CCG Contextual Labels in Hierarchical Phrase-Based SMT

Hala Almaghout, Jie Jiang, Andy Way

CNGL, School of Computing, Dublin City University
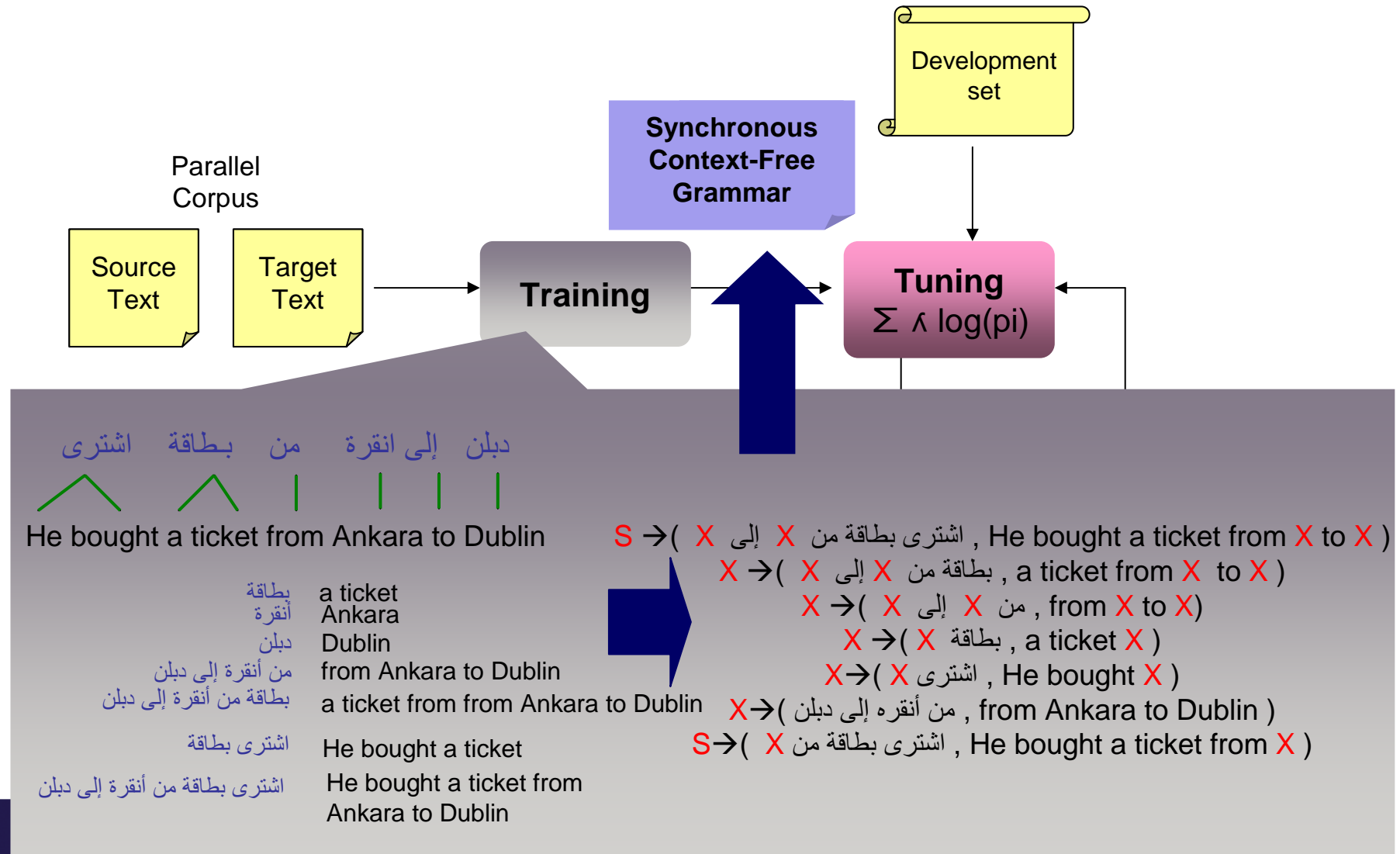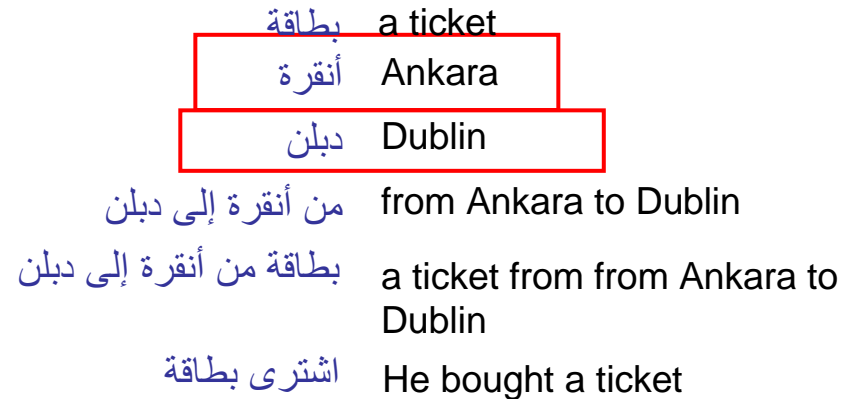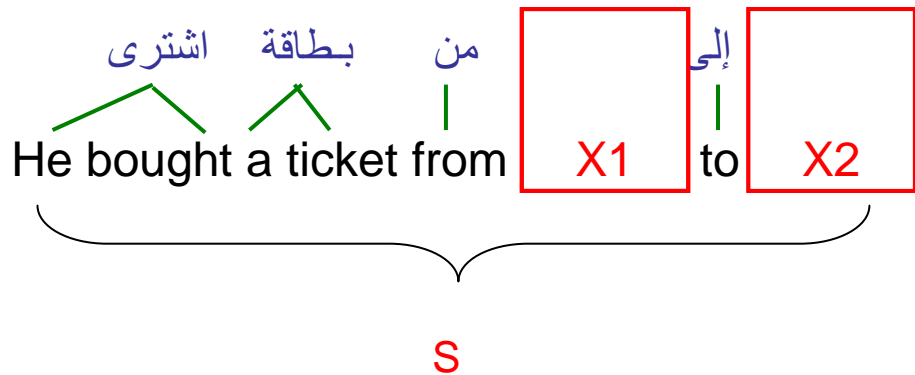
Outline

