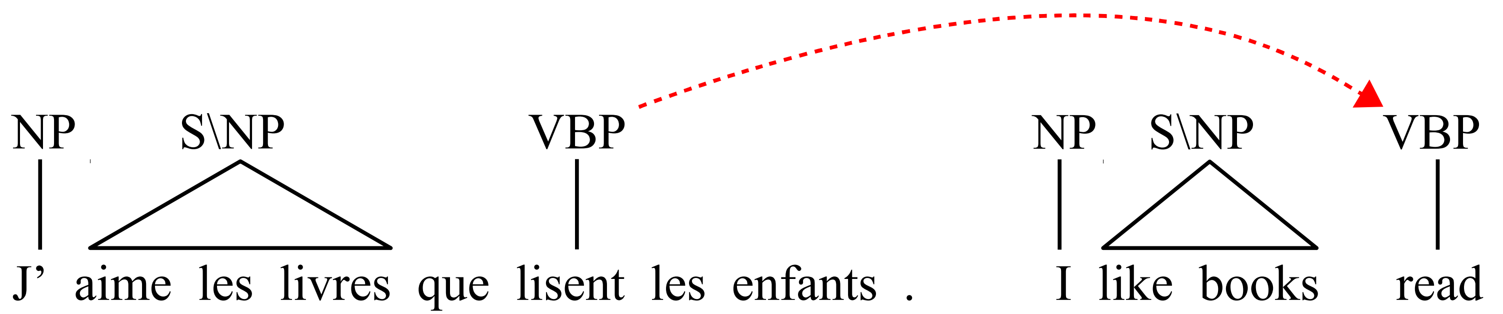
S → [S_1 $SBAR_2$.]::[S_1 $SBAR_2$.]

S\NP → [aime les livres]::[like books]

SBAR → [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

SBAR → [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

VBP → [lisent]::[read]



Translating with SAMT

NNS → [enfants]::[children]

NP → [J’]::[I]

S → [NP_1 $S\backslash NP_2$]::[NP_1 $S\backslash NP_2$]

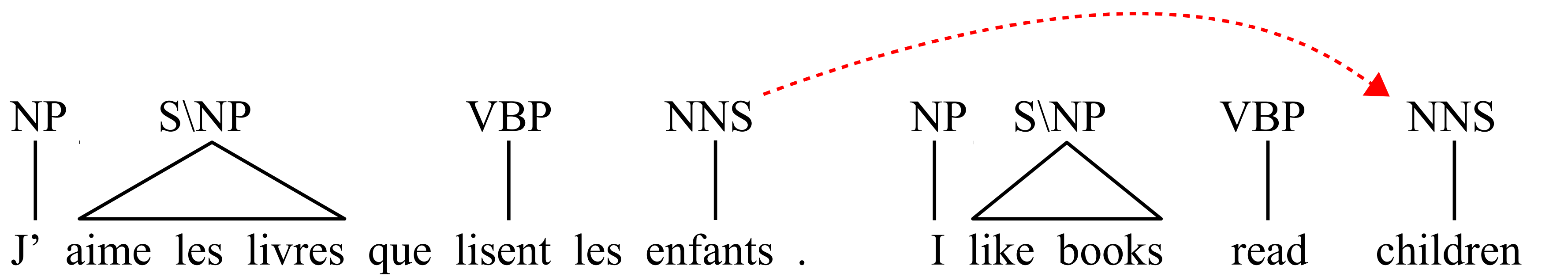
S → [S_1 $SBAR_2$.]::[S_1 $SBAR_2$.]

S\NP → [aime les livres]::[like books]

SBAR → [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

SBAR → [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

VBP → [lisent]::[read]



Translating with SAMT

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

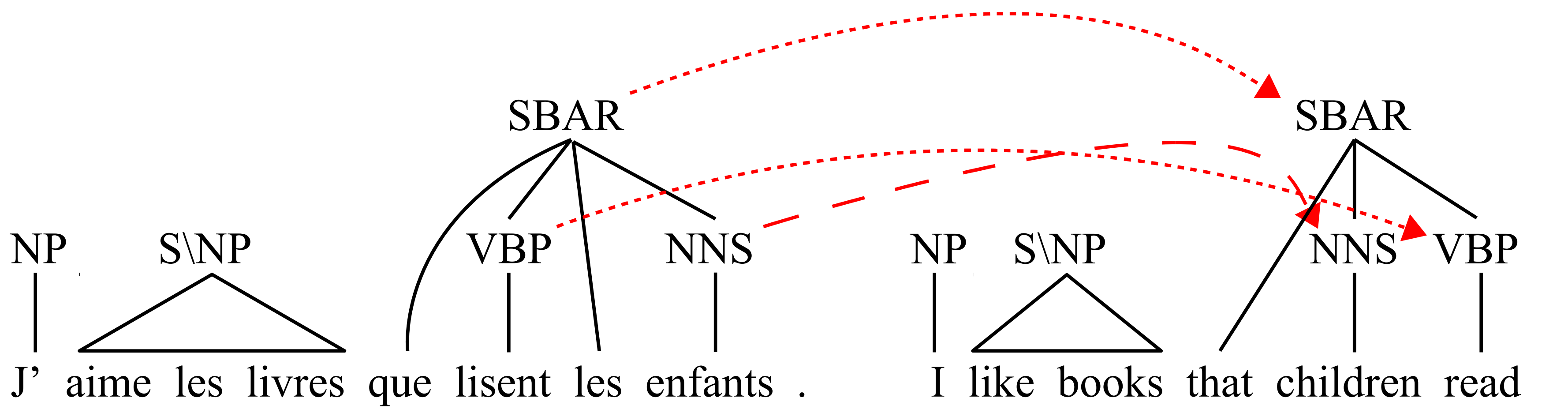
$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow$ [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

$SBAR \rightarrow$ [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

$VBP \rightarrow$ [lisent]::[read]



Translating with SAMT

NNS → [enfants]::[children]

NP → [J']::[I]

S → [NP$_1$ S\NP$_2$]::[NP$_1$ S\NP$_2$]

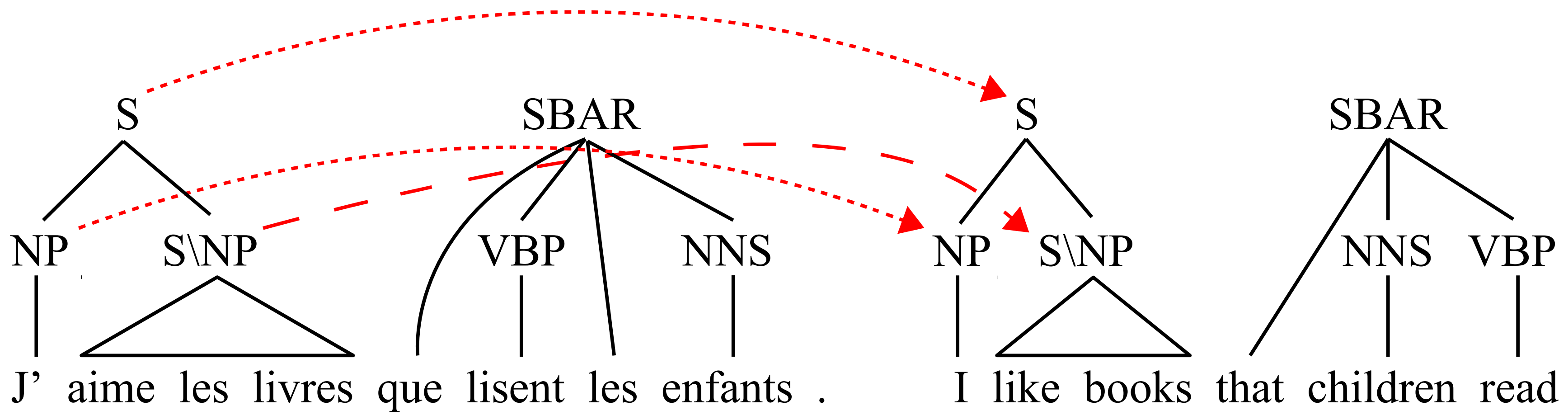
S → [S$_1$ SBAR$_2$.]::[S$_1$ SBAR$_2$.]

S\NP → [aime les livres]::[like books]

SBAR → [que VBP$_1$ les NNS$_2$]::[that NNS$_2$ VBP$_1$]

SBAR → [que S/NP$_1$ les NNS$_2$]::[that S/NP$_1$ the NNS$_2$]

VBP → [lisent]::[read]



Translating with SAMT

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

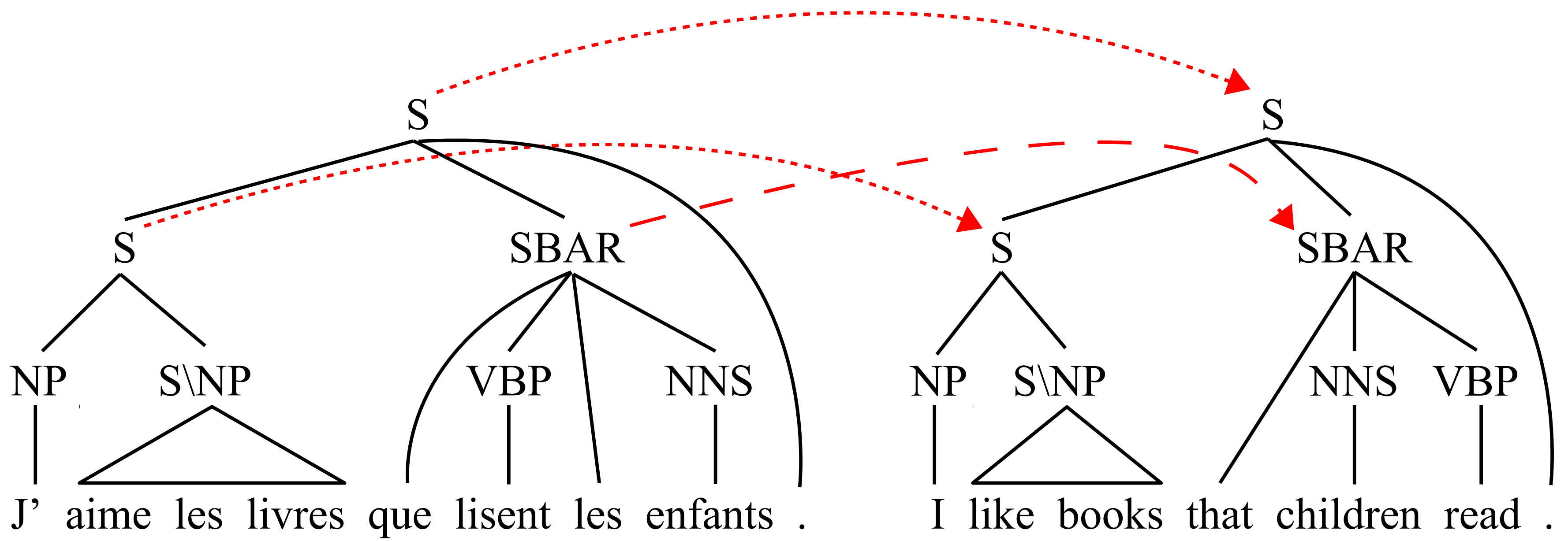
$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow [\text{que } VBP_1 \text{ les } NNS_2]::[\text{that } NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que } S/NP_1 \text{ les } NNS_2]::[\text{that } S/NP_1 \text{ the } NNS_2]$

$VBP \rightarrow$ [lisent]::[read]



Translating with SAMT

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

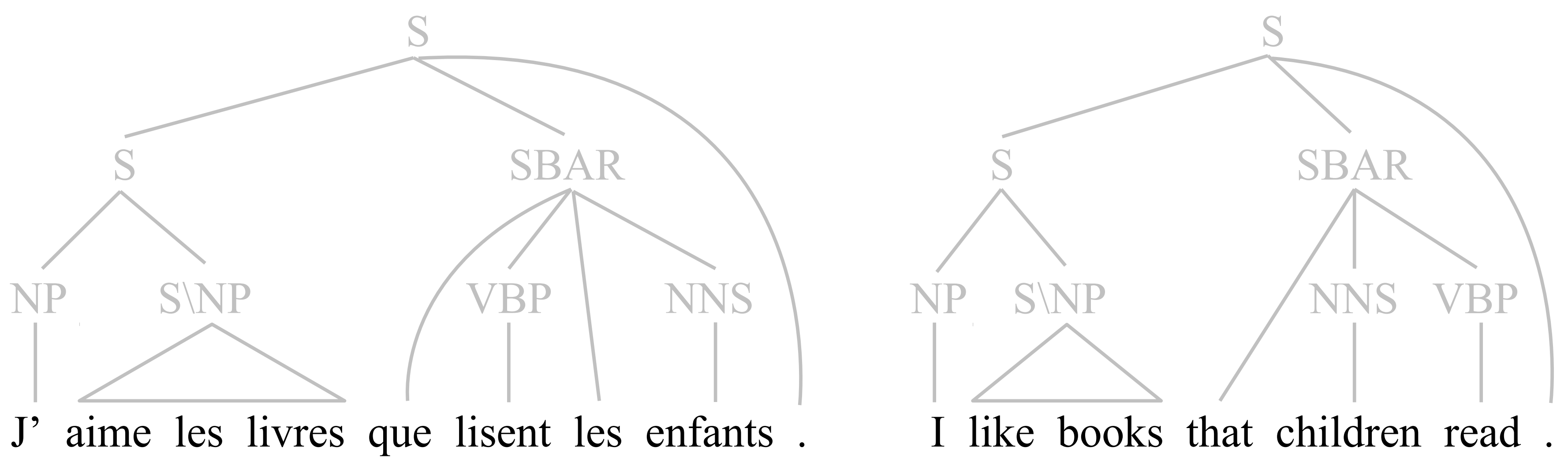
$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow [\text{que } VBP_1 \text{ les } NNS_2]::[\text{that } NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que } S/NP_1 \text{ les } NNS_2]::[\text{that } S/NP_1 \text{ the } NNS_2]$

$VBP \rightarrow$ [lisent]::[read]



The Importance of Labels

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow [\text{que } VBP_1 \text{ les } NNS_2]::[\text{that } NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que } S/NP_1 \text{ les } NNS_2]::[\text{that } S/NP_1 \text{ the } NNS_2]$

$\cancel{VBP} \rightarrow$ [lisent]::[read]

$VB \rightarrow$ [lisent]::[read]

J' aime les livres que lisent les enfants .

