Hieu Hoang

MT Marathon 2011 Trento

MIXED SYNTAX TRANSLATION

Contents

- What is a syntactic model?
- What's wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

What is a syntactic model?

- Hierarchical Phrase-Based Model
  - String-to-string
  - Non-terminals are unlabelled

    X → habe X_1 gegessen # have eaten X_1

- Tree-to-string Model
  - Source non-terminals are labelled
    - match input parse tree

    S → habe **NP_1** gegessen # have eaten NP_1

What is a syntactic model?

- Hierarchical Phrase-Based Model
  - String-to-string
  - Non-terminals are unlabelled

    X → habe X_1 gegessen # have eaten X_1

- Tree-to-string Model
  - Source non-terminals are labelled
    - match input parse tree

    S → habe **NP_1** gegessen # have eaten NP_1

What is a syntactic model?

- Hierarchical Phrase-Based Model
  - String-to-string
  - Non-terminals are unlabelled

    X → habe X_1 gegessen # have eaten X_1

- Tree-to-string Model
  - Source non-terminals are labelled
    - match input parse tree

    S → habe **NP_1** gegessen # have eaten NP_1

What is a syntactic model?

- Hierarchical Phrase-Based Model
  - String-to-string
  - Non-terminals are unlabelled

    X → habe X_1 gegessen # have eaten X_1

- Tree-to-string Model
  - Source non-terminals are labelled
    - match input parse tree

    S → habe **NP_1** gegessen # have eaten NP_1

What is a syntactic model?

- Hierarchical Phrase-Based Model
  - String-to-string
  - Non-terminals are unlabelled

    X → habe X_1 gegessen # have eaten X_1

- Tree-to-string Model
  - Source non-terminals are labelled
    - match input parse tree

    S → habe **NP_1** gegessen # have eaten NP_1

Contents

- What is a syntactic model?
- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

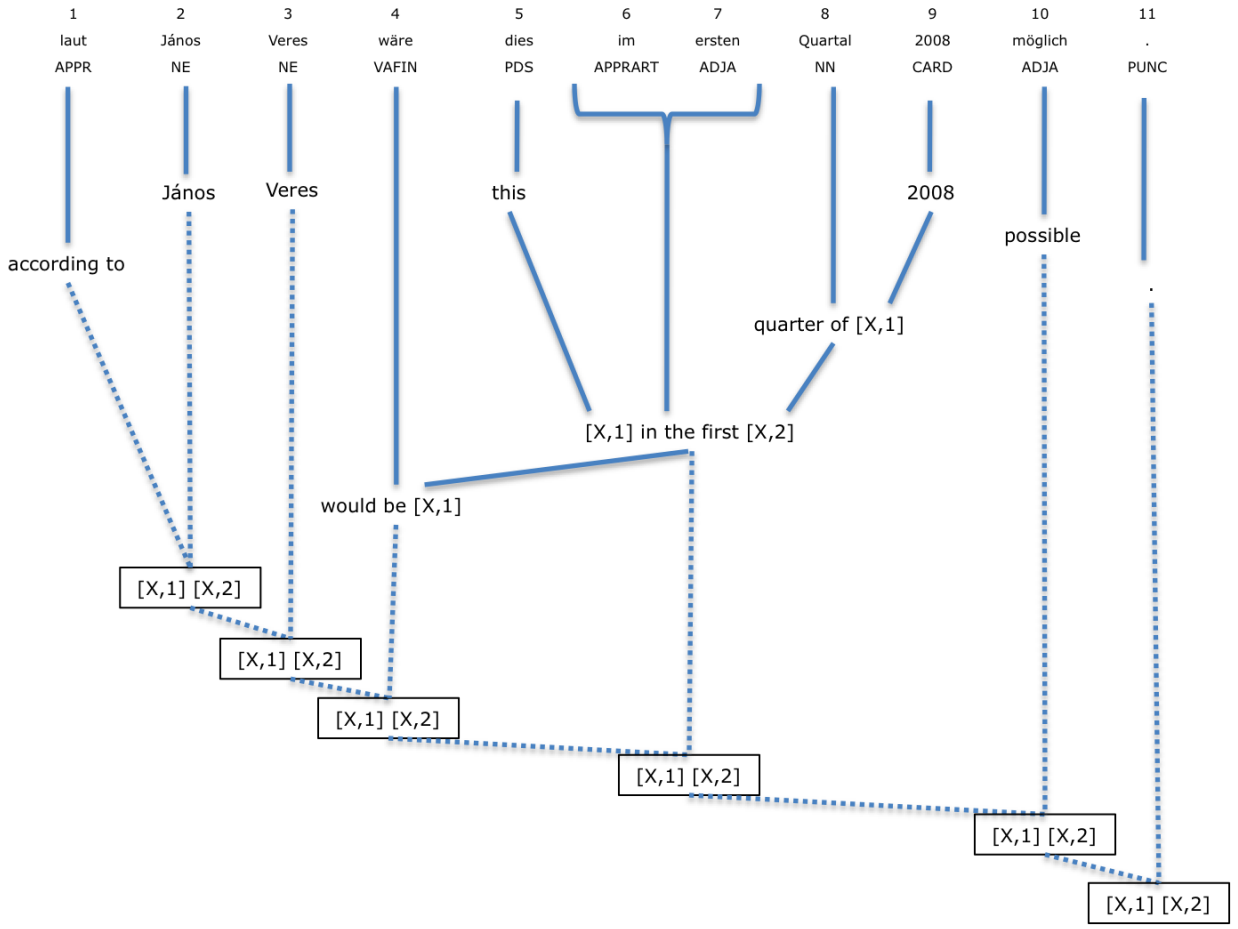
What's Wrong with Syntax?

	BLEU	METEOR
Tree-to-string	27.02	57.68
Tree-to-tree	22.23	54.05
Moses (phrase-based)	30.18	58.13

Evaluation of French-English MT System

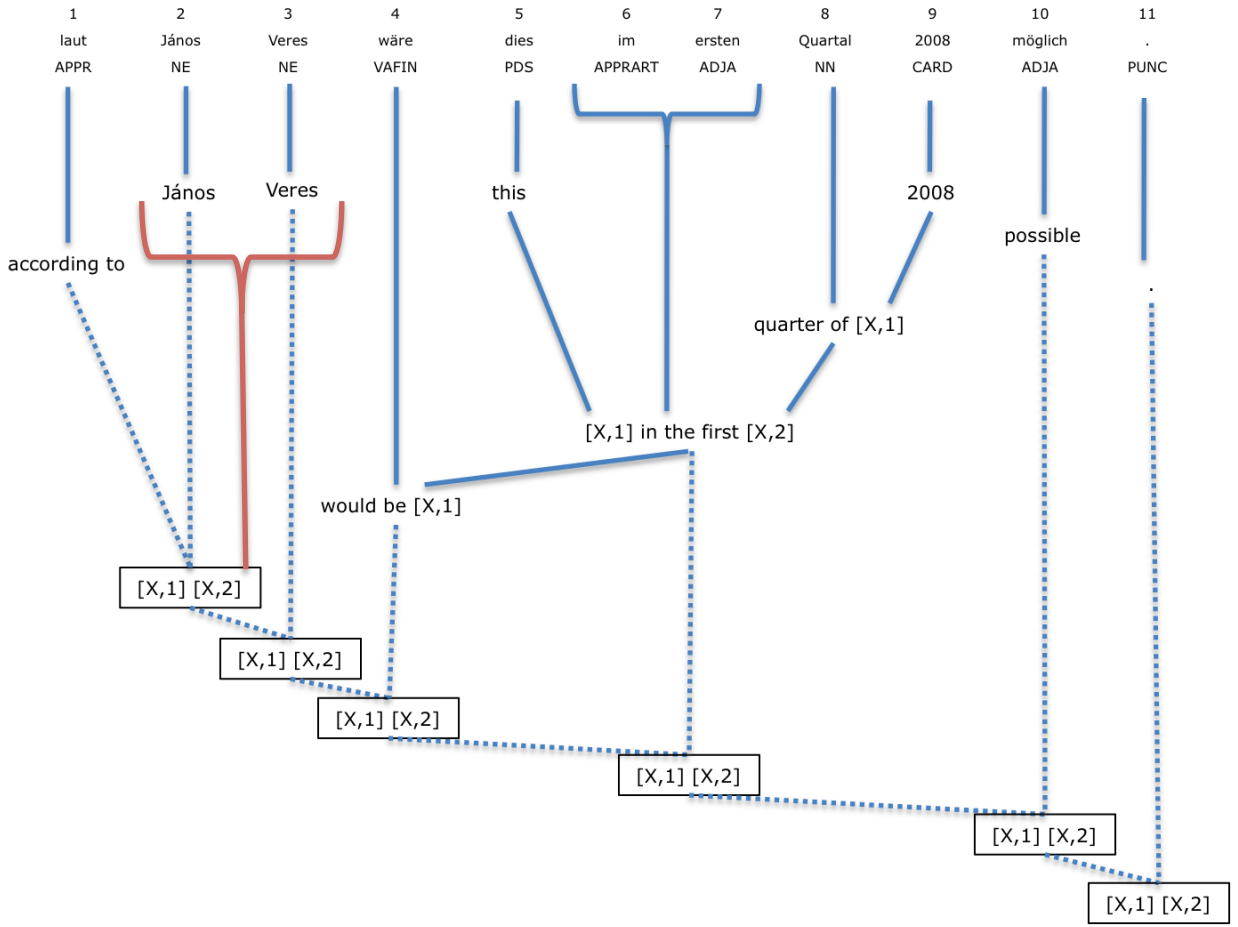
(Ambati and Lavie, 2009)