- **Introduction**
- Related Work
- Our Approach
- Experiments
- Analysis
- Conclusion
- Future Work

Hierarchical Phrase-Based SMT (Chiang, 2005)

Development set

Synchronous Context-Free Grammar

Parallel Corpus

Source Text

Target Text

Training

Tuning
$\sum \lambda \log(p_i)$

اشترى بطاقة من أنقرة إلى انقرة دبلن

He bought a ticket from Ankara to Dublin

بطاقة	a ticket
أنقرة	Ankara
دبلن	Dublin
من أنقرة إلى دبلن	from Ankara to Dublin
بطاقة من أنقرة إلى دبلن	a ticket from from Ankara to Dublin
اشترى بطاقة	He bought a ticket
اشترى بطاقة من أنقرة إلى دبلن	He bought a ticket from Ankara to Dublin

S → (اشترى بطاقة من X إلى X , He bought a ticket from X to X)

X → (بطاقة من X إلى X , a ticket from X to X)

X → (من X إلى X , from X to X)

X → (بطاقة X , a ticket X)

X → (اشترى X , He bought X)

X → (من أنقره إلى دبلن , from Ankara to Dublin)

S → (اشترى بطاقة من X , He bought a ticket from X)

Hierarchical Rule Extraction

He bought a ticket from [X1] to [X2]

اشترى بطاقة من إلى

S

بطاقة	a ticket
أنقرة	Ankara
دبلن	Dublin
من أنقرة إلى دبلن	from Ankara to Dublin
بطاقة من أنقرة إلى دبلن	a ticket from from Ankara to Dublin
اشترى بطاقة	He bought a ticket
اشترى بطاقة من أنقرة إلى دبلن	He bought a ticket from Ankara to Dublin

source side nonterminals

S → (X2 إلى X1 اشترى بطاقة من , He bought a ticket from X1 to X2)

source side
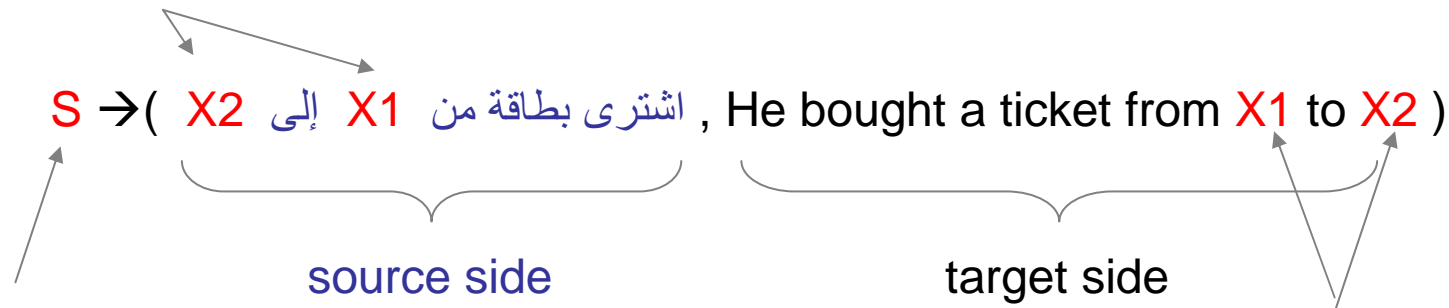
target side

Left hand side nonterminal

target side nonterminals

Limitations of HPB SMT

- There are *no syntactic constraints* imposed on phrases replacing nonterminals in hierarchical rules.