The Importance of Labels

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

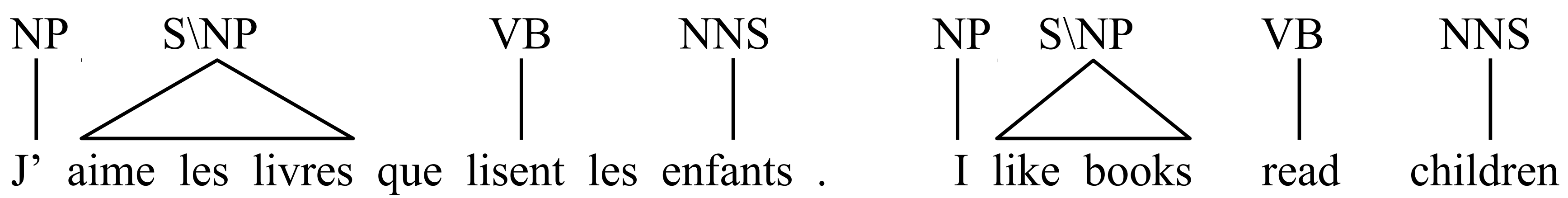
$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow [\text{que } VBP_1 \text{ les } NNS_2]::[\text{that } NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que } S/NP_1 \text{ les } NNS_2]::[\text{that } S/NP_1 \text{ the } NNS_2]$

$VB \rightarrow$ [lisent]::[read]



The Importance of Labels

NNS $\rightarrow$ [enfants]::[children]

NP $\rightarrow$ [J’]::[I]

S $\rightarrow$ [NP_1 $S\backslash NP_2$]::[NP_1 $S\backslash NP_2$]

S $\rightarrow$ [S_1 $SBAR_2$.]::[S_1 $SBAR_2$.]

S\NP $\rightarrow$ [aime les livres]::[like books]

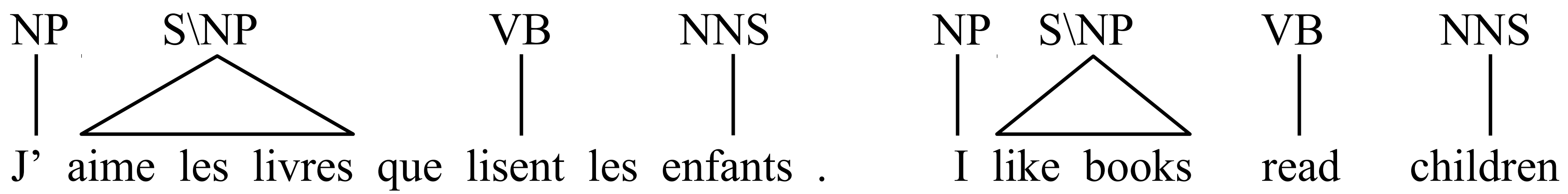
SBAR $\rightarrow$ [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

SBAR $\rightarrow$ [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

VB $\rightarrow$ [lisent]::[read]

VB ≠ VBP

No match!



The Importance of Labels

$NNS \rightarrow$ [enfants]::[children]

$NP \rightarrow$ [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

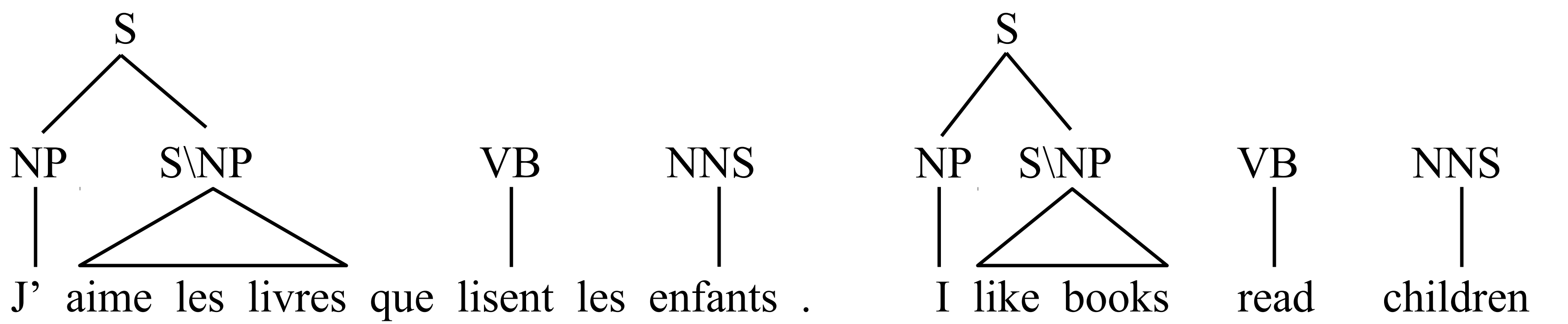
$S\backslash NP \rightarrow$ [aime les livres]::[like books]

$SBAR \rightarrow$ [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

$SBAR \rightarrow$ [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

$VB \rightarrow$ [lisent]::[read]

How do we finish parsing and translating?

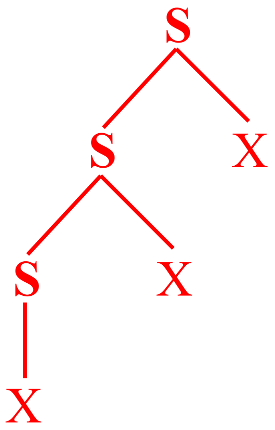


The Importance of Labels

- Label of constituent being parsed must exactly match right-hand side of rule
- If no match in grammar, need a fallback: **glue rules**
- Finish translation without "real" syntax

$$\mathbf{S} \rightarrow [X_1]::[X_1]$$

$$\mathbf{S} \rightarrow [\mathbf{S}_1\ X_2]::[\mathbf{S}_1\ X_2]$$



The Importance of Labels

- Label of constituent being parsed must exactly match right-hand side of rule
- If no match in grammar, need a fallback: **glue rules**
- Finish translation without “real” syntax

$$\mathbf{S} \rightarrow [X_1]::[X_1]$$

$$\mathbf{S} \rightarrow [\mathbf{S}_1\ X_2]::[\mathbf{S}_1\ X_2]$$

$$\mathbf{S} \rightarrow [JJ{+}NN_1]::[JJ{+}NN_1]$$

$$\mathbf{S} \rightarrow [\mathbf{S}_1\ JJ{+}NN_2]::[\mathbf{S}_1\ JJ{+}NN_2]$$

$$\mathbf{S} \rightarrow [NP_1]::[NP_1]$$

$$\mathbf{S} \rightarrow [\mathbf{S}_1\ NP_2]::[\mathbf{S}_1\ NP_2]$$

$$\mathbf{S} \rightarrow [VP\backslash VBZ_1]::[VP\backslash VBZ_1]$$

$$\mathbf{S} \rightarrow [\mathbf{S}_1\ VP\backslash VBZ_2]::[\mathbf{S}_1\ VP\backslash VBZ_2]$$

The Importance of Labels

- Label of constituent being parsed must exactly match right-hand side of rule
- If no match in grammar, need a fallback: **glue rules**
- Finish translation without “real” syntax

$\mathbf{S} \rightarrow [X_1]::[X_1]$ … $]::[JJ{+}NN_1]$

$\mathbf{S} \rightarrow [\mathbf{S}_1\ X_2]::[\mathbf{S}$ … $]::[\mathbf{S}_1\ JJ{+}NN_2]$

$\mathbf{S} \rightarrow [NP_1]::[N$ … $Z_1]::[VP\backslash VBZ_1]$

$\mathbf{S} \rightarrow [\mathbf{S}_1\ NP_2]::[\mathbf{S}_1\ N$ … $_2]$ … $[\mathbf{S}_1\ VP\backslash VBZ_2]::[\mathbf{S}_1\ VP\backslash VBZ_2]$

One pair for each nonterminal in the grammar!

Translating with Glue Rules

NNS → [enfants]::[children]

NP → [J']::[I]

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

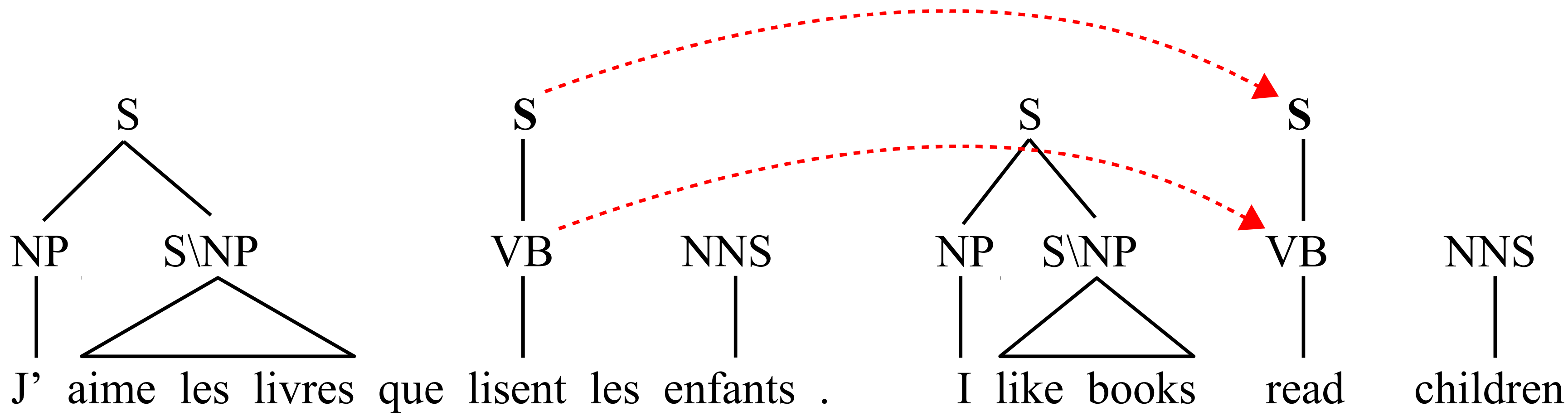
S\NP → [aime les livres]::[like books]

$SBAR \rightarrow [\text{que } VBP_1 \text{ les } NNS_2]::[\text{that } NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que } S/NP_1 \text{ les } NNS_2]::[\text{that } S/NP_1 \text{ the } NNS_2]$

VB → [lisent]::[read]

$\mathbf{S} \rightarrow [VB_1]::[VB_1]$



Translating with Glue Rules

NNS → [enfants]::[children]

NP → [J’]::[I]

S → [NP$_1$ S\NP$_2$]::[NP$_1$ S\NP$_2$]

S → [S$_1$ SBAR$_2$.]::[S$_1$ SBAR$_2$.]

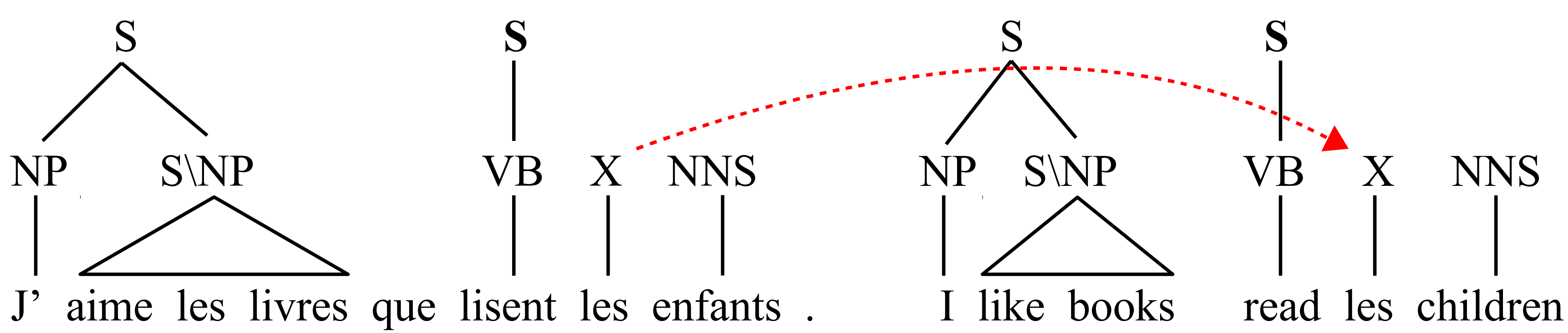
S\NP → [aime les livres]::[like books]

SBAR → [que VBP$_1$ les NNS$_2$]::[that NNS$_2$ VBP$_1$]

SBAR → [que S/NP$_1$ les NNS$_2$]::[that S/NP$_1$ the NNS$_2$]

VB → [lisent]::[read]

X → [les]::[les]



Translating with Glue Rules

NNS → [enfants]::[children]

NP → [J']::[I]

S → [NP_1 $S\backslash NP_2$]::[NP_1 $S\backslash NP_2$]