Hierarchical Model



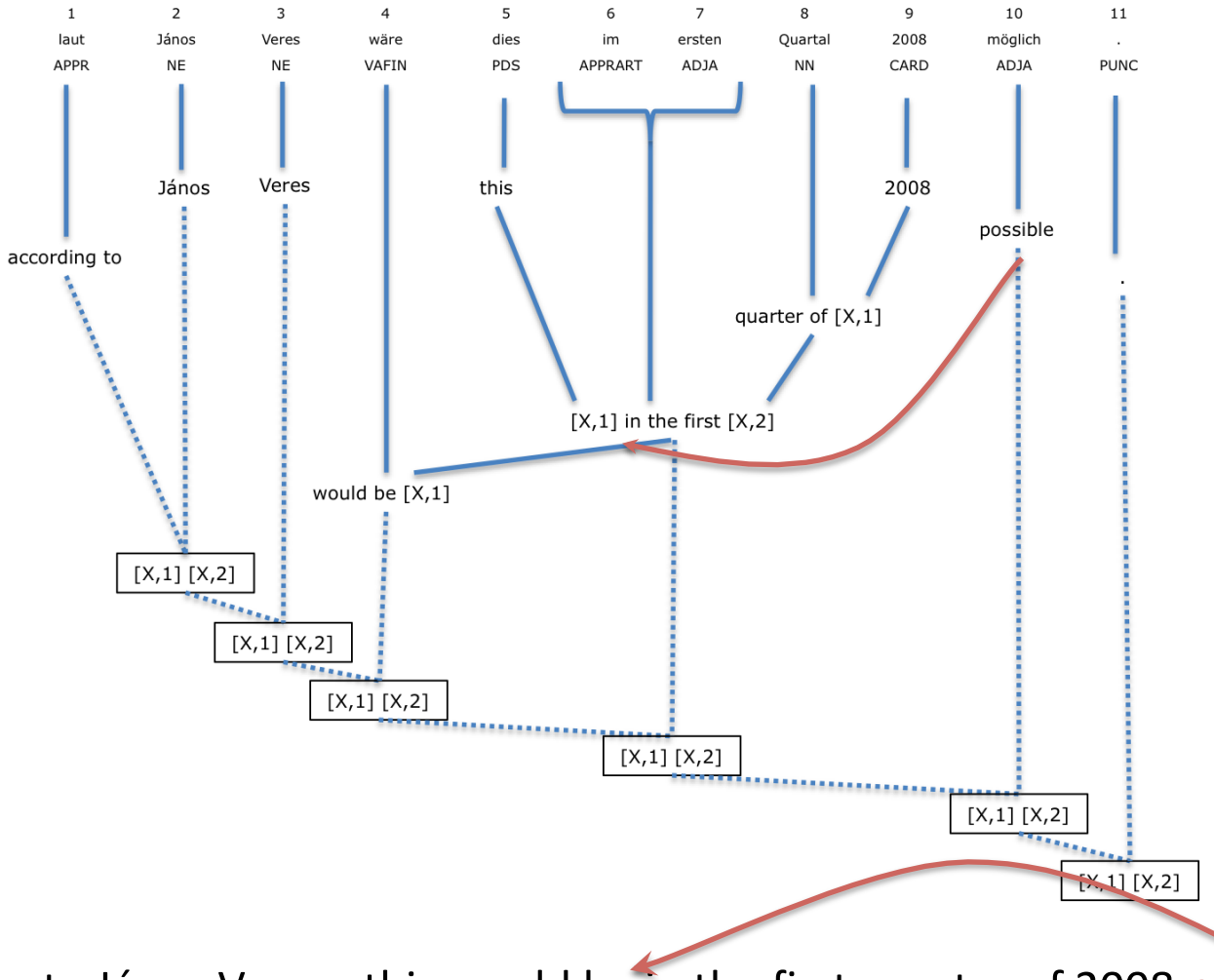
according to János Veres , this would be in the first quarter of 2008 **possible** .

Hierarchical Model



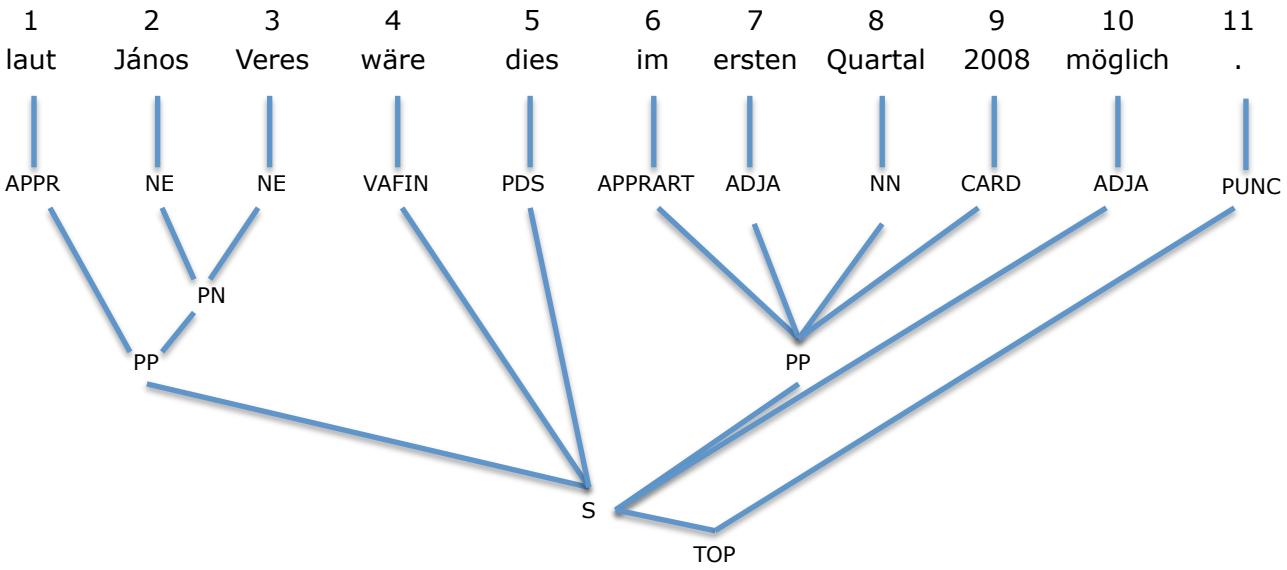
according to János Veres , this would be in the first quarter of 2008 **possible** .

Hierarchical Model

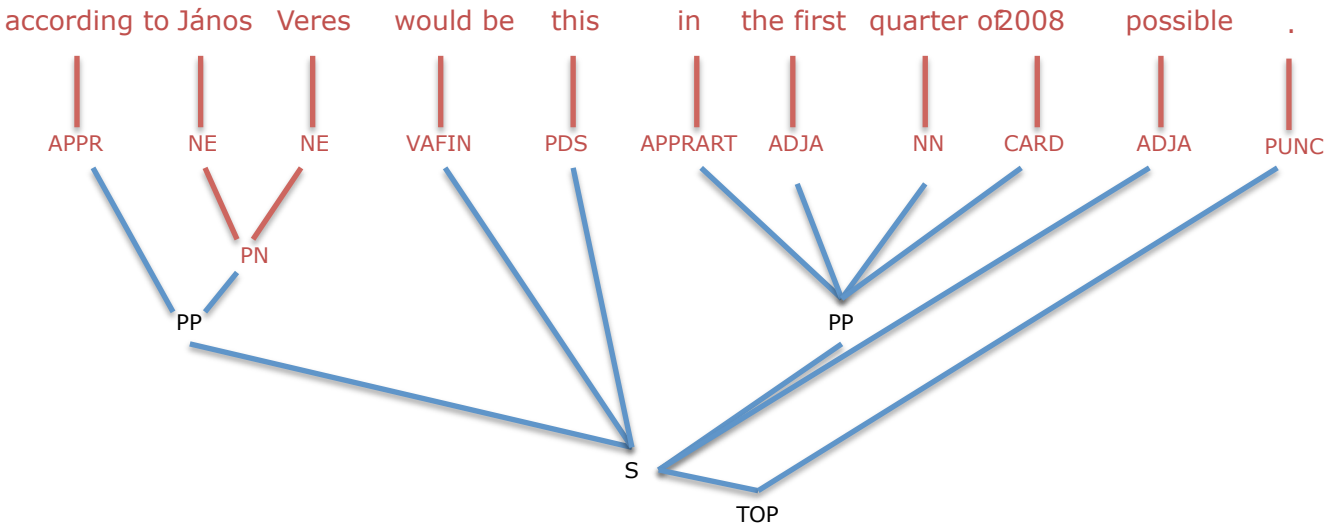


according to János Veres , this would be in the first quarter of 2008 **possible** .