S= اشترى بطاقة

S→ (X اشترى , He bought X)

He bought ticket

He bought a ticket

بطاقة   ticket   N
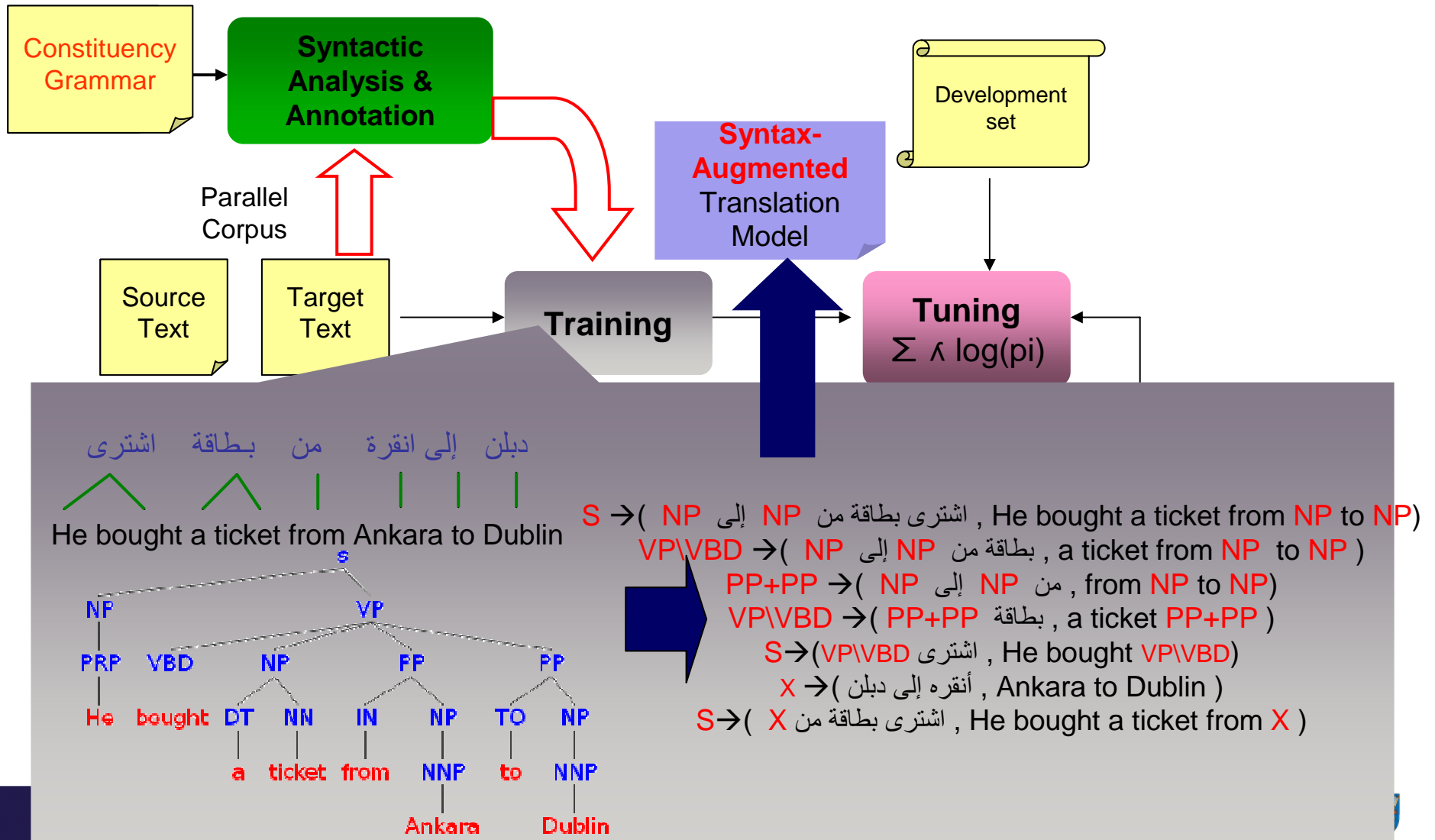
بطاقة   a ticket   NP
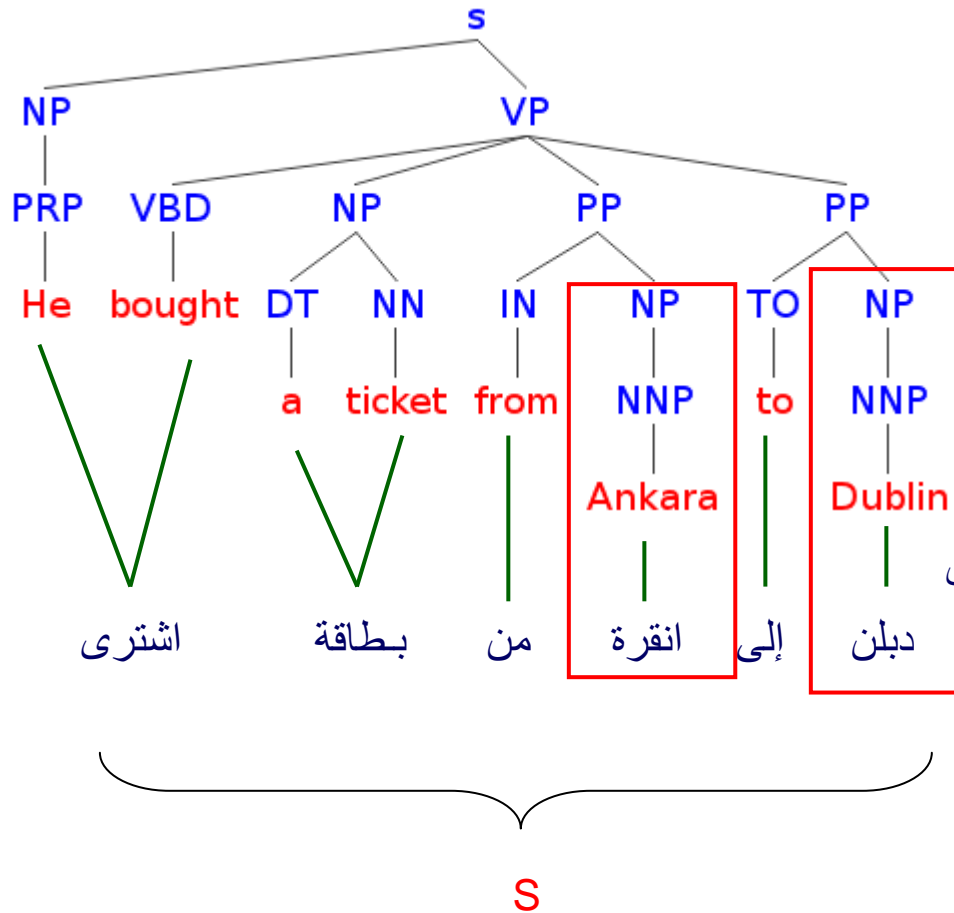
Solution:

S→ (X اشترى , He bought NP)

He bought

Syntax Augmented Machine Translation (SAMT) (Zollmann and Venugopal, 2006)

Syntax Augmented Machine Translation (SAMT)

SAMT Rule Extraction

S → (اشترى بطاقة من NP1 إلى NP2 , He bought a ticket from NP1 to NP2)

SAMT Rule Extraction

VP\VBD

S

PP+PP

X

NP — VP

PRP   VBD   NP   PP   PP

He   bought   DT   NN   IN   NP   TO   NP

a   ticket   from NNP   to   NNP

Ankara   Dublin

اشترى   بطاقة   من   انقرة   إلى   دبلن

بطاقة	a ticket	NP
أنقرة	Ankara	NP
دبلن	Dublin	NP
من أنقرة إلى دبلن	from Ankara to Dublin	PP+PP
بطاقة من أنقرة إلى دبلن	a ticket from from Ankara to Dublin	VP\VBD
اشترى بطاقة	He bought a ticket	X
اشترى بطاقة من أنقرة إلى دبلن	He bought a ticket from Ankara to Dublin	S
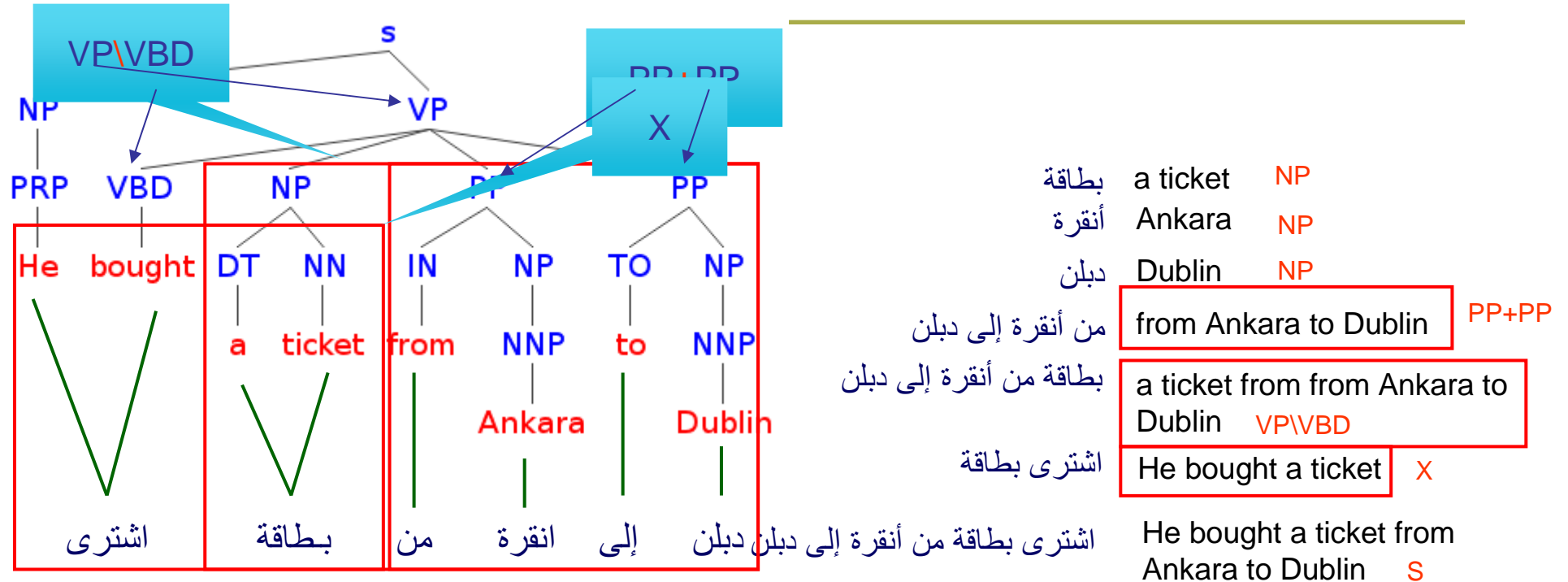
S →(NP إلى NP من اشترى بطاقة , He bought a ticket from NP to NP)