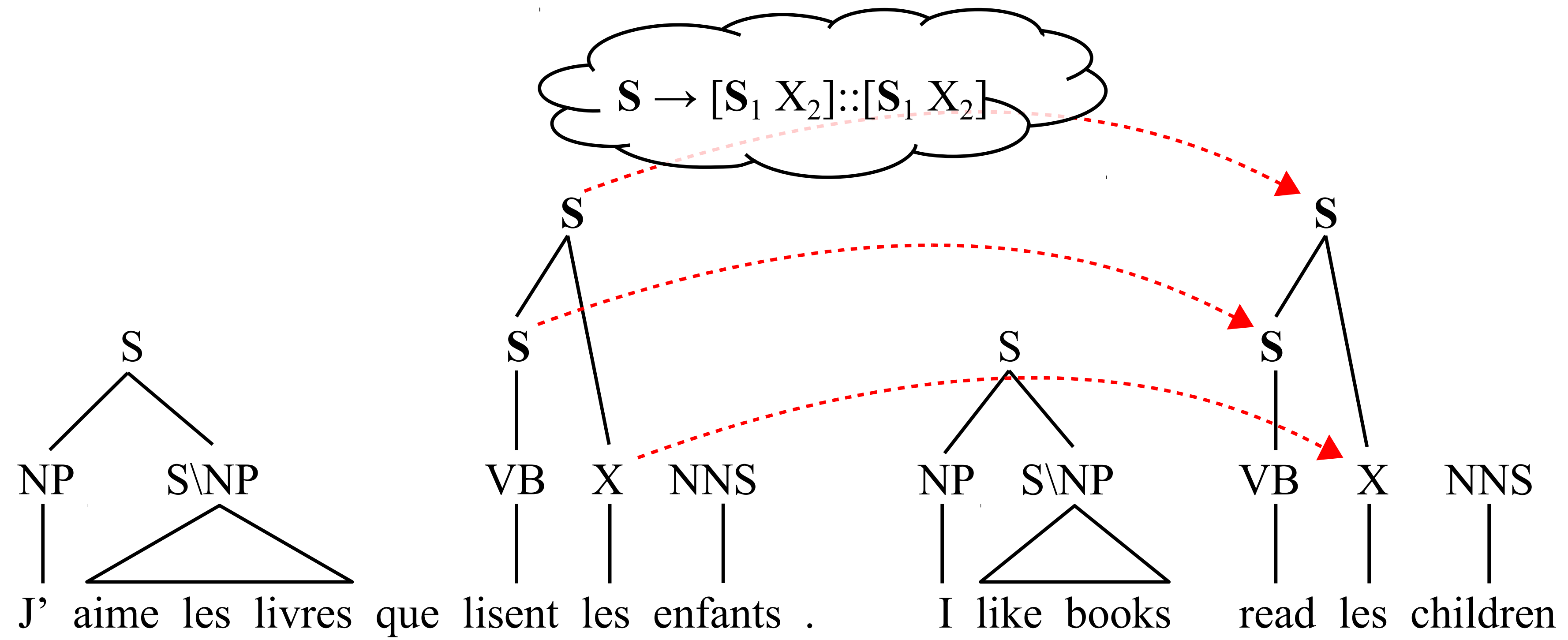
S → [S_1 $SBAR_2$.]::[S_1 $SBAR_2$.]

S\NP → [aime les livres]::[like books]

SBAR → [que VBP_1 les NNS_2]::[that NNS_2 VBP_1]

SBAR → [que S/NP_1 les NNS_2]::[that S/NP_1 the NNS_2]

VB → [lisent]::[read]



Translating with Glue Rules

$NNS \rightarrow [\text{enfants}]::[\text{children}]$

$NP \rightarrow [\text{J'}]::[\text{I}]$

$S \rightarrow [NP_1\ S\backslash NP_2]::[NP_1\ S\backslash NP_2]$

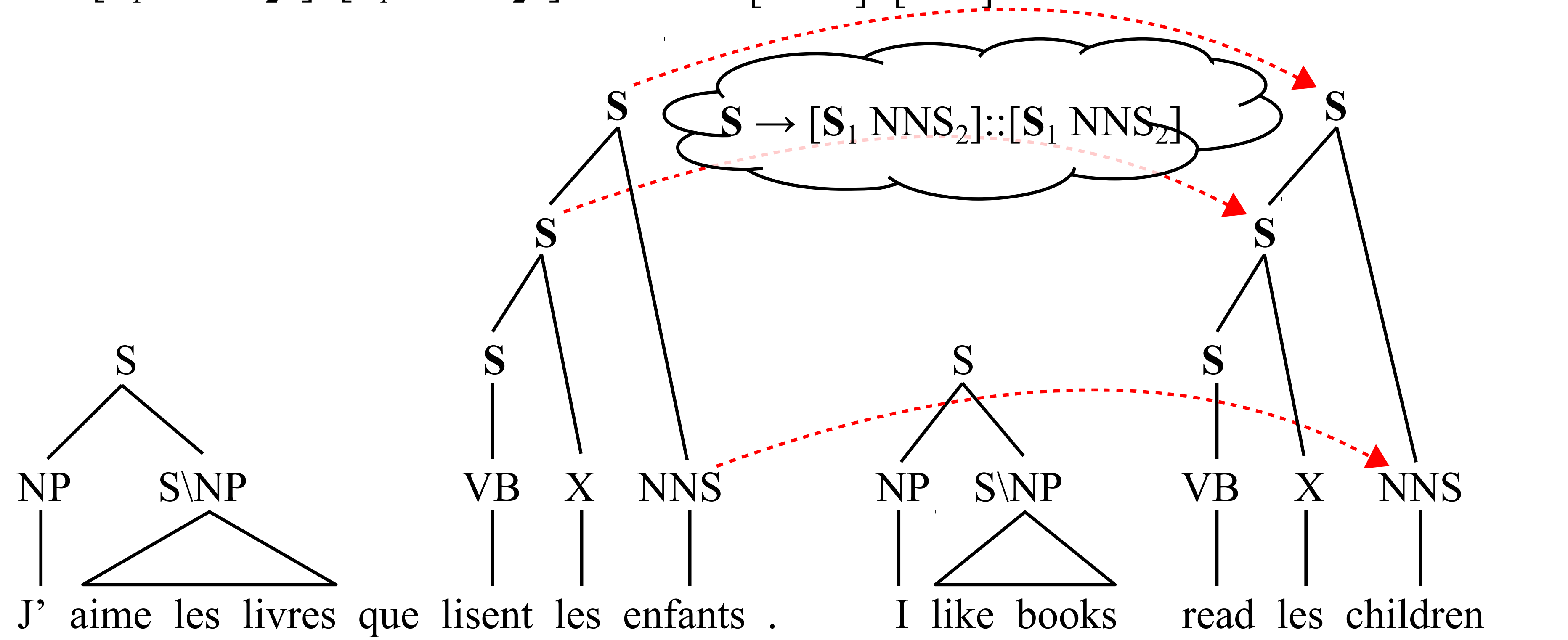
$S \rightarrow [S_1\ SBAR_2\ .]::[S_1\ SBAR_2\ .]$

$S\backslash NP \rightarrow [\text{aime les livres}]::[\text{like books}]$

$SBAR \rightarrow [\text{que}\ VBP_1\ \text{les}\ NNS_2]::[\text{that}\ NNS_2\ VBP_1]$

$SBAR \rightarrow [\text{que}\ S/NP_1\ \text{les}\ NNS_2]::[\text{that}\ S/NP_1\ \text{the}\ NNS_2]$

VB $\rightarrow$ [lisent]::[read]



SAMT Summary

- Label Hiero rules with nonterminals derived from parse trees
- Decisions made:
  - Syntax on: Target side
  - Rule formalism: SCFG
  - Input: String
  - Label space: Treebank + combinations
- Extensions: Venugopal et al. [2009], Chiang [2010], Zollmann [2011]

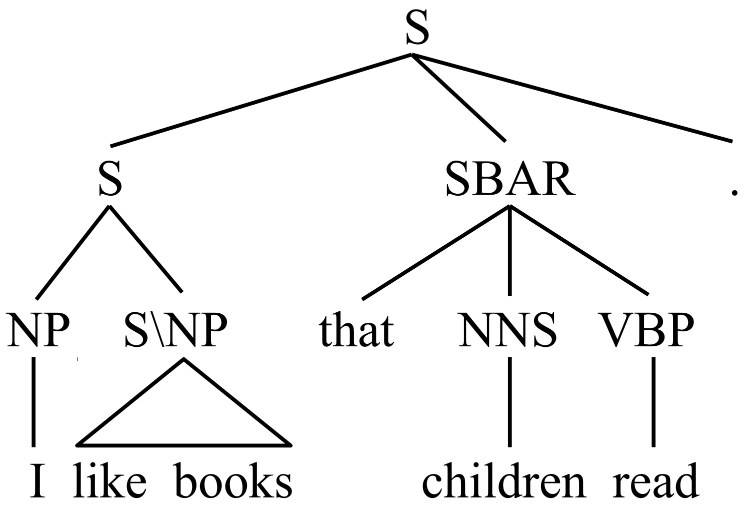
SAMT Pros and Cons

SAMT Pros and Cons

- No coverage loss from phrase-based MT
- Distinguishes different syntactic contexts

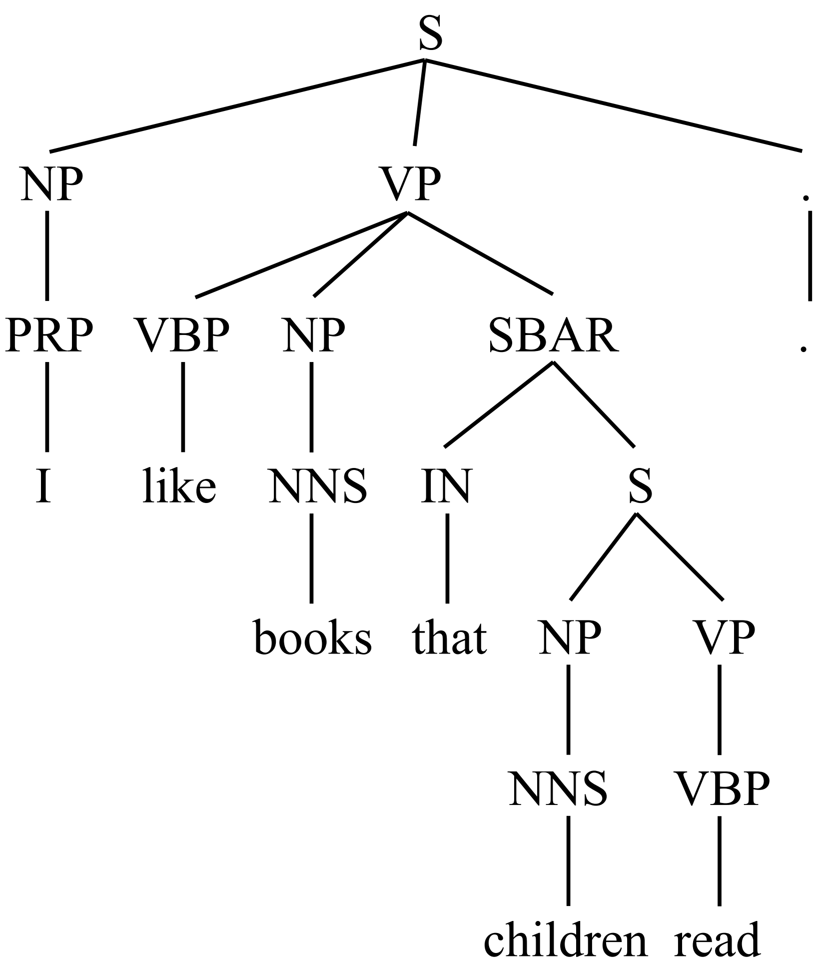
- Very large label set = sparse rules
- Still may need to restrict rule shape (Hiero)
- Produces weird tree structure

SAMT vs. “Native” Trees

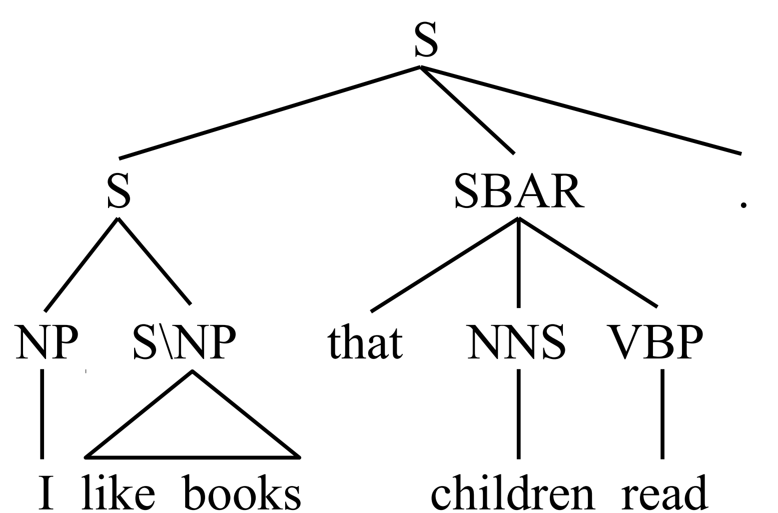


SAMT vs. "Native" Trees

Stanford Parser



SAMT decoding



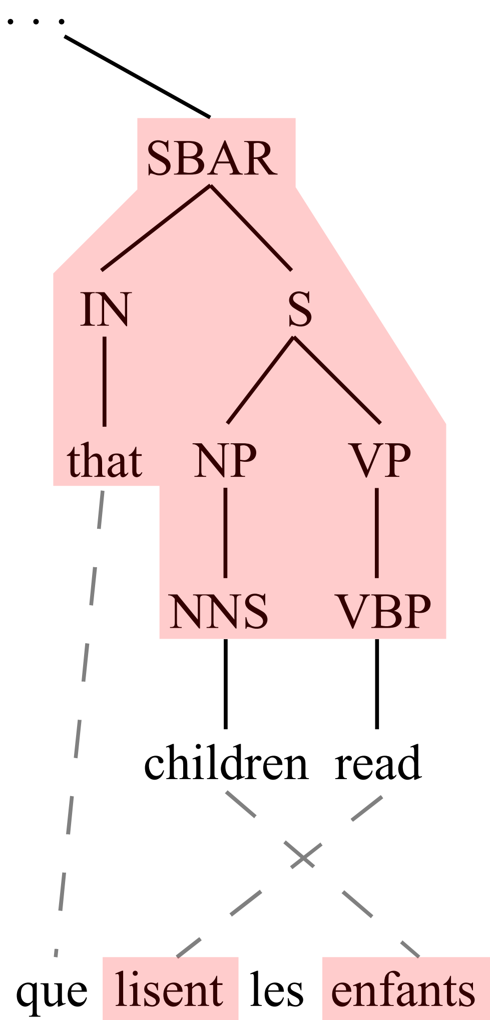
Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]