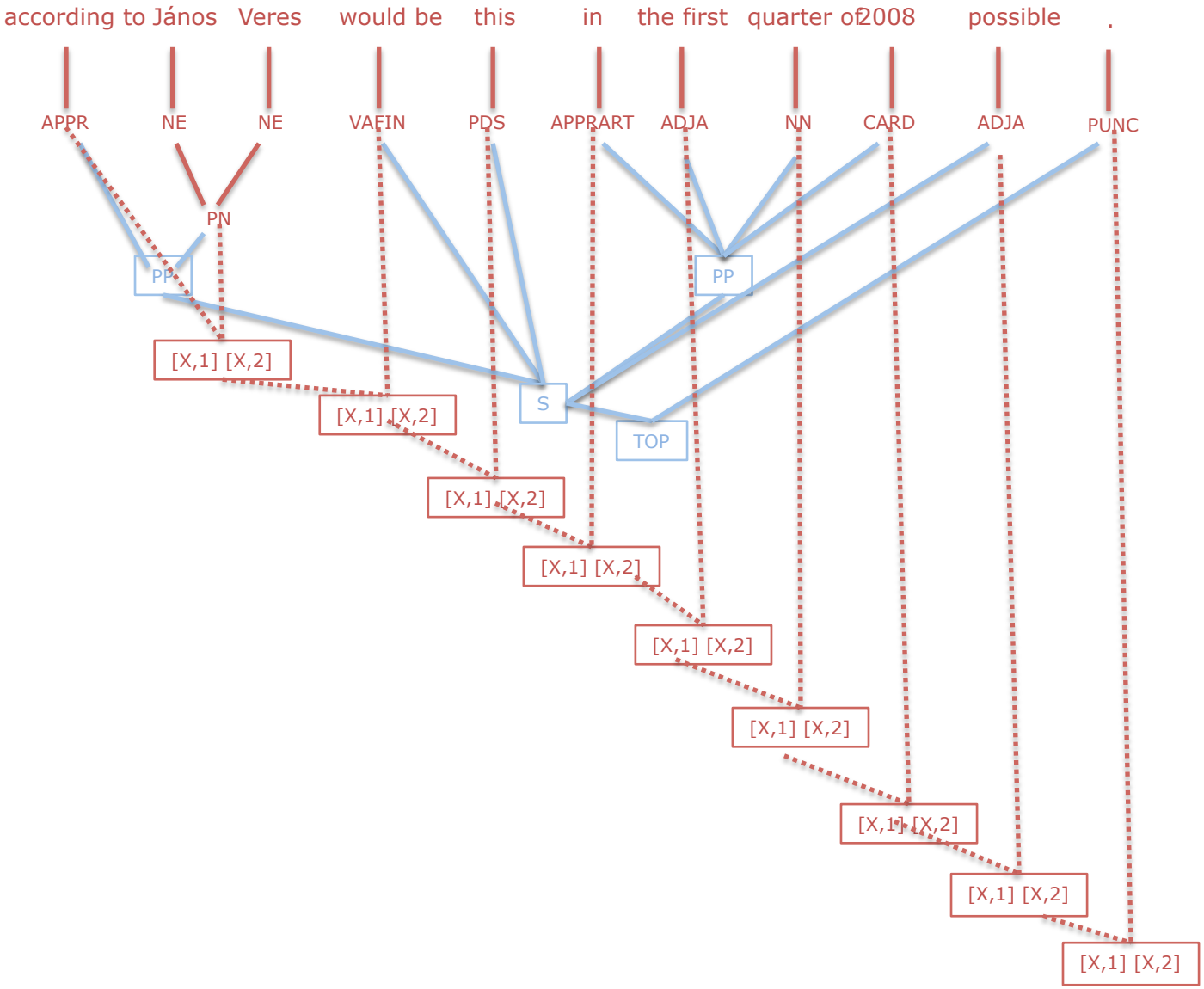
Tree-to-String Model



Tree-to-String Model



Tree-to-String Model



Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)

- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)

- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)

- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)

- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)

- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)

- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)

- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)

- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)

- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)
- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)
- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)
- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)
- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)
- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)
- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)

Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

Why Use Syntactic Models?

- Decrease decoding time
  - Derivation constrained by source parse tree

Why Use Syntactic Models?

- Decrease decoding time
  - Derivation constrained by source parse tree
- Long-range reordering during decoding
  - rules covering more words than max-span limit

Why Use Syntactic Models?

- Decrease decoding time
  - Derivation constrained by source parse tree
- Long-range reordering during decoding
  - rules covering more words than max-span limit
- Other rule-forms
  - 3+ non-terminals
  - consecutive non-terminals
  - non-lexicalized rules

Why Use Syntactic Models?

- Decrease decoding time
  - Derivation constrained by source parse tree
- Long-range reordering during decoding
  - rules covering more words than max-span limit
- Other rule-forms
  - 3+ non-terminals
  - consecutive non-terminals
  - non-lexicalized rules

$X \rightarrow S_1 O_2 V_3$ # $S_1 V_3 O_2$

$X \rightarrow PRO_1\ PRO_2$ aime bien # PRO_1 like PRO_2

Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)
- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)
- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)
- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)

Other Syntactic Models

- Syntax-Augmented MT (SAMT)
  - Not constrained only to parse tree
  - (Zollmann and Venugopal, 2006)
- Binarization
  - Restructure and relabel parse parse tree
  - (Wang et al, 2010)
- Forest-based translation
  - Recover from parse errors
  - (Mi et al, 2008)
- Soft constraint
  - Reward/Penalize derivations which follows parse structure
  - (Chiang 2010)
- Ignore Syntax (occasionally)

Mixed-Syntax Model

- Tree-to-string model
  - input is a parse tree
- Roles of non-terminals
  - Constrain derivation to parse constituents
  - State information
    - Consistent node label on target derivation
    - hypotheses with different head NT cannot be recombined

Mixed-Syntax Model

- Tree-to-string model
  - input is a parse tree
- Roles of non-terminals
  - Constrain derivation to parse constituents
    - Can sometime have no constraints
  - State information
    - Consistent node label on target derivation
    - hypotheses with different head NT cannot be recombined
    - always X

Mixed-Syntax Model

Example Translation Rules

- Naïve syntax model

  VP → VVFIN1 zu $VVINF_2$ # to $VVFIN_2$ $VVINF_1$

- Mixed-Syntax Model

  VP → X_1 zu $VVINF_2$ # X→ to X_2 X_1

Mixed-Syntax Model

Example Translation Rules

- Naïve syntax model

  VP → VVFIN1 zu $VVINF_2$ # to $VVFIN_2$ $VVINF_1$

- Mixed-Syntax Model

  VP → X_1 zu $VVINF_2$ # X→ to X_2 X_1

Mixed-Syntax Model

Example Translation Rules

- Naïve syntax model

  VP → VVFIN1 zu $VVINF_2$ # to $VVFIN_2$ $VVINF_1$

- Mixed-Syntax Model

  VP → X_1 zu $VVINF_2$ # X→ to X_2 X_1

Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

Extraction

- Allow rules
  - Max 3 non-terminals
  - Adjacent non-terminals
    - At least 1 NT must be syntactic
  - Non-lexicalized rules

Example Rules Extracted

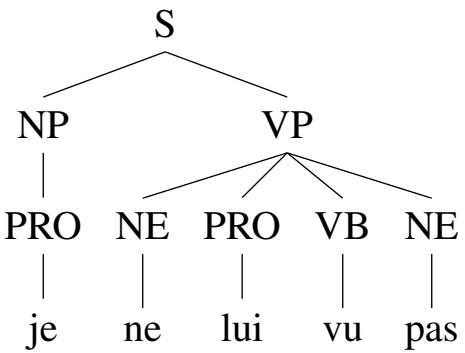
Rule	Factional Count	p(t | s)
Syntactic Rules		
VP → NP_1 $VVINF_2$ # X→X_2 X_1	167.3	68%
Mixed Rules		
VP → X_1 VZ_2 # X→ X_2 X_1	63.3	64%
VP → X_1 zu $VVINF_2$ # X→ to X_2 X_1	39.9	56%
TOP → NP_1 X_2 # X→ X_1 X_2	43.1	92%

Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

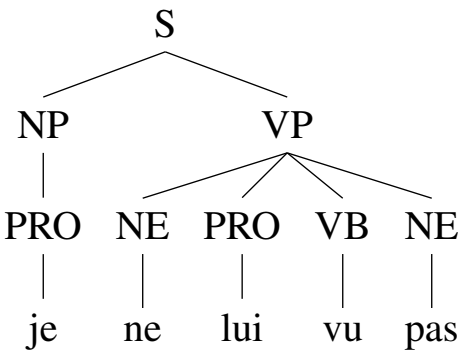
Synchronous CFG

Input:



Synchronous CFG

Input:



Rules:

S →NP_1 VP_2 # NP_1 VP_2

NP →je # I

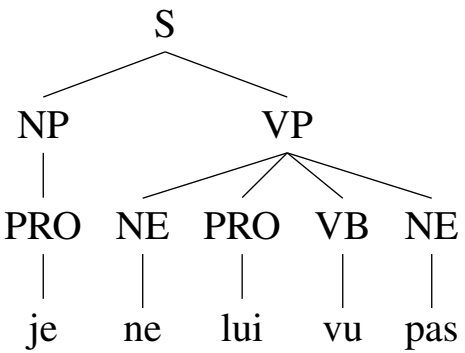
PRO →lui # him

VB →vu # see

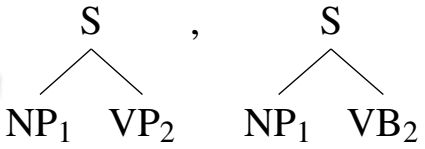
VP →ne PRO_1 VB_2 pas # did not VB_2 PRO_1

Synchronous CFG

Input:



Derivation:



Rules:

S →NP_1 VP_2 # NP_1 VP_2

NP →je # I

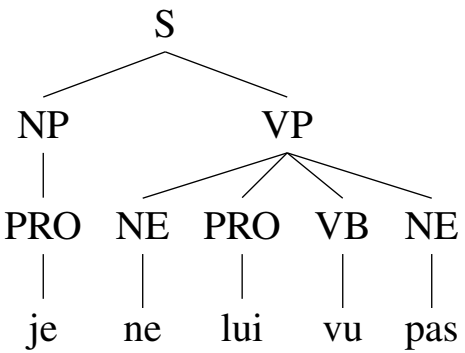
PRO →lui # him

VB →vu # see

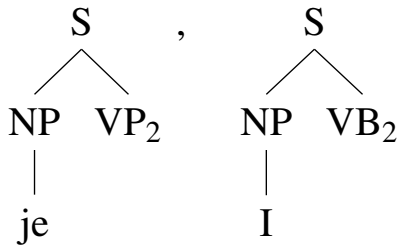
VP →ne PRO_1 VB_2 pas # did not VB_2 PRO_1