VP\VBD →(NP إلى NP من بطاقة , a ticket from NP to NP)

PP+PP →(NP إلى NP من , from NP to NP)

VP\VBD →(PP+PP بطاقة , a ticket PP+PP)

S→(VP\VBD اشترى , He bought VP\VBD)

X →(أنقره إلى دبلن , Ankara to Dublin)

S→(X اشترى بطاقة من , He bought a ticket from X)

SAMT Limitations

- **Label sparsity**: using many different nonterminal labels. This results from using SAMT combinatory operators.

S → (X بطاقة اشترى , He bought a ticket X)

S → (PP بطاقة اشترى , He bought a ticket

S → (PP+PP بطاقة اشترى , He bou

S → (SBAR بطاقة اشترى , He bought a ticket SBAR)   0.015

> • SAMT has larger translation model.
> • Low-probability rules weaken the system's ability to generalize, and damage the performance of the system.

- **Label coverage**: failing to find a syntactic label expressing the syntactic function of some of SMT phrases.
- **Label accuracy**: how accurate the SAMT labels are in reflecting the real syntactic function of the phrases.

SAMT

Producing more grammatical translation by imposing syntactic constraints on nonterminal replacement
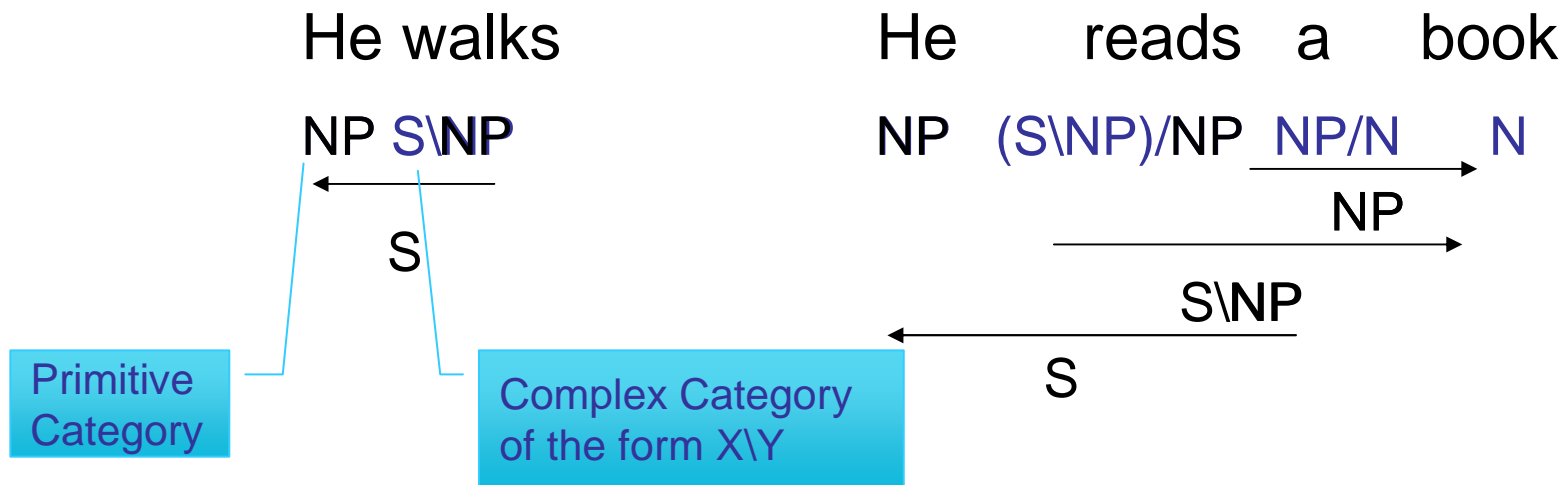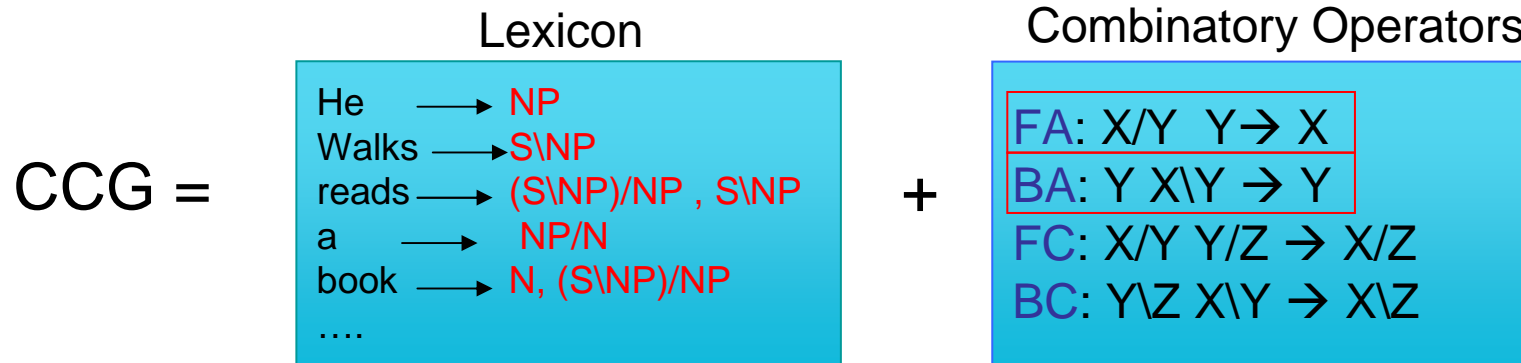
Constituency grammar rigid structures →Label coverage
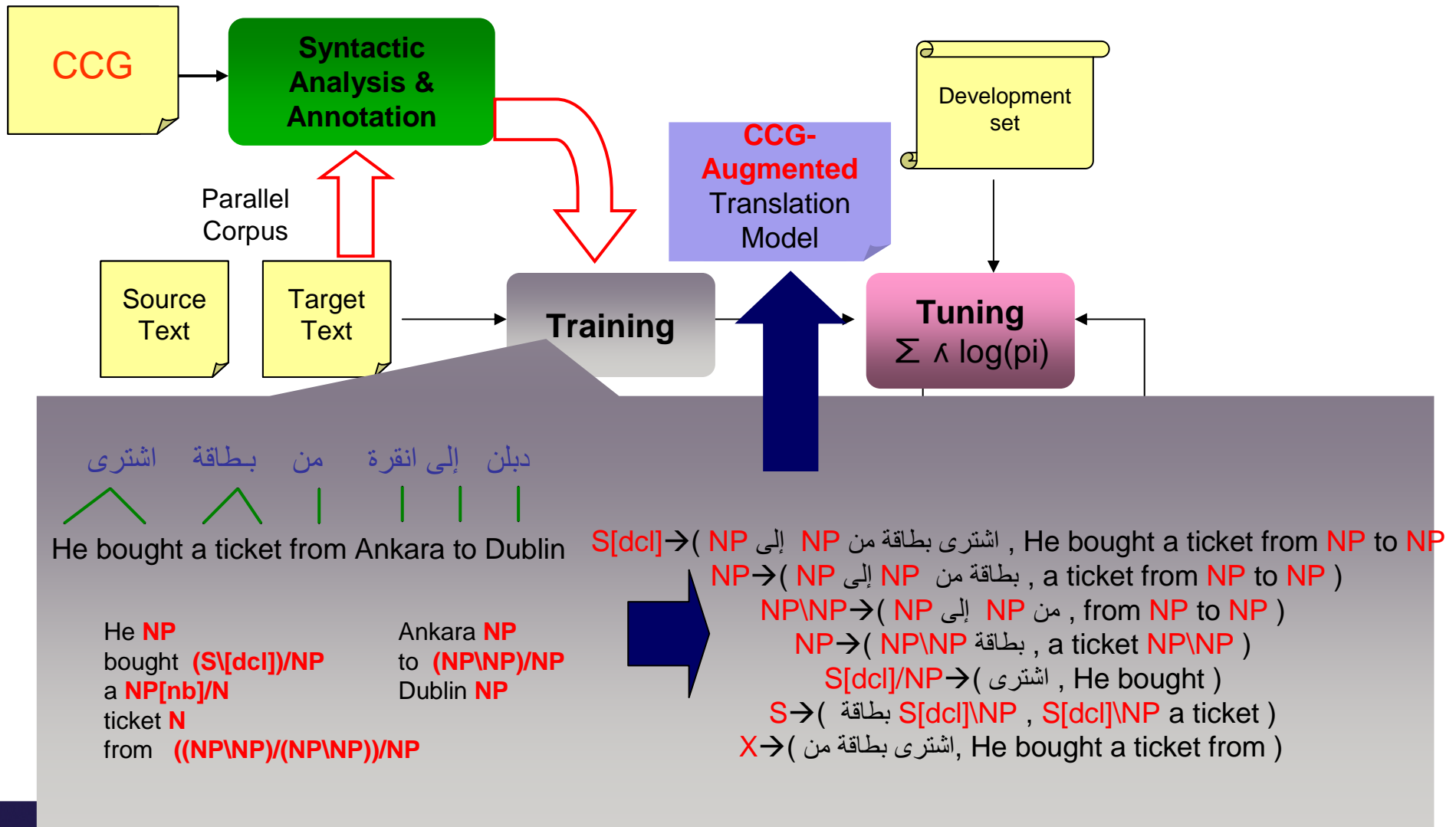SAMT combinatory operators →Label sparsity & accuracy