- Let’s use larger fragments of tree structure!
- Keep internal nodes in the rule
- Require exact constituent match, but relax Hiero/SAMT rule format constraints
- Needs generalization of SCFG: tree-to-string transducer rules (“xRs”) or synchronous tree-substitution grammar

Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]

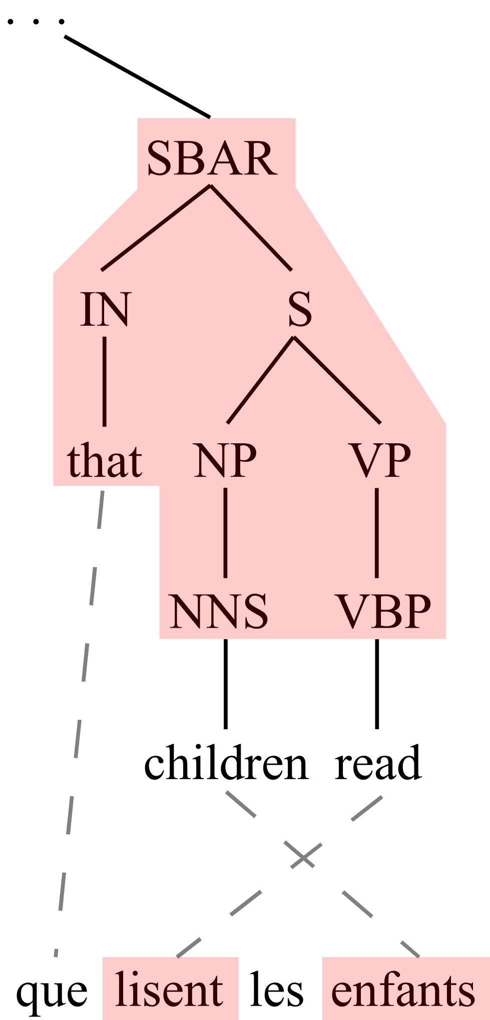


Extracted SAMT rule (SCFG)

$$\text{SBAR} \rightarrow [\text{que VBP}_1 \text{ les NNS}_2]::$$
$$[\text{that NNS}_2 \text{ VBP}_1]$$

Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]

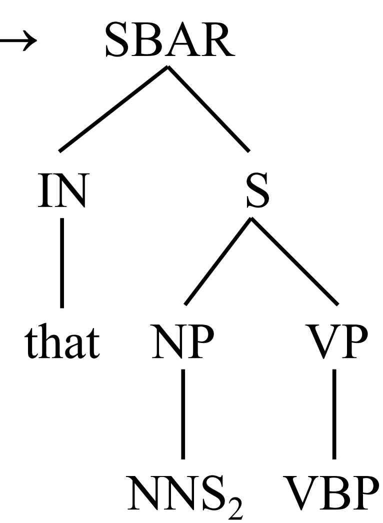


Extracted SAMT rule (SCFG)

$$\text{SBAR} \rightarrow [\text{que VBP}_1 \text{ les NNS}_2] :: [\text{that NNS}_2 \text{ VBP}_1]$$

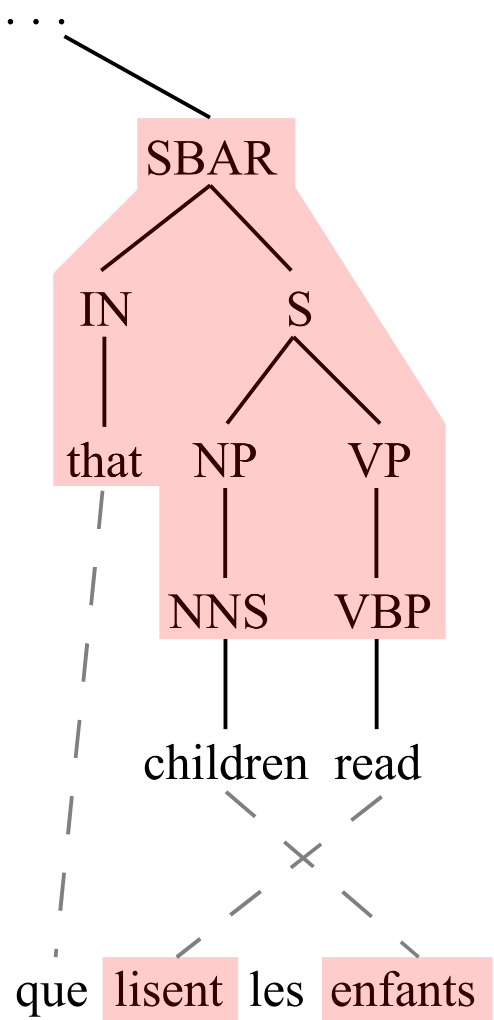
Extracted GHKM rule (xRs)

que VBP_1 les NNS_2 $\rightarrow$



Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]



Extracted SAMT rule (SCFG)

$$\text{SBAR} \rightarrow [\text{que VBP}_1 \text{ les NNS}_2]:: [\text{that NNS}_2 \text{ VBP}_1]$$

Extracted GHKM rule (xRs)

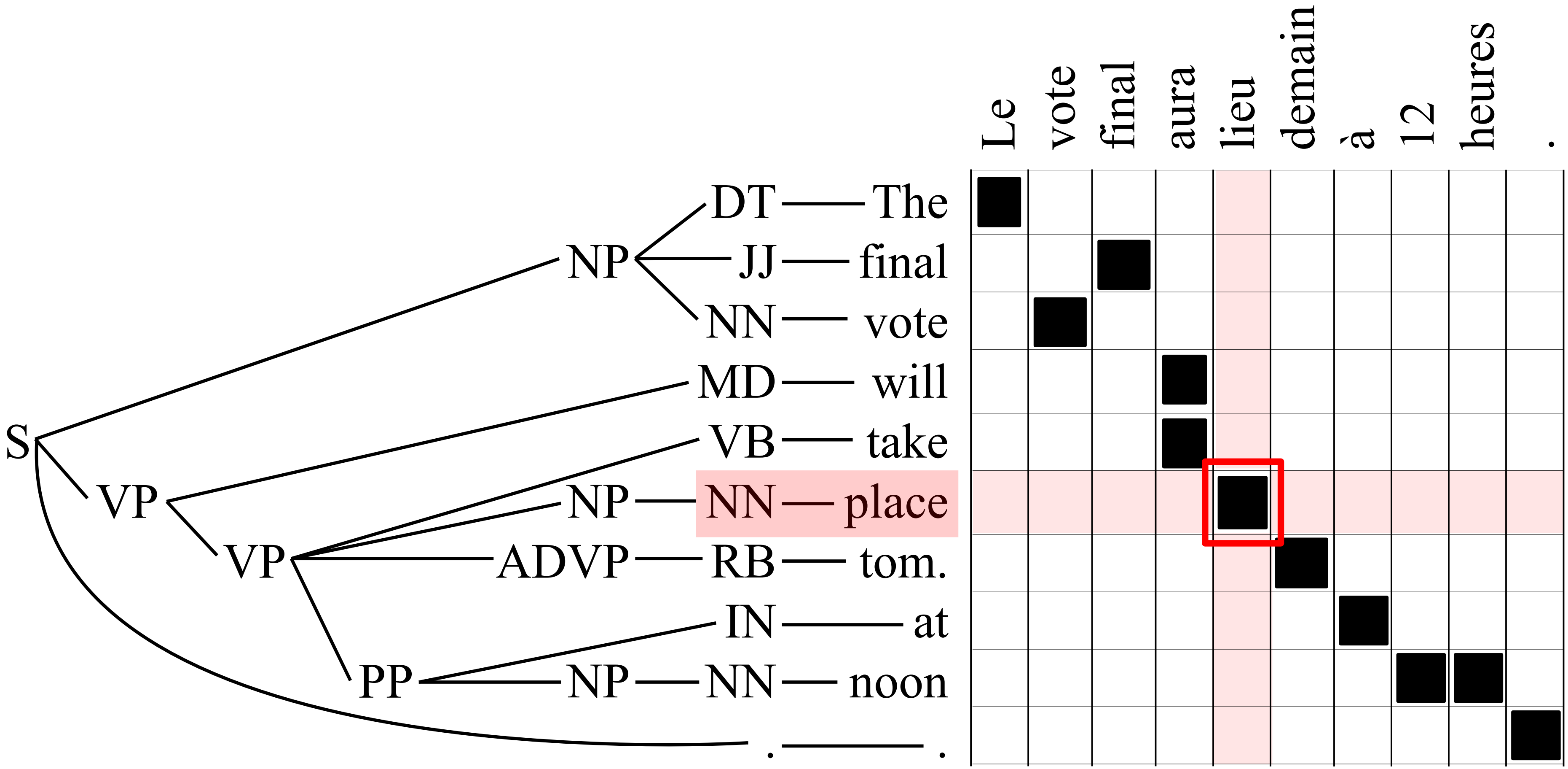
$$\text{que VBP}_1 \text{ les NNS}_2 \rightarrow \text{SBAR(IN(that) S(NP(NNS}_2\text{) VP(VBP}_1\text{)))}$$

Galley, Hopkins, Knight, Marcu

[Galley et al., 2004; Galley et al., 2006]

- Two extraction methods:
  - Minimal: break tree apart into smallest possible fragments
  - Composed: combine minimal fragments into larger ones that might overlap
- What about phrases that aren't constituents? Do we lose coverage?

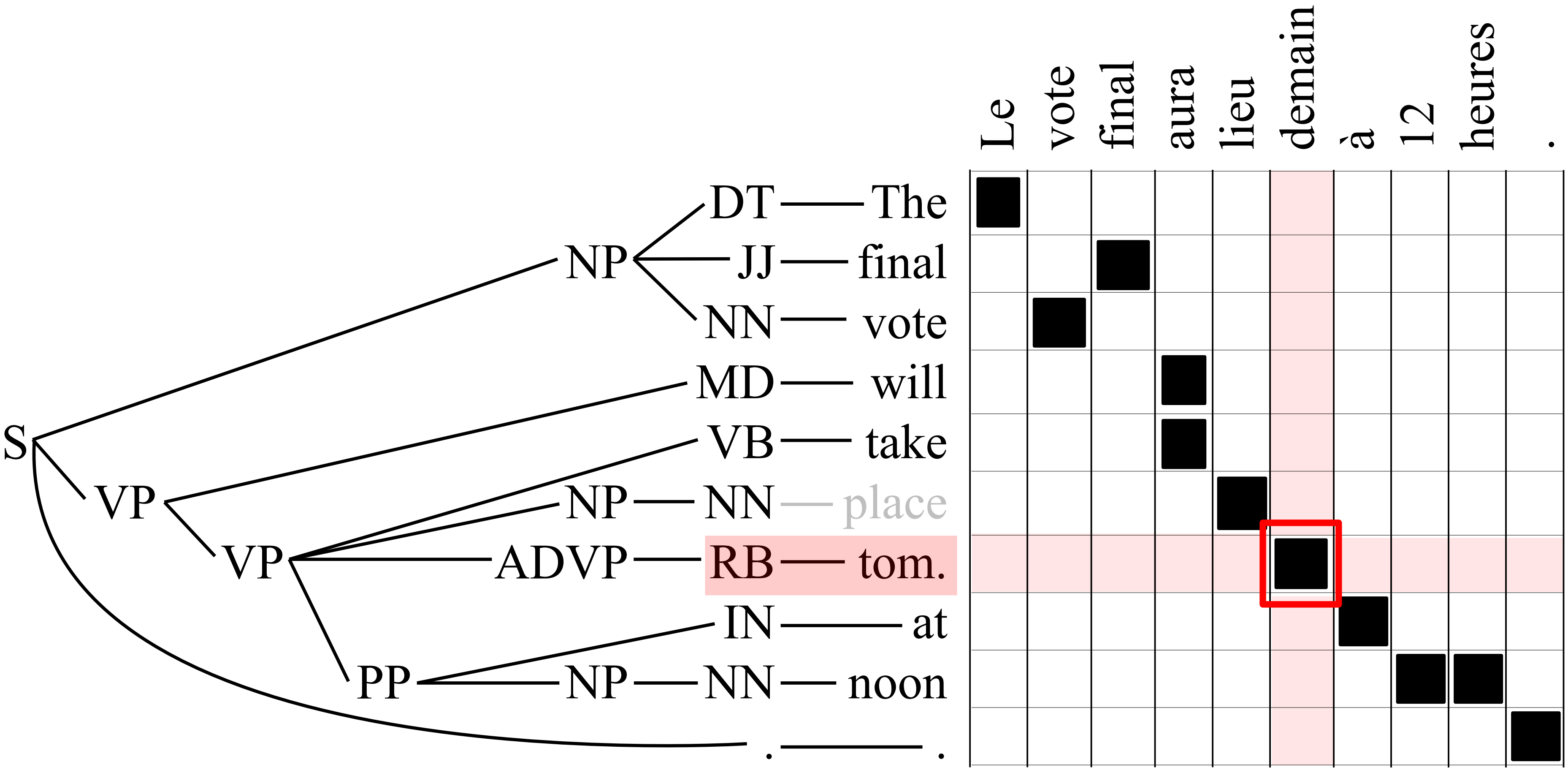
GHKM Minimal Rule Extraction



lieu → NN(place)