Synchronous CFG

Input:



Derivation:



Rules:

S →NP_1 VP_2 # NP_1 VP_2

NP →je # I

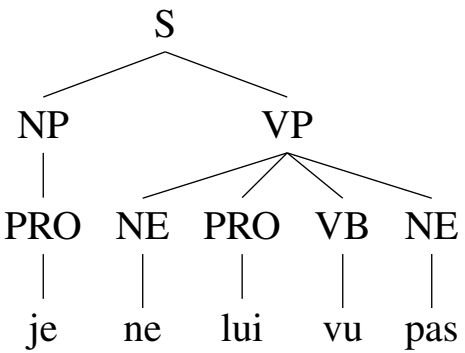
PRO →lui # him

VB →vu # see

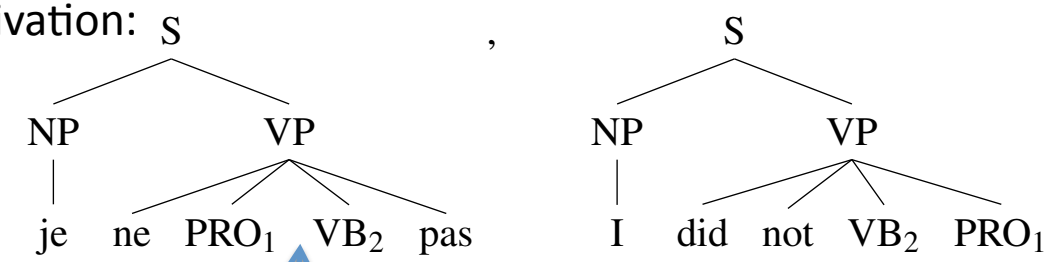
VP →ne PRO_1 VB_2 pas # did not VB_2 PRO_1

Synchronous CFG

Input:



Derivation:



Rules:

S →NP_1 VP_2 # NP_1 VP_2

NP →je # I

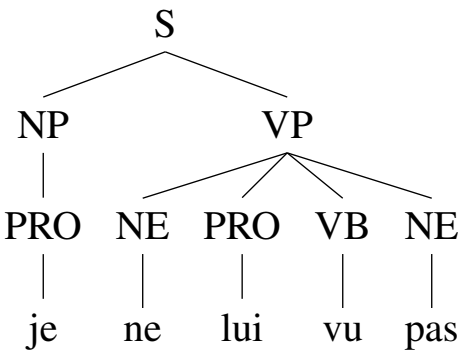
PRO →lui # him

VB →vu # see

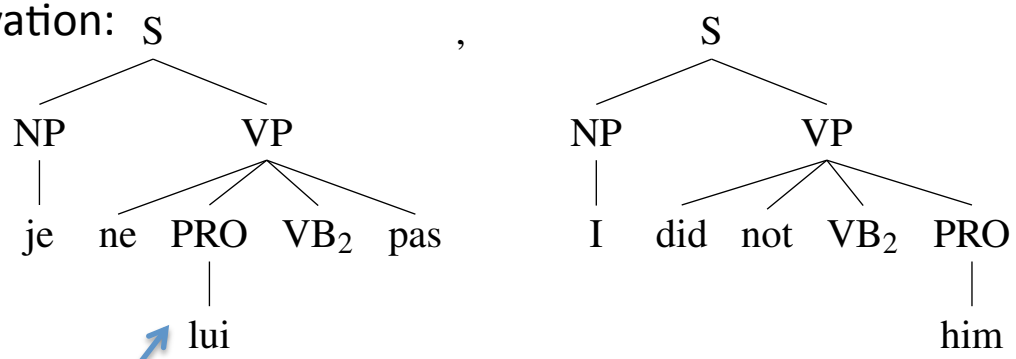
VP →ne PRO_1 VB_2 pas # did not VB_2 PRO_1

Synchronous CFG

Input:



Derivation:



Rules:

S →NP_1 VP_2 # NP_1 VP_2

NP →je # I

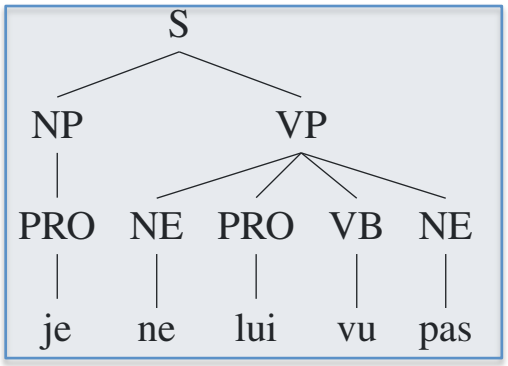
PRO →lui # him

VB →vu # see

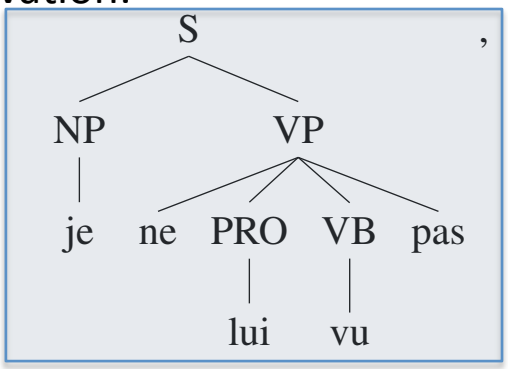
VP →ne PRO_1 VB_2 pas # did not VB_2 PRO_1

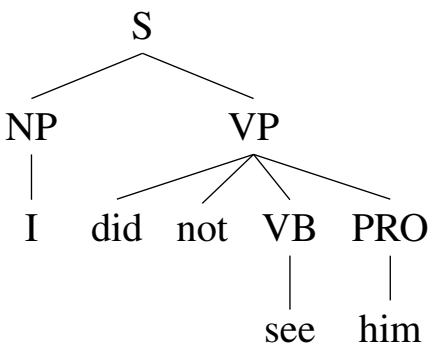
Synchronous CFG

Input:



Derivation:





Rules:

S → NP_1 VP_2 # NP_1 VP_2

NP → je # I

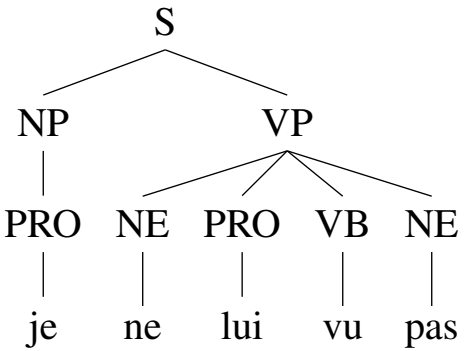
PRO → lui # him

VB → vu # see

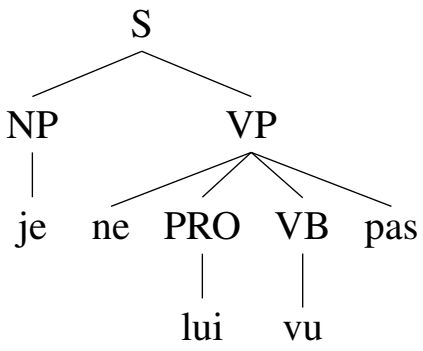
VP → ne PRO_1 VB_2 pas # did not VB_2 PRO_1

Synchronous CFG

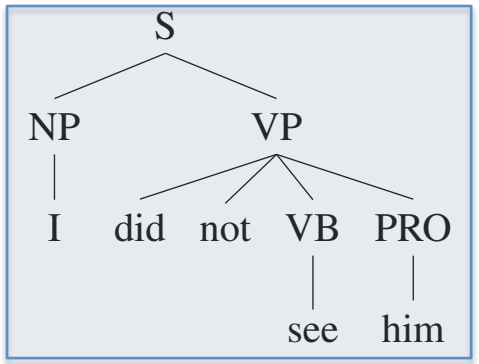
Input:



Derivation:



,



Rules:

S →NP_1 VP_2 # NP_1 VP_2

NP →je # I

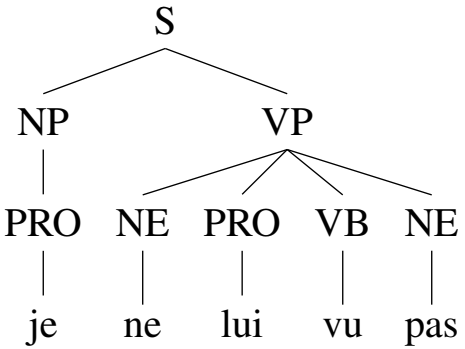
PRO →lui # him

VB →vu # see

VP →ne PRO_1 VB_2 pas # did not VB_2 PRO_1

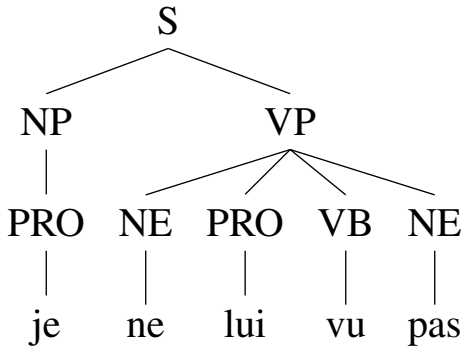
Mixed-Syntax Model

Input:



Mixed-Syntax Model

Input:

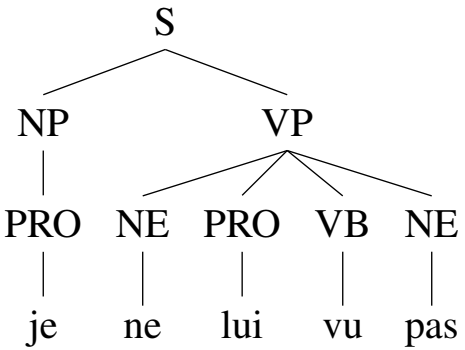


Rules:

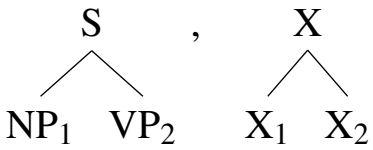
S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X→ him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

Mixed-Syntax Model

Input:



Derivation:

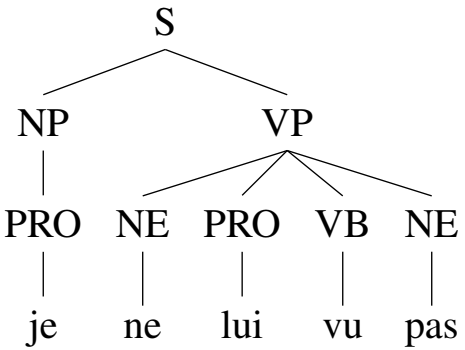


Rules:

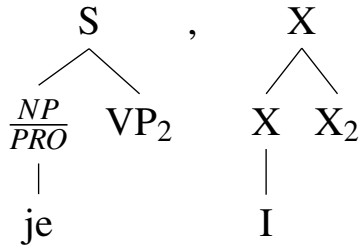
S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X→ him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

Mixed-Syntax Model

Input:



Derivation:

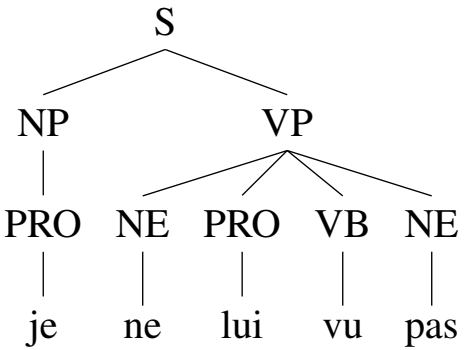


Rules:

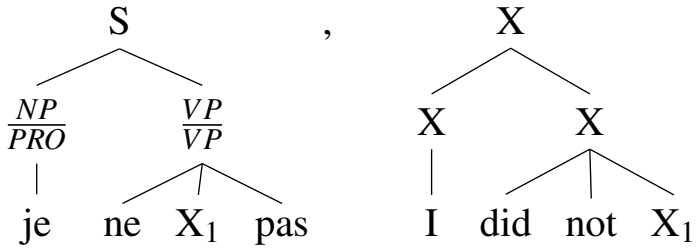
S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X→ him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

Mixed-Syntax Model

Input:



Derivation:

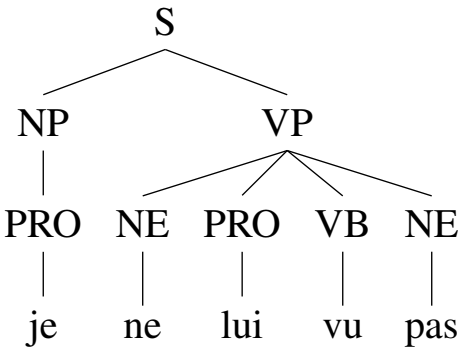


Rules:

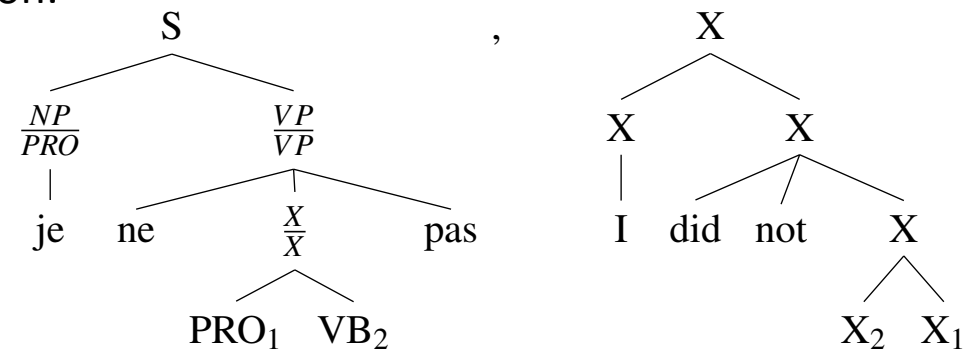
S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X→ him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

Mixed-Syntax Model

Input:



Derivation:

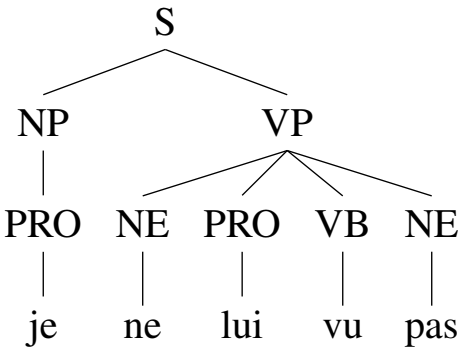


Rules:

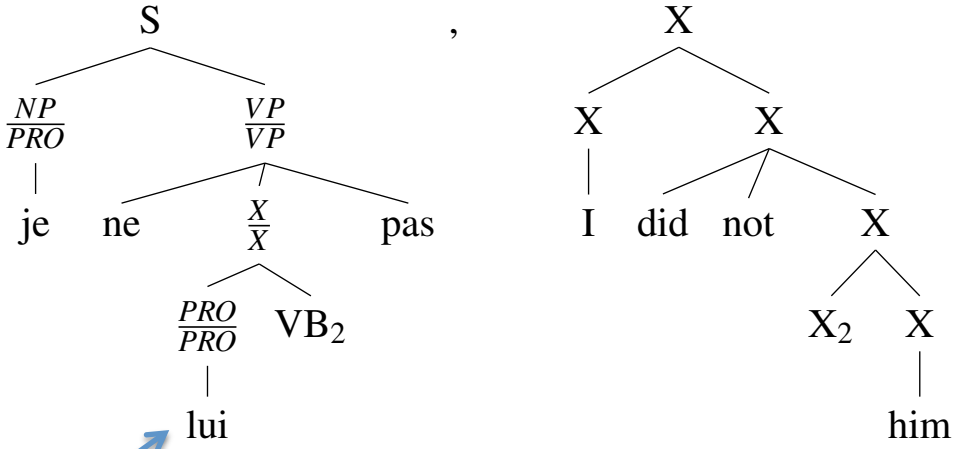
S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X→ him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

Mixed-Syntax Model

Input:



Derivation:

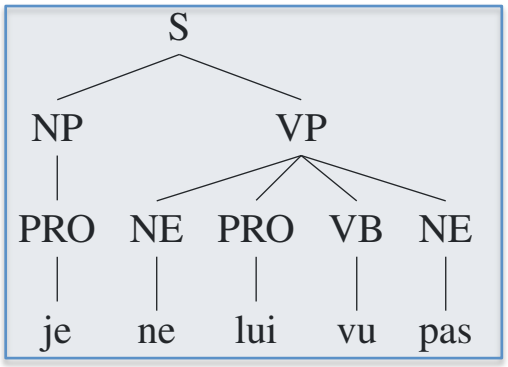


Rules:

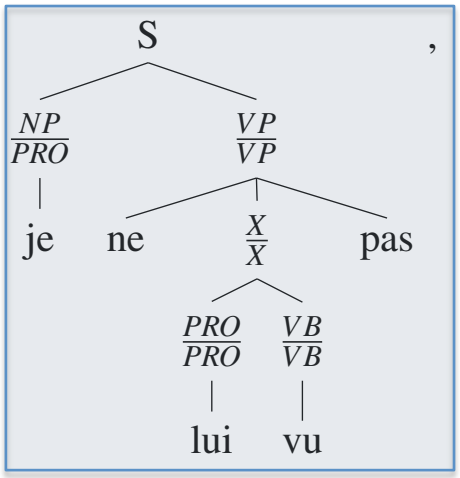
S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X → him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

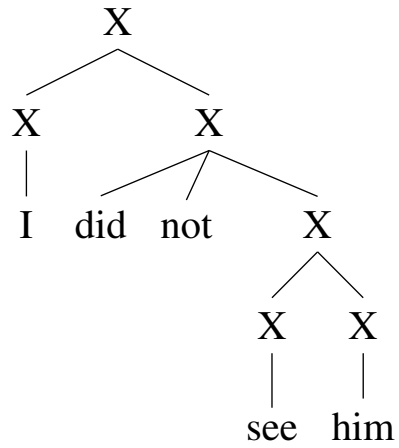
Mixed-Syntax Model

Input:



Derivation:





Rules:

S	→ NP_1 VP_2	# X → NP_1 VP_2
PRO	→ je	# X → I
PRO	→ lui	# X→ him
VB	→ vu	# X → see
VP	→ ne X_1 pas	# did not X_1
X	→ PRO_1 VB_2	# X → X_2 X_1

Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

Experiment

German-English

Corpus

		German	English
Train	Sentences	82,306	
	Words	2,034,373	1,965,325
Tune	Sentences	2000	
Test	Sentences	1026	