CCG-Augmented HPB SMT

Using Combinatory Categorial Grammar (CCG) to label target-side phrases instead of constituency grammar

+ CCG more flexible and richer structures → Label coverage

+ CCG supertags reflect rich syntactic information at the lexical level → Label sparsity & accuracy
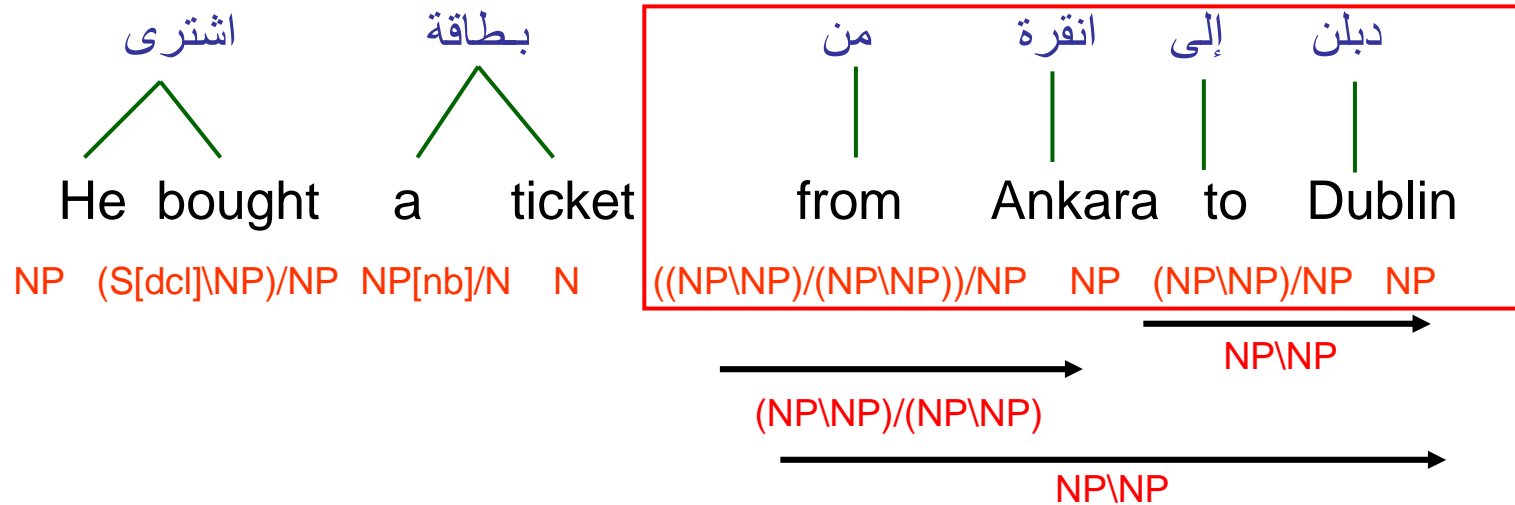
+ CCG is efficiently parsed

Combinatory Categorial Grammar (CCG)

Lexicon

He ⟶ NP
Walks ⟶ S\NP
reads ⟶ (S\NP)/NP , S\NP
a ⟶ NP/N
book ⟶ N, (S\NP)/NP
….

Combinatory Operators

FA: X/Y Y→ X
BA: Y X\Y → Y
FC: X/Y Y/Z → X/Z
BC: Y\Z X\Y → X\Z

CCG = +

He walks

NP S\NP
S\

He   reads   a   book

NP   (S\NP)/NP   NP/N   N
NP
S\NP
S

Primitive Category

Complex Category of the form X\Y

CCG-Augmented Hierarchical Phrase-Based SMT

CCG-Augmented Hierarchical Rule Extraction

S[dcl]→(اشترى بطاقة من NP إلى NP , He bought a ticket from NP to NP)
NP→(بطاقة من NP إلى NP , a ticket from NP to NP)
NP\NP→(من NP إلى NP , from NP to NP)
NP→(بطاقة NP\NP , a ticket NP\NP)
S[dcl]/NP→(اشترى , He bought)
S→(بطاقة S[dcl]\NP , S[dcl]\NP a ticket)
X→(اشترى بطاقة من , He bought a ticket from)