NN → [lieu]::[place]

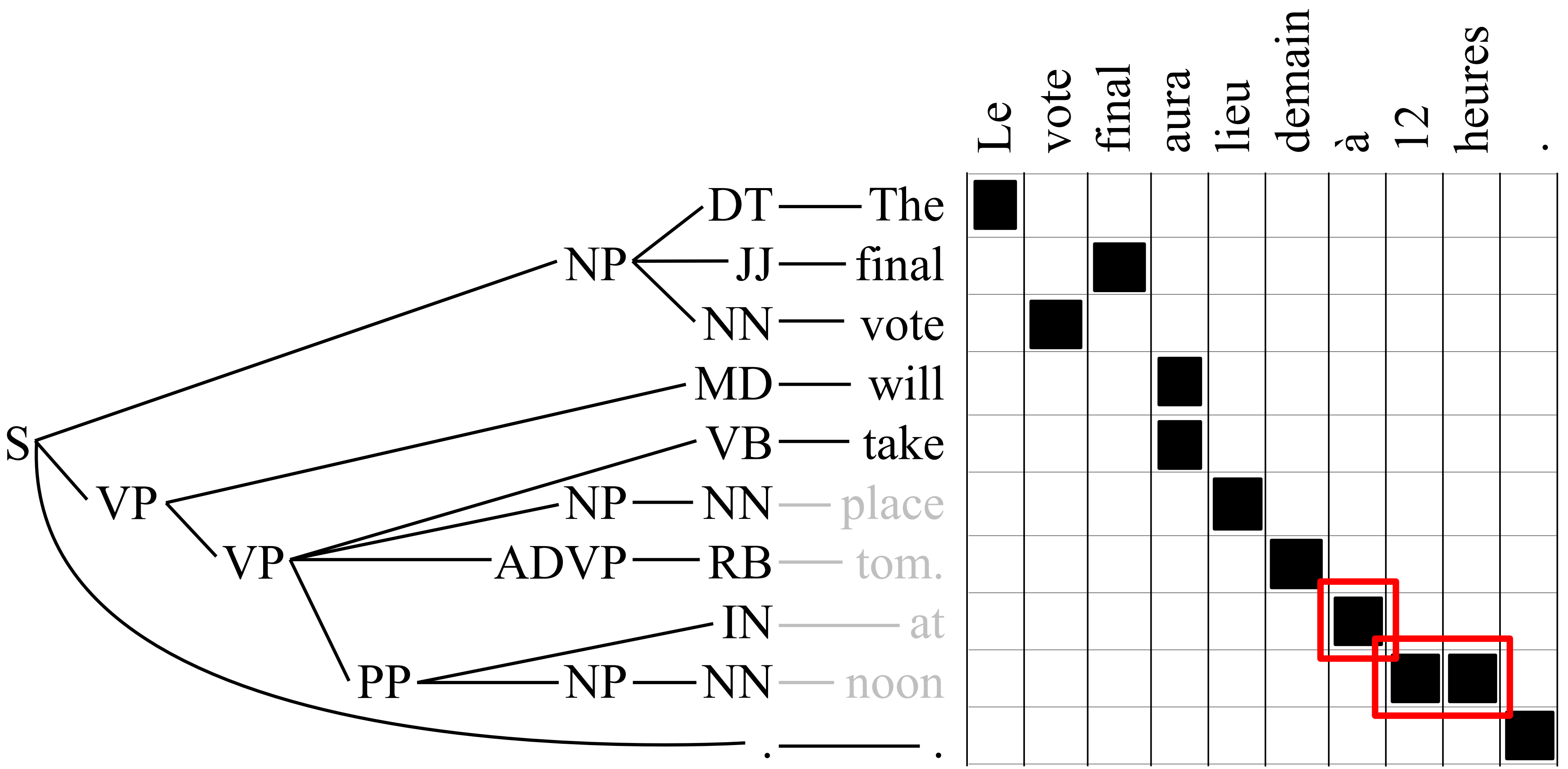
GHKM Minimal Rule Extraction



lieu → NN(place)

demain → RB(tomorrow)

GHKM Minimal Rule Extraction



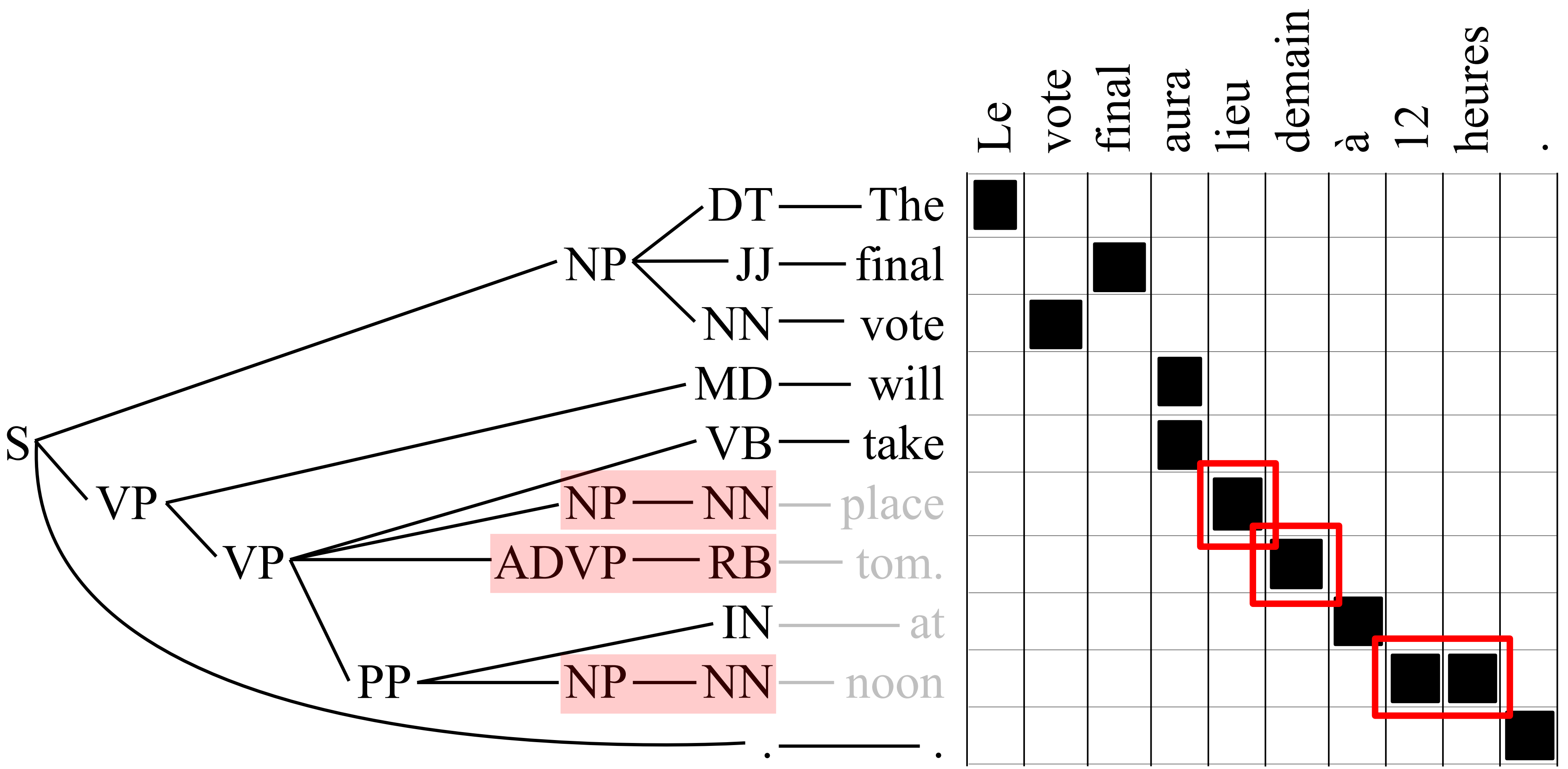
lieu → NN(place)

demain → RB(tomorrow)

à → IN(at)

12 heures → NN(noon)

GHKM Minimal Rule Extraction



lieu → NN(place)

demain → RB(tomorrow)

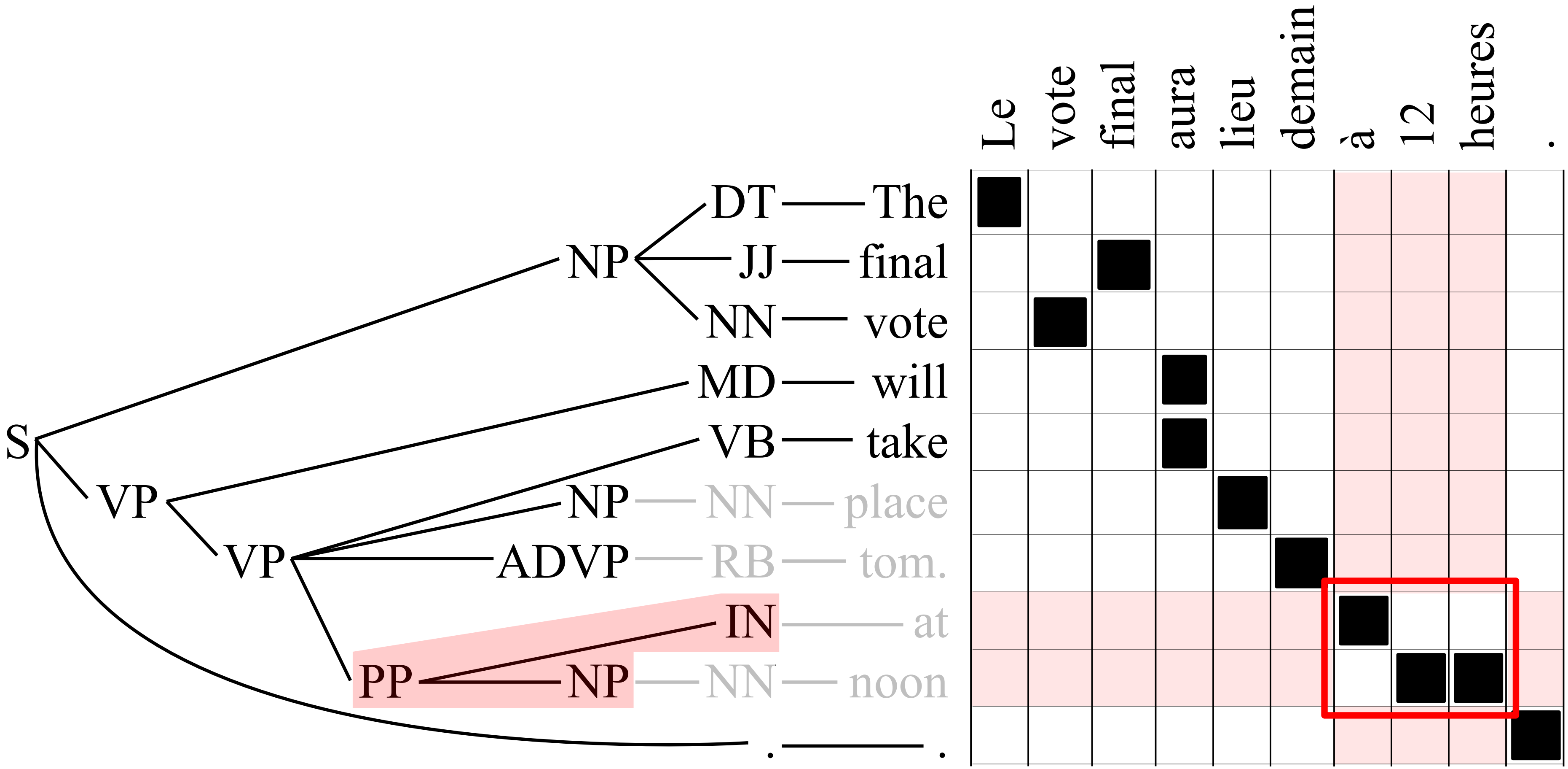
à → IN(at)

12 heures → NN(noon)