Trained: News Commentary 2009
Tuned: held out set
Tested: *nc test2007*

Results

Using constituent parse

German-English

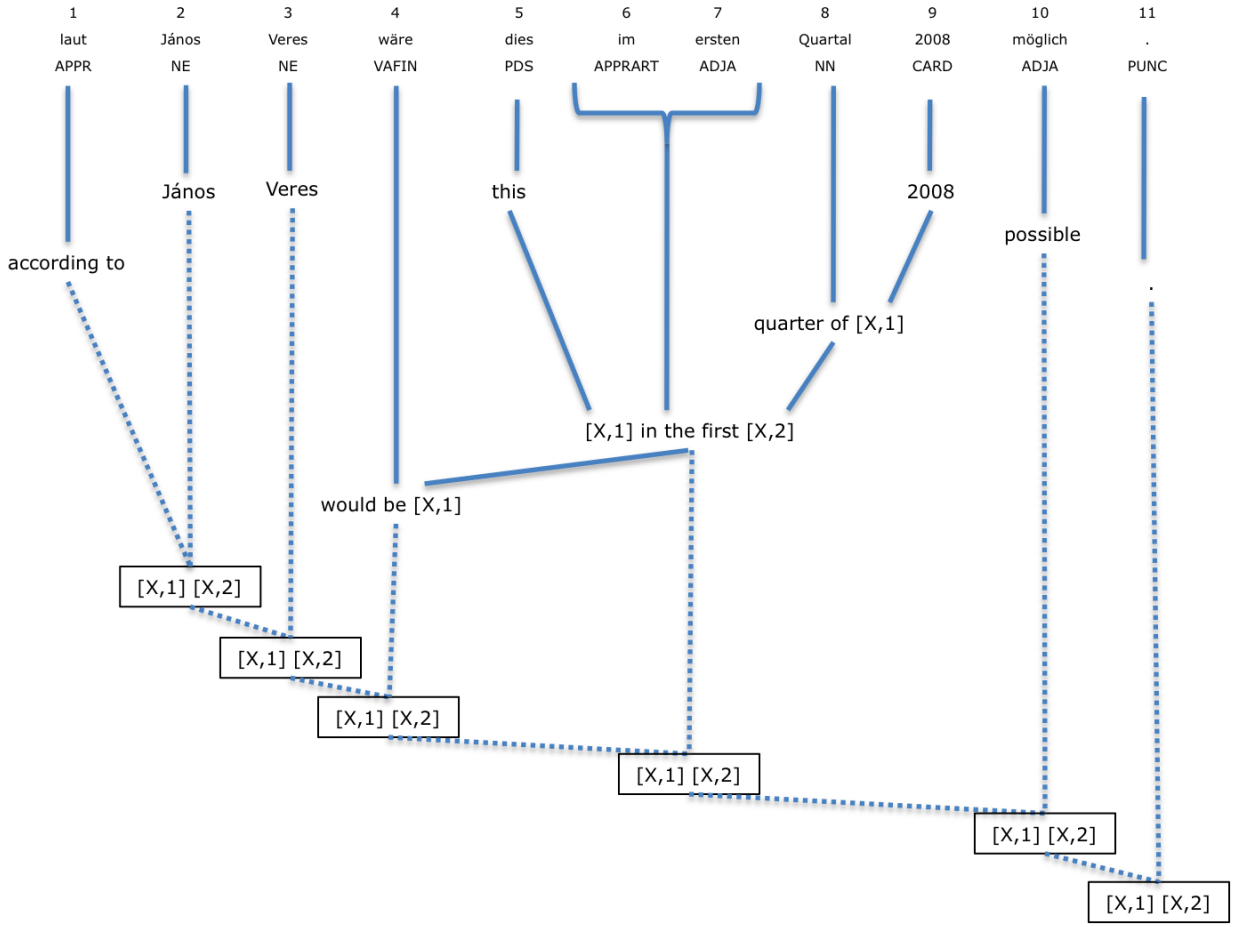
Model	# rules	%BLEU
Hierarchical	61.2m	15.9
Tree-to-String	4.7m	14.9
Mixed Syntax	128.7m	16.7

English-German

Model	# rules	%BLEU
Hierarchical	84.6m	10.2
Mixed Syntax	175.0m	10.6

Example

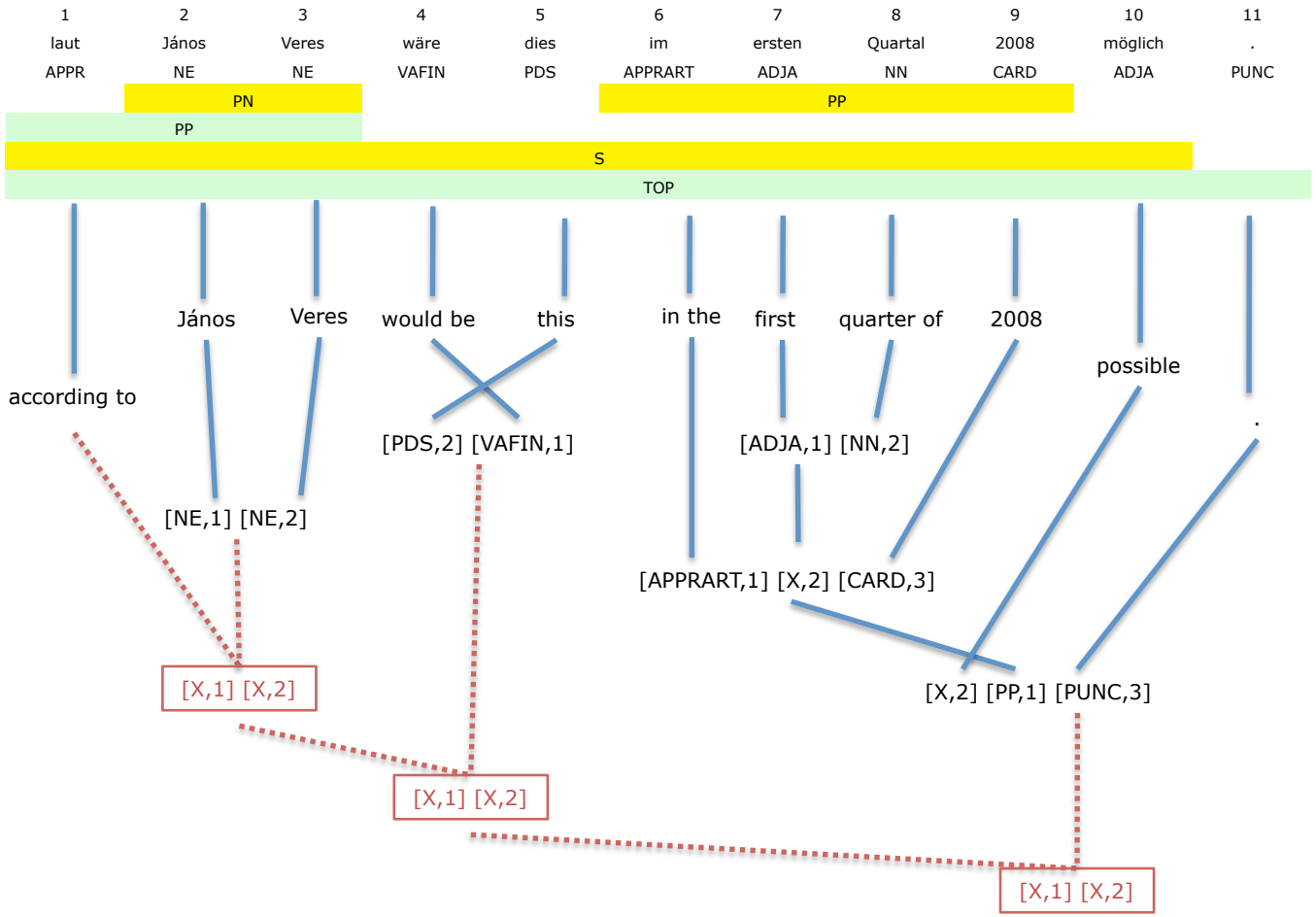
Hierarchical Model



according to János Veres , this would be in the first quarter of 2008 **possible** .