$NN_1 \rightarrow NP(NN_1)$ x2

$RB_1 \rightarrow ADVP(RB_1)$

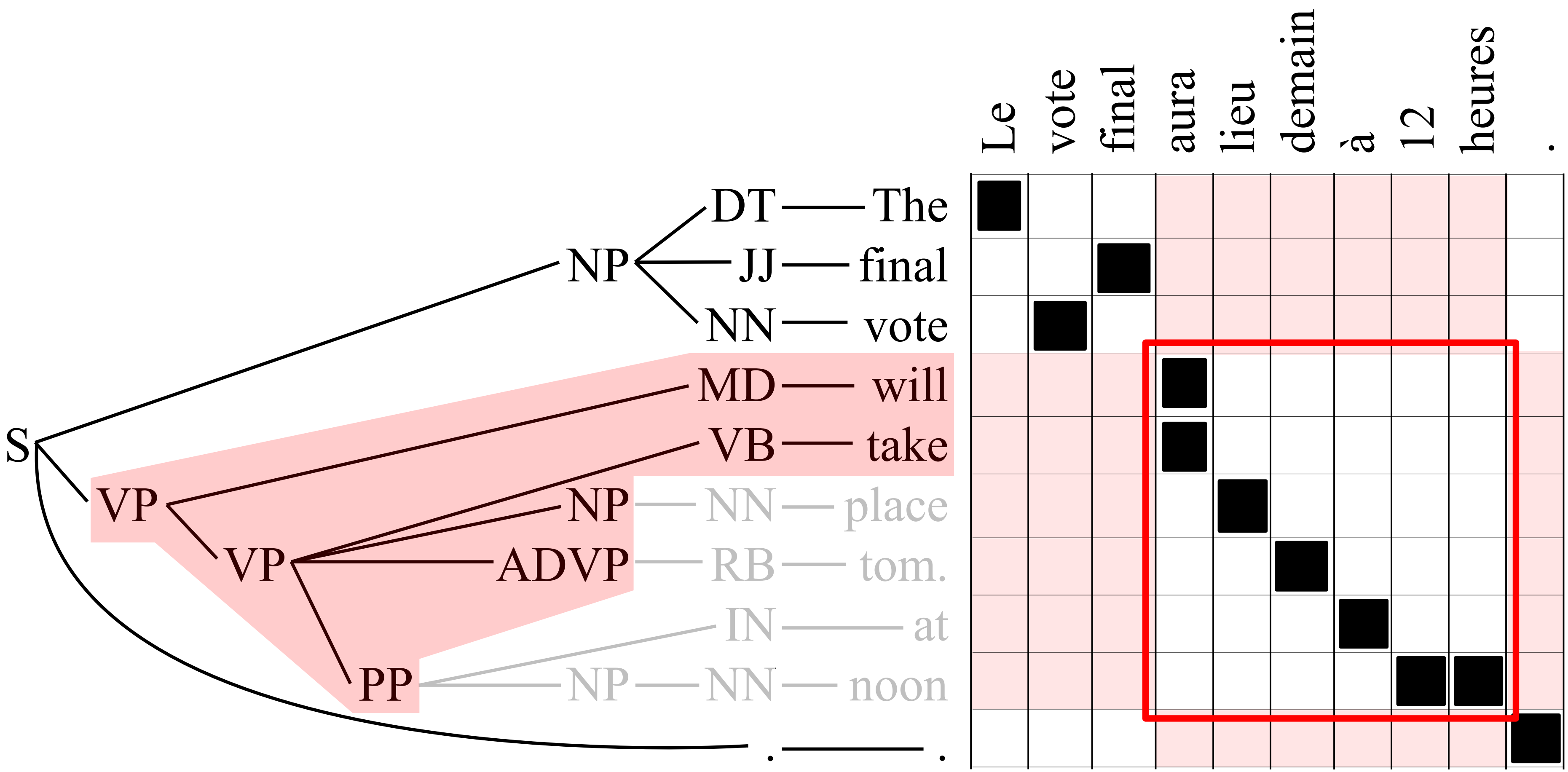
GHKM Minimal Rule Extraction



$\mathrm{IN}_1\ \mathrm{NP}_2 \rightarrow \mathrm{PP}(\mathrm{IN}_1\ \mathrm{NP}_2)$

$\mathrm{PP} \rightarrow [\mathrm{IN}_1\ \mathrm{NP}_2]::[\mathrm{IN}_1\ \mathrm{NP}_2]$

GHKM Minimal Rule Extraction

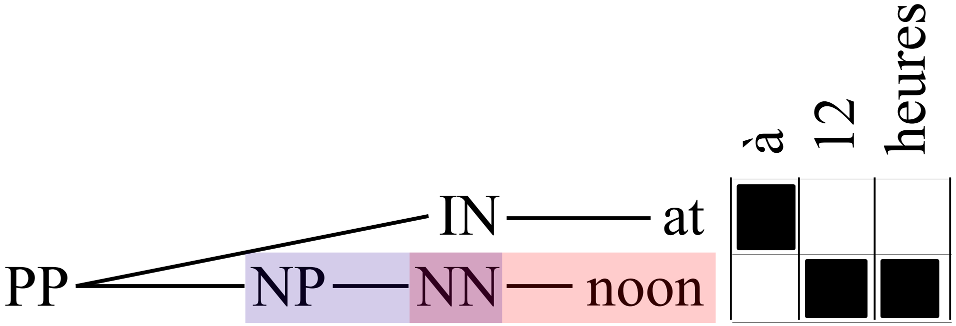


$IN_1\ NP_2 \rightarrow PP(IN_1\ NP_2)$

aura $NP_1\ ADVP_2\ PP_3 \rightarrow VP(MD(will)\ VP(VB(take)\ NP_1\ ADVP_2\ PP_3))$

GHKM Composed Rules

- When extracting minimal rules from a tree, plug two or more of them together



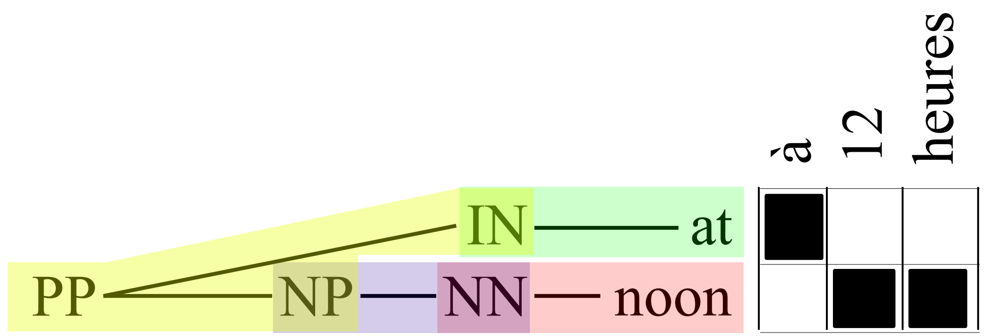
$$12 \text{ heures} \rightarrow \mathrm{NN}(\text{noon})$$

$$+ \quad \mathrm{NN}_1 \rightarrow \mathrm{NP}(\mathrm{NN}_1)$$

$$12 \text{ heures} \rightarrow \mathrm{NP}(\mathrm{NN}(\text{noon}))$$

GHKM Composed Rules

- When extracting minimal rules from a tree, plug two or more of them together



$$\text{12 heures} \rightarrow \text{NN(noon)}$$

$$\text{NN}_1 \rightarrow \text{NP}(\text{NN}_1)$$

$$\text{à} \rightarrow \text{IN(at)}$$

$$+ \quad \text{IN}_1\ \text{NP}_2 \rightarrow \text{PP}(\text{IN}_1\ \text{NP}_2)$$

$$\text{à 12 heures} \rightarrow \text{PP(IN(at) NP(NN(noon)))}$$

Consider the Following

- Where do unaligned words go?

- What happens at the top of the tree for a very long/flat sentence?

- Is the decomposition of the tree unique?

Consider the Following

- Where do unaligned words go?
  - Minimal: Attach at highest-level rule by fiat
  - Composed: Try attaching to all rules
- What happens at the top of the tree for a very long/flat sentence?
  - You get a very long (and likely sparse) S rule
- Is the decomposition of the tree unique?
  - Minimal: Yes, given unaligned words fiat
  - Composed: Definitely not!

Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow VP(VBP_1\ NP_2\ SBAR_3)$

Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$VP(VBP_1\ NP_2\ SBAR_3)$

J' aime les livres que lisent les enfants .

Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$VP(VBP_1\ NP_2\ SBAR_3)$

PRP

I aime les livres que lisent les enfants .

Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J’ $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$VP(VBP_1\ NP_2\ SBAR_3)$

PRP VBP

I like les livres que lisent les enfants .

Translating with GHKM

$\text{aime} \rightarrow \text{VBP(like)}$

$\text{enfants} \rightarrow \text{NNS(children)}$

$\text{J'} \rightarrow \text{PRP(I)}$

$\text{lisent} \rightarrow \text{VBP(read)}$

$\text{livres} \rightarrow \text{NNS(books)}$

$\text{que} \rightarrow \text{IN(that)}$

$. \rightarrow .(.)$

$\text{les NNS}_1 \rightarrow \text{NP(NNS}_1\text{)}$

$\text{IN}_1\ \text{S}_2 \rightarrow \text{SBAR(IN}_1\ \text{S}_2\text{)}$

$\text{NP}_1\ \text{VP}_2\ ._3 \rightarrow \text{S(NP}_1\ \text{VP}_2\ ._3\text{)}$

$\text{PRP}_1 \rightarrow \text{NP(PRP}_1\text{)}$

$\text{VBP}_1 \rightarrow \text{VP(VBP}_1\text{)}$

$\text{VP}_1\ \text{NP}_2 \rightarrow \text{S(NP}_2\ \text{VP}_1\text{)}$

$\text{VBP}_1\ \text{NP}_2\ \text{SBAR}_3 \rightarrow$
$\quad \text{VP(VBP}_1\ \text{NP}_2\ \text{SBAR}_3\text{)}$

PRP	VBP		NNS	IN	VBP		NNS	.
I	like	les	books	that	read	les	children	.

Translating with GHKM

aime → VBP(like)

enfants → NNS(children)

J’ → PRP(I)

lisent → VBP(read)

livres → NNS(books)

que → IN(that)

. → .(.)

les NNS_1 → $NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

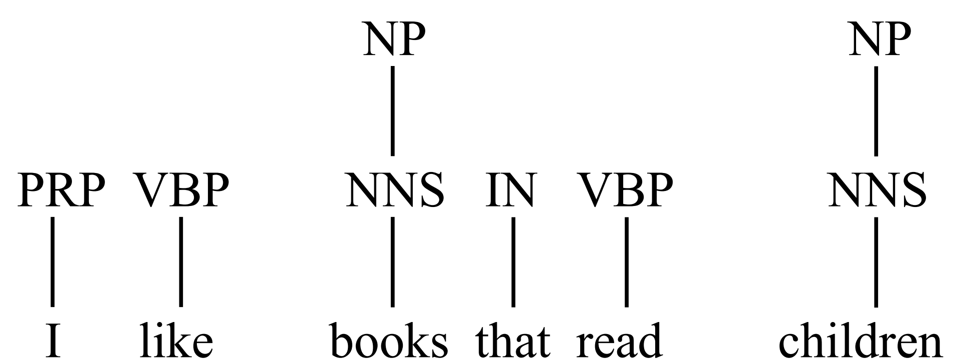
$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$VP(VBP_1\ NP_2\ SBAR_3)$



Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

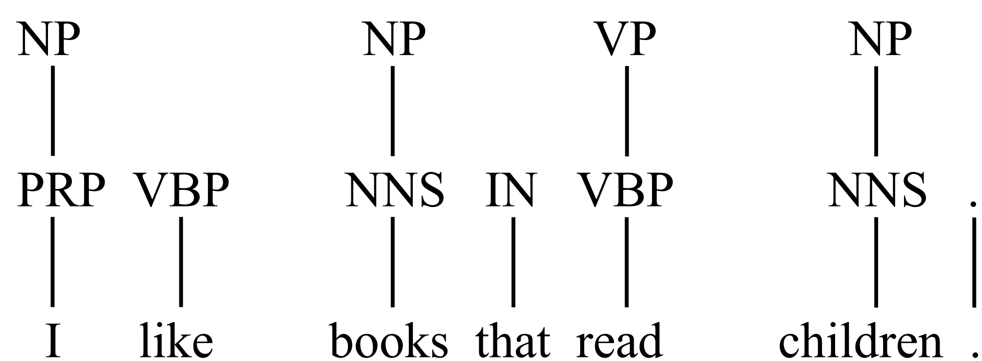
$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$VP(VBP_1\ NP_2\ SBAR_3)$



Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

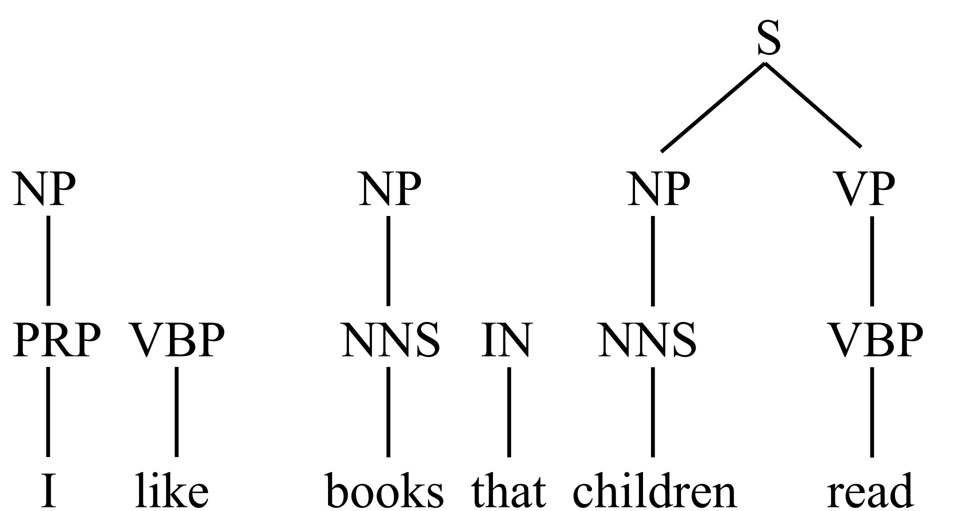
$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$\quad VP(VBP_1\ NP_2\ SBAR_3)$



Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

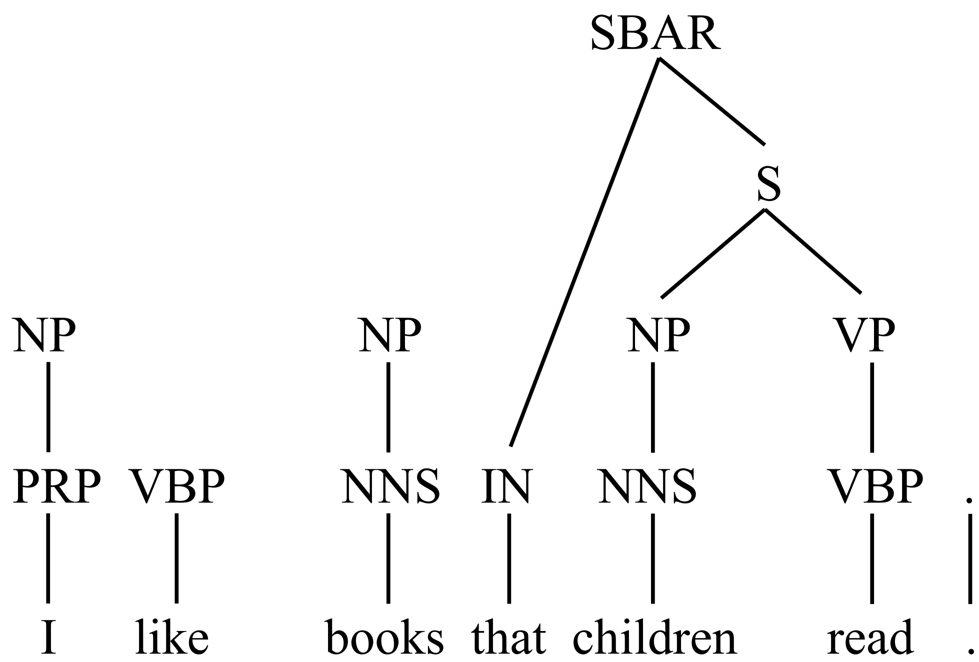
$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$\quad VP(VBP_1\ NP_2\ SBAR_3)$



Translating with GHKM

aime → VBP(like)

enfants → NNS(children)

J' → PRP(I)

lisent → VBP(read)

livres → NNS(books)

que → IN(that)

. → .(.)

les NNS_1 → $NP(NNS_1)$

IN_1 S_2 → $SBAR(IN_1\ S_2)$

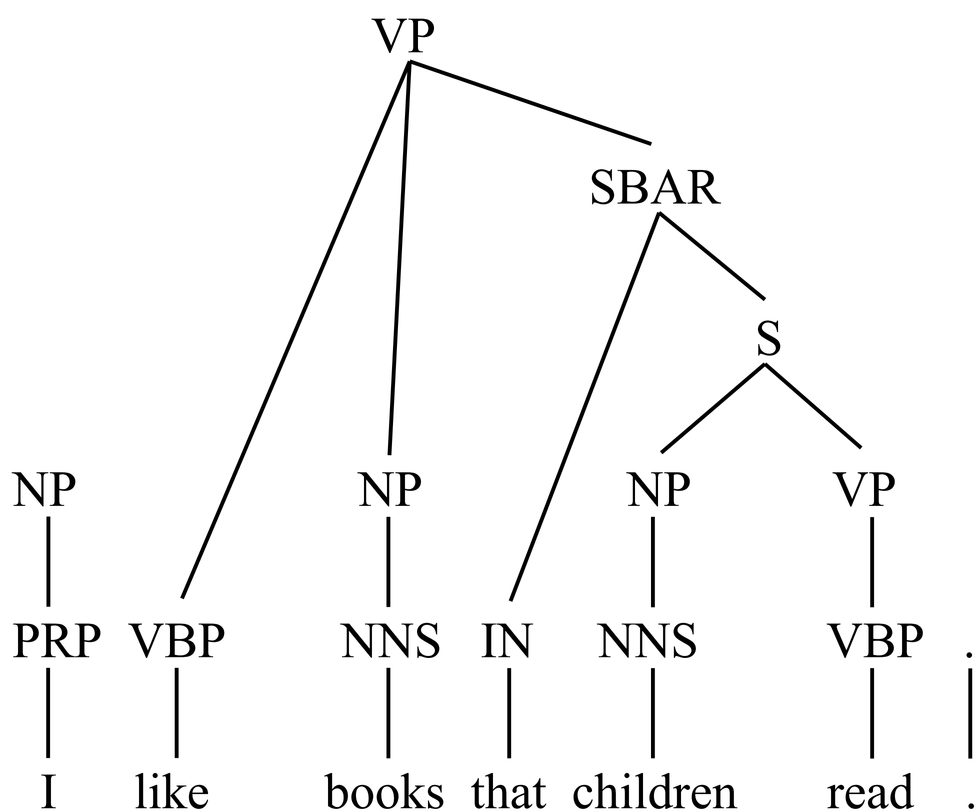
NP_1 VP_2 $._3$ → $S(NP_1\ VP_2\ ._3)$

PRP_1 → $NP(PRP_1)$

VBP_1 → $VP(VBP_1)$

VP_1 NP_2 → $S(NP_2\ VP_1)$

VBP_1 NP_2 $SBAR_3$ →
$VP(VBP_1\ NP_2\ SBAR_3)$



Translating with GHKM

aime $\rightarrow$ VBP(like)

enfants $\rightarrow$ NNS(children)

J' $\rightarrow$ PRP(I)

lisent $\rightarrow$ VBP(read)

livres $\rightarrow$ NNS(books)

que $\rightarrow$ IN(that)

. $\rightarrow$.(.)

les $NNS_1 \rightarrow NP(NNS_1)$

$IN_1\ S_2 \rightarrow SBAR(IN_1\ S_2)$

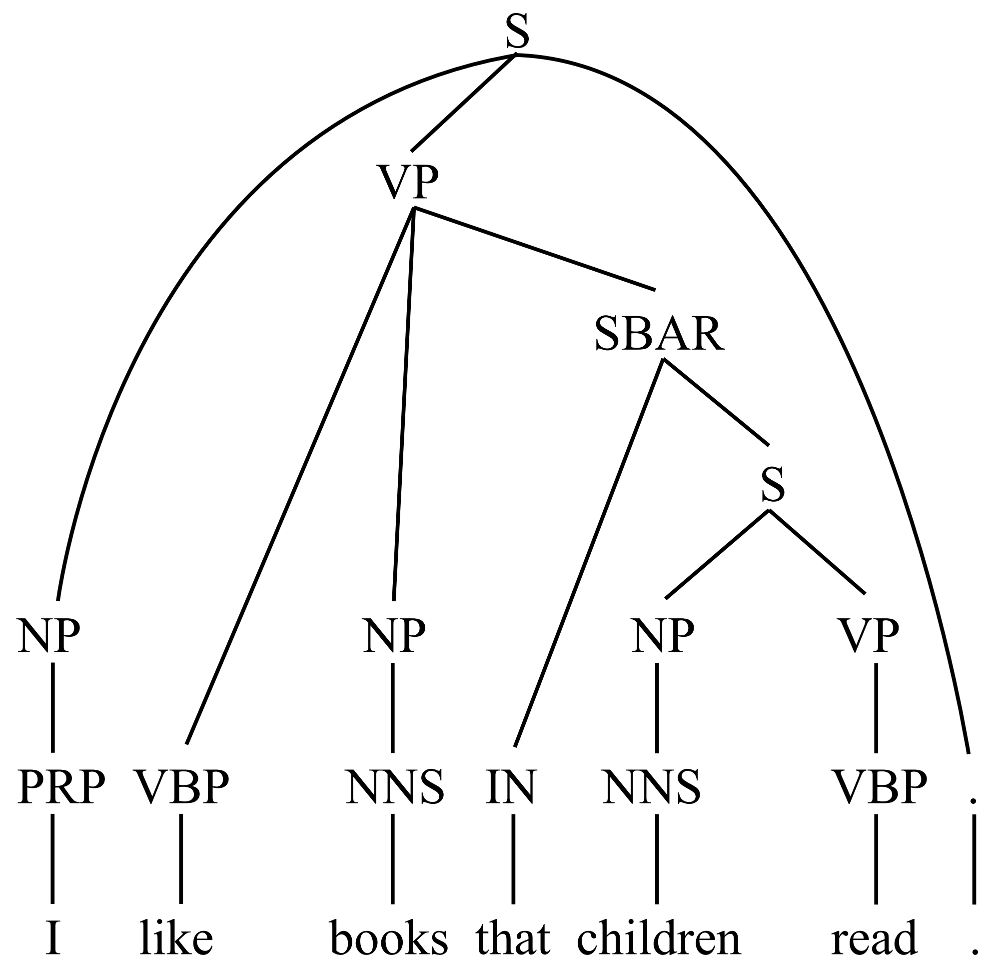
$NP_1\ VP_2\ ._3 \rightarrow S(NP_1\ VP_2\ ._3)$

$PRP_1 \rightarrow NP(PRP_1)$

$VBP_1 \rightarrow VP(VBP_1)$

$VP_1\ NP_2 \rightarrow S(NP_2\ VP_1)$

$VBP_1\ NP_2\ SBAR_3 \rightarrow$
$VP(VBP_1\ NP_2\ SBAR_3)$



GHKM Summary

- Extract chunks of syntactic structure; recursively transform string into tree
- Decisions made:
  - Syntax on: Target side
  - Rule formalism: String-to-tree transducer
  - Input: String
  - Label space: Treebank
- Extensions and analysis: DeNeefe et al. [2007], Huck et al. [2014]

GHKM Pros and Cons

GHKM Pros and Cons

- More “native-looking” trees (LMs, etc.)
- No restrictions on rule shape or size

- Smaller label set than SAMT

- Can need big rules at top of tree
- No coverage for non-constituents
- Unary chains permit many label changes

Part 2

Tree-to-Tree Methods

Parse Trees on Both Sides

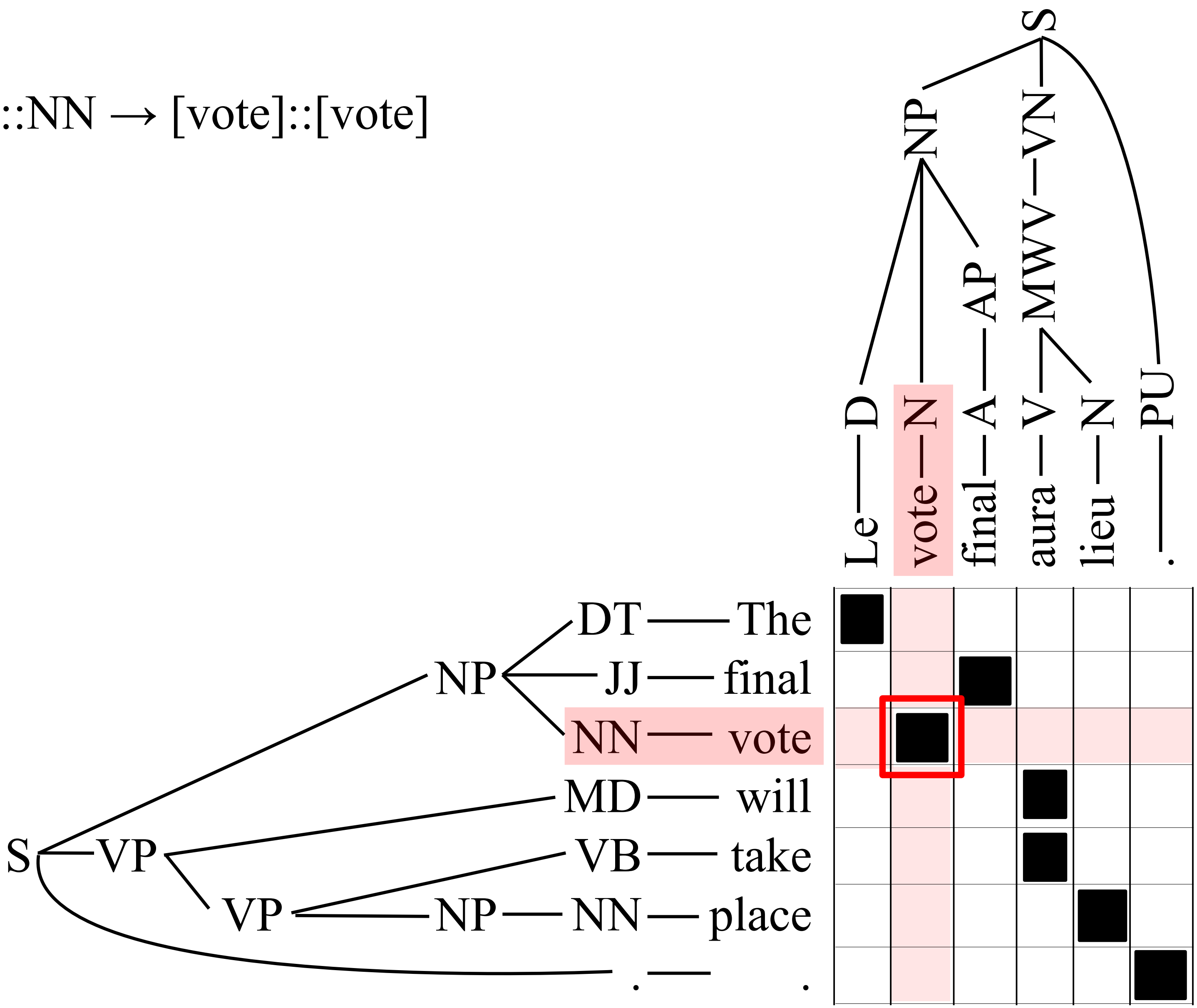
- Why add the source-side tree?
- Why could that also be a bad idea?