Example

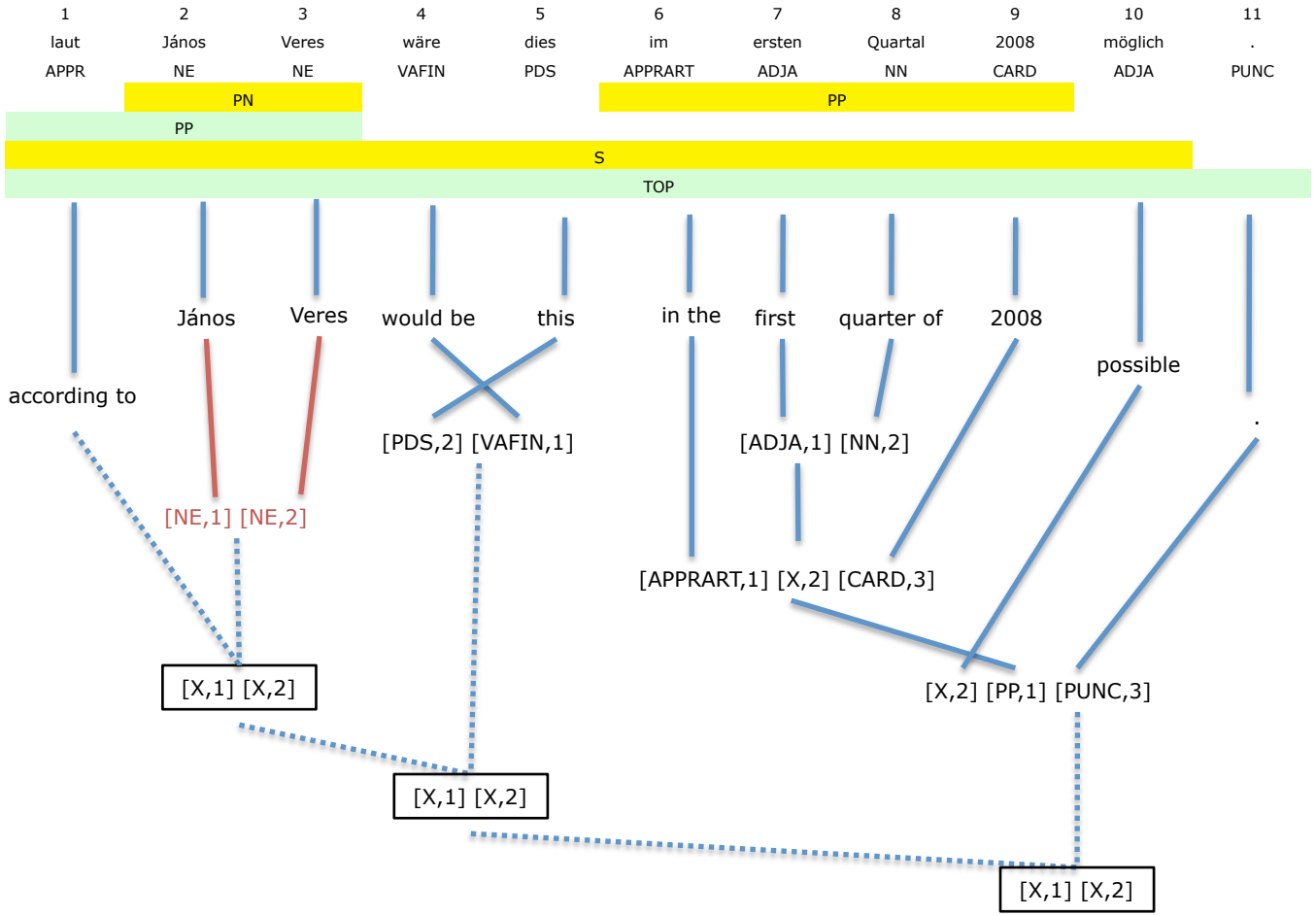
Mixed-Syntax Model



according to János Veres this would be **possible** in the first quarter of 2008.

Example

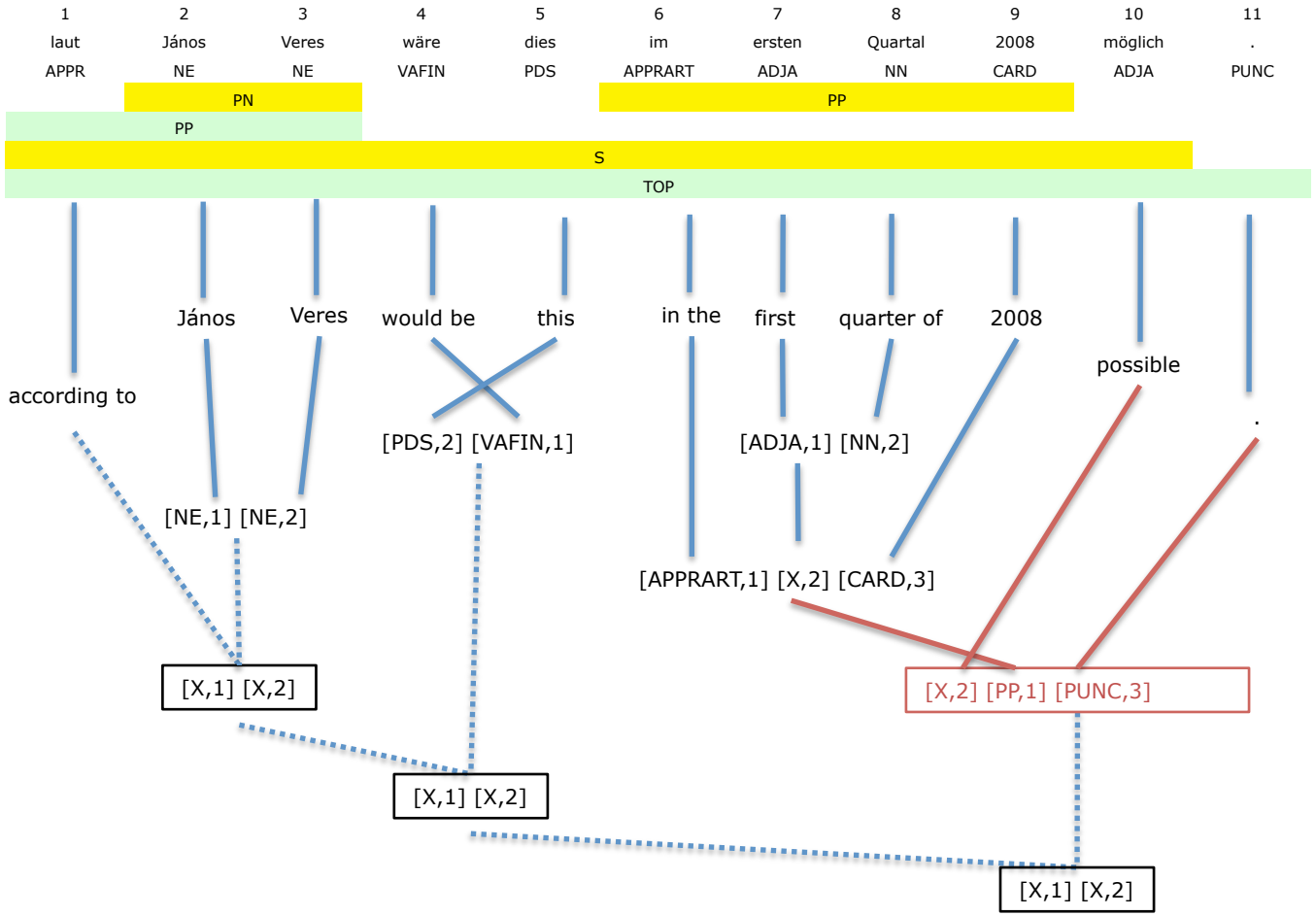
Mixed Syntax



according to János Veres this would be **possible** in the first quarter of 2008.

Example

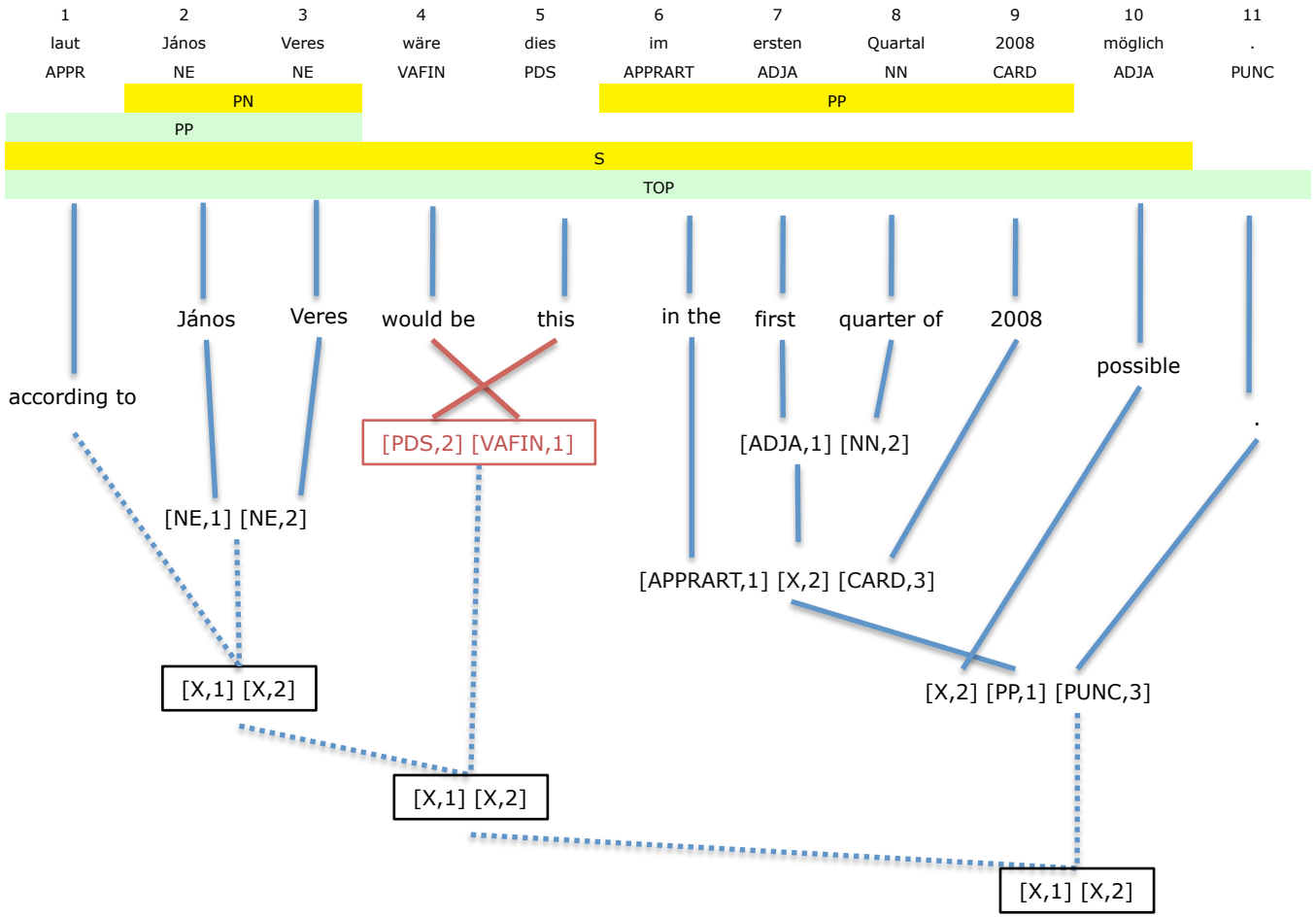
Mixed Syntax



according to János Veres this would be **possible** in the first quarter of 2008.