Parse Trees on Both Sides

- Why add the source-side tree?
  - Further disambiguate syntactic context
  - Could parse input and transform it: don't have to guess constituent boundaries
- Why could that also be a bad idea?
  - Could increase rule sparsity (e.g. labels)
  - Could hurt coverage if required to exactly match constituents on both sides
  - Only works for source and target languages that both have parsers

Tree-to-Tree Rule Extraction

N::NN → [vote]::[vote]



Tree-to-Tree Rule Extraction

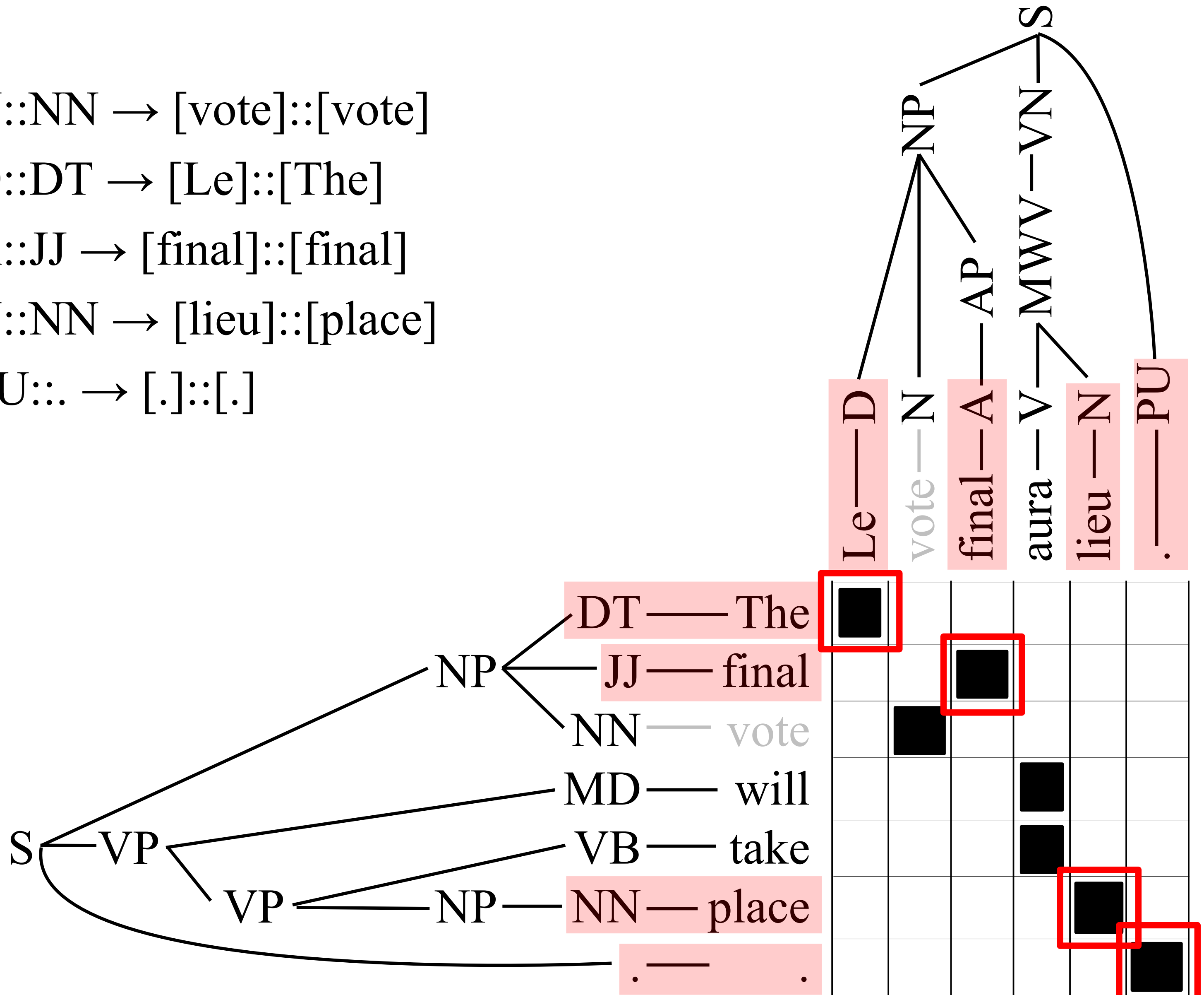
N::NN → [vote]::[vote]

D::DT → [Le]::[The]

A::JJ → [final]::[final]

N::NN → [lieu]::[place]

PU::. → [.]::[.]



Tree-to-Tree Rule Extraction

$\text{N::NN} \rightarrow \text{[vote]::[vote]}$

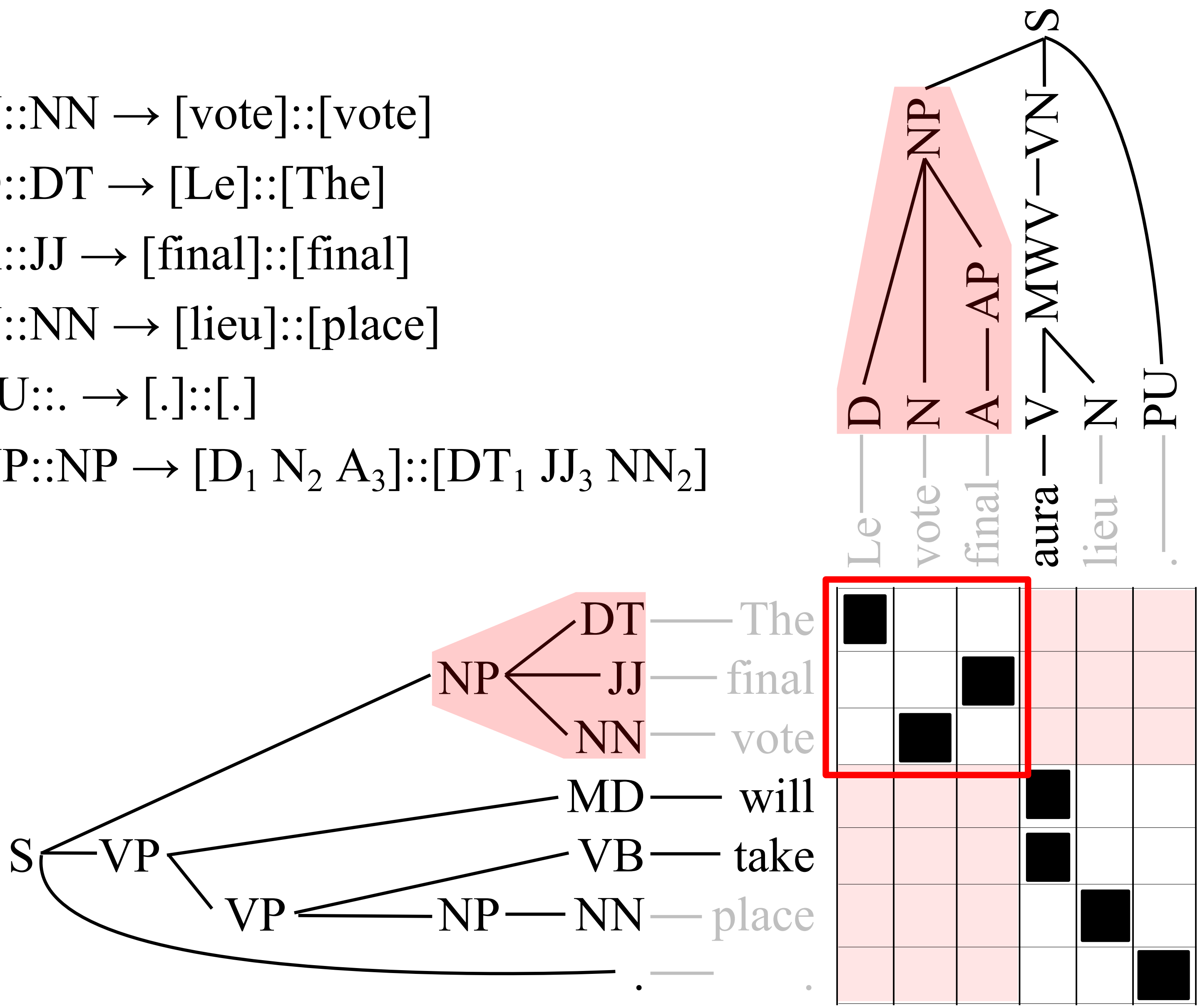
$\text{D::DT} \rightarrow \text{[Le]::[The]}$

$\text{A::JJ} \rightarrow \text{[final]::[final]}$

$\text{N::NN} \rightarrow \text{[lieu]::[place]}$

$\text{PU::.} \rightarrow \text{[.]::[.]}$

$\text{NP::NP} \rightarrow [\text{D}_1\ \text{N}_2\ \text{A}_3]\text{::}[\text{DT}_1\ \text{JJ}_3\ \text{NN}_2]$



Tree-to-Tree Methods

- Input is a string to jointly parse and translate (SCFG) [Lavie et al., 2008]
- Input is a tree to translate via tree transduction (STSG) [Xia and McCord, 2004]
- Input is a forest of packed variant trees to translate via transduction [Liu et al., 2009]

Tree-to-Tree Relaxations

- Use SAMT labeling instead [Chiang, 2010]
- Insert "virtual" nodes into the tree to cover partial constituents [Ambati and Lavie, 2008; Hanneman et al., 2011]
- Allow rules to replace one subtree with multiple (STSSG) [Zheng et al., 2008]
- Or ignore the target-side tree entirely…

Part 3

Tree-to-String Methods

Tree-to-String Translation

- Extract chunks of syntactic structure; recursively transform tree into string
- Why model only the source tree?
- Do we need a new rule formalism?

Tree-to-String Translation

- Extract chunks of syntactic structure; recursively transform tree into string
- Why model only the source tree?
  - Take tree input at run time: don’t have to guess constituent bounds
  - Therefore very fast and no pruning!
- Do we need a new rule formalism?
  - Still STSG: just flip direction of GHKM-like rules
  - Could also use SCFG

Tree-to-String Methods

- Input is a tree to translate via transduction [Huang et al., 2006; Liu et al., 2006]
- Input is a tree where nodes can be inserted, permuted, or translated [Yamada and Knight, 2001]
- Input is a forest of packed variant trees to translate via transduction [Mi et al., 2008]
- Analysis: Neubig and Duh [2014]

Part 4

Grand Summary

You Should Now Know…

- Why (linguistic) syntax in MT is desirable
- Some of the decision points involved
- Two ways of deciding: SAMT and GHKM
- How to extract/apply SCFG and STSG rules
- About trade-offs involving labels and tree constraints
- That if you want to know more…

There are Lots of Good Papers

- Ambati and Lavie (2008), "Improving Syntax Driven Translation Models by Re-Structuring Divergent and Non-Isomorphic Parse Tree Structures." Proceedings of AMTA.
- Chiang (2010), "Learning to Translate with Source and Target Syntax." Proceedings of ACL.
- DeNeefe et al. (2007), "What Can Syntax-Based MT Learn from Phrase-Based MT?" Proceedings of EMNLP.
- Galley et al. (2004), "What's in a Translation Rule?" Proceedings of NAACL.
- Galley et al. (2006), "Scalable Inference and Training of Context-Rich Syntactic Translation Models." Proceedings of ACL.
- Hanneman et al. (2011), "A General-Purpose Rule Extractor for SCFG-Based Machine Translation." Proceedings of SSST-5.
- Huang et al. (2006), "Statistical Syntax-Directed Translation with Extended Domain of Locality." Proceedings of AMTA.
- Huck et al. (2014), "Preference Grammars and Soft Syntax Constraints for GHKM Syntax-Based Statistical Machine Translation." Proceedings of SSST-8.
- Lavie et al. (2008), "Syntax-Driven Learning of Sub-Sentential Translation Equivalents and Translation Rules from Parsed Parallel Corpora." Proceedings of SSST-2.
- Liu et al. (2006), "Tree-to-String Alignment Template for Statistical Machine Translation." Proceedings of ACL.

There are Lots of Good Papers

- Liu et al. (2009), "Improving Tree-to-Tree Translation with Packed Forests." Proceedings of ACL.
- Mi et al. (2008), "Forest-Based Translation." Proceedings of ACL.
- Neubig and Duh (2014), "On the Elements of an Accurate Tree-to-String Machine Translation System." Proceedings of ACL.
- Venugopal et al. (2009), "Preference Grammars: Softening Syntactic Constraints to Improve Statistical Machine Translation." Proceedings of NAACL.
- Xia and McCord (2004), "Improving a Statistical MT System with Automatically Learned Rewrite Patterns." Proceedings of COLING.
- Zhang et al. (2008), "A Tree Sequence Alignment-Based Tree-to-Tree Translation Model." Proceedings of ACL.
- Zollmann and Venugopal (2006), "Syntax Augmented Machine Translation via Chart Parsing." Proceedings of WMT.
- Zollmann (2011), "Learning Multiple-Nonterminal Synchronous Grammars for Statistical Machine Translation." Ph.D. thesis, Carnegie Mellon University.

Any questions?