Example

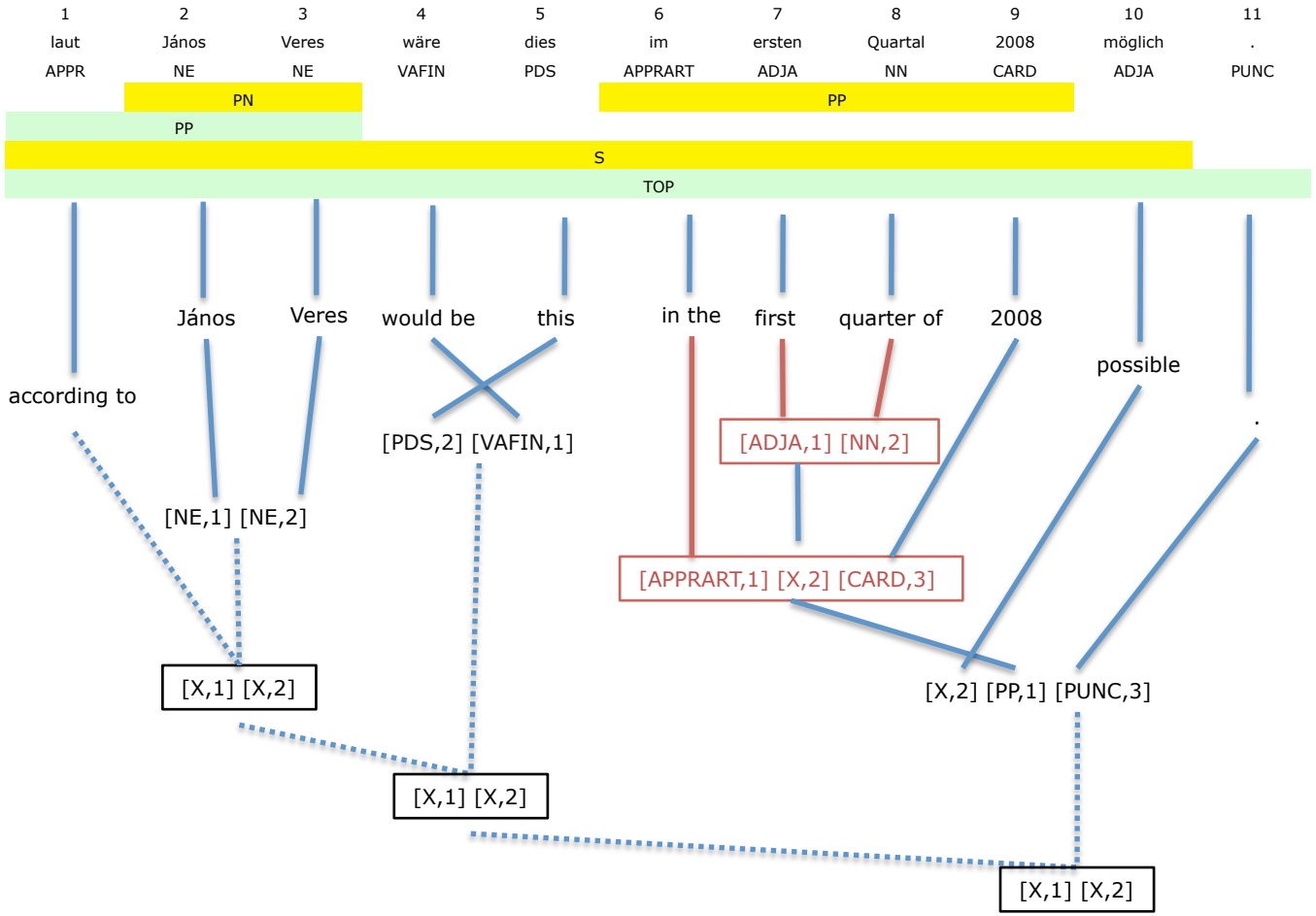
Mixed Syntax



according to János Veres this would be **possible** in the first quarter of 2008.

Example

Mixed Syntax



according to János Veres this would be **possible** in the first quarter of 2008.

Chunk Tags

- Advantages of Shallow Tags
  - Don’t need Treebank
  - More reliable
- Disadvantages
  - Not a tree structure
    - We don’t rely on tree structure

Results

Shallow Tags

German-English

Model	# rules	%BLEU
Hierarchical	64.3m	16.3
Mixed Syntax	254.5m	16.8

Larger Training Corpus

German-English

Corpus

		German	English	Corpus
Train	Sentences	1,446,224		Europarl v5
	Words	37,420,876	39,464,626	
Tune	Sentences	1910		dev2006
Test (in-domain) (out-of-domain)	Sentences	 1920 1042		 nc test2007 v2 devtest2006

Larger Training Corpus

German-English

Model	# rules	In-domain (BLEU)	Out-of-domain (BLEU)
Hierarchical	500m	22.1	16.5
Mixed Syntax (original)	2664m	21.6	16.3
Mixed Syntax (new extraction)	1435m	22.7	17.8

Contents

- What's Wrong with Syntax?
- Which syntax model to use?
- Why use syntactic models?
- Mixed-Syntax Model
  - Extraction
  - Decoding
  - Results
- Future Work

Create your own label

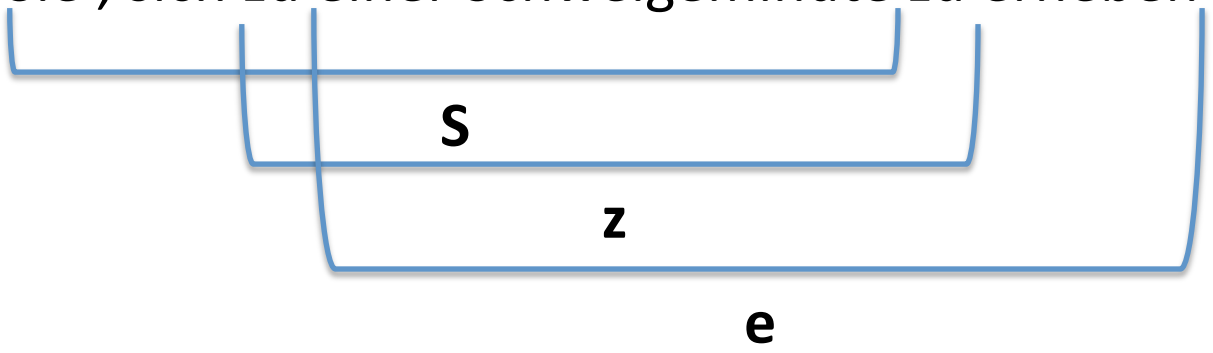
Dumb labels

ich bitte Sie , sich zu einer Schweigeminute zu erheben .

Create your own label

Dumb labels

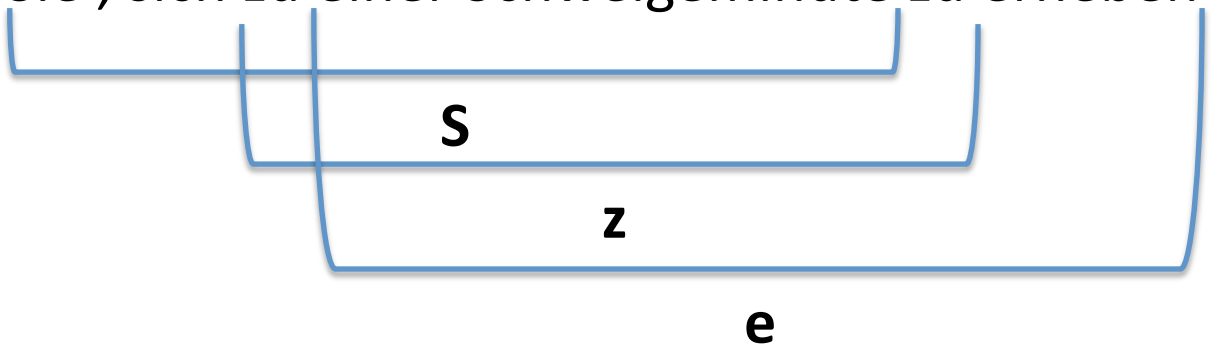
ich bitte Sie , sich zu einer Schweigeminute zu erheben .



Create your own labels

Dumb labels

ich bitte Sie , sich zu einer Schweigeminute zu erheben .



Model	In-domain (BLEU)	Out-of-domain (BLEU)
Hierarchical	22.1	16.5
Dumb Labels	22.0	16.3

Create your own labels

Labels motivated by reordering

Labelling patterns:

1. VMFIN...VVINF EOS
2. VVINF und ... VVINF
3. VAFIN ... (VVPP or VVINF) EOS
4. , PRELS ... VVINF EOS
5. EOS ... zu VVINF

Example:

ich bitte Sie , sich zu einer Schweigeminute zu erheben .

label 5

… werde ich dem Vorschlag von Herrn Evans folgen .

label 3

Create your own labels

Labels motivated by reordering

Model	In-domain (BLEU)	Out-of-domain (BLEU)
Hierarchical	22.1	16.5
Dumb Labels	22.0	16.3
Reordering Labels	22.1	16.9

Conclusion

- Mixed-Syntax Model
  - SCFG-based decoding
  - Hierarchical phrase-based v. tree-to-string
  - Generality v. specificity
- Syntax Models
  - Many variations
  - Won't automatically make MT better
  - Question
    - which syntactic information?
    - how do we use it?
    - why use it?

END