بطاقة	a ticket	NP
أنقرة	Ankara	NP
دبلن	Dublin	NP
من أنقرة إلى دبلن	from Ankara to Dublin	NP\NP
بطاقة من أنقرة إلى دبلن	a ticket from from Ankara to Dublin	NP
اشترى بطاقة	He bought a ticket	S[dcl]

CCG-Augmented HPB SMT vs. SAMT

AE news

CE news

Number of different syntactic labels

Percentage of phrase-pairs without syntactic label

CCG supertag labels are less sparse and are able to label more phrases than SAMT labels

CCG-Augmented HPB system could not outperform the HPB baseline→ CCG supertag labels still suffer from label sparsity problem
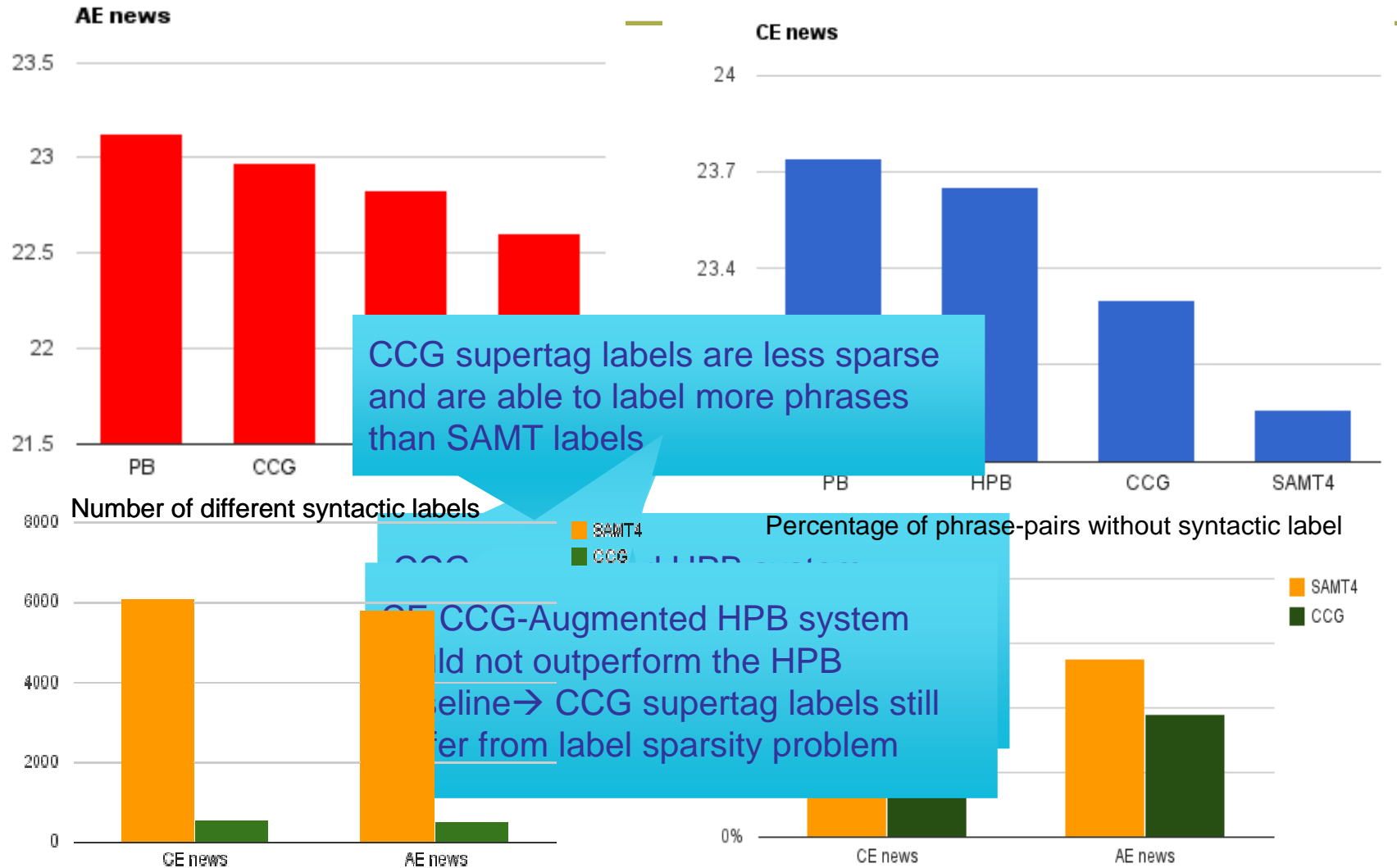
Solution

- *Softening* CCG supertags labels by employing *part of* the information represented in them.
- Two softening methods:
  - CCG contextual labels
  - Feature-removed CCG labels
- Goal:
  - Reduce label sparsity.
  - Loosen syntactic constraints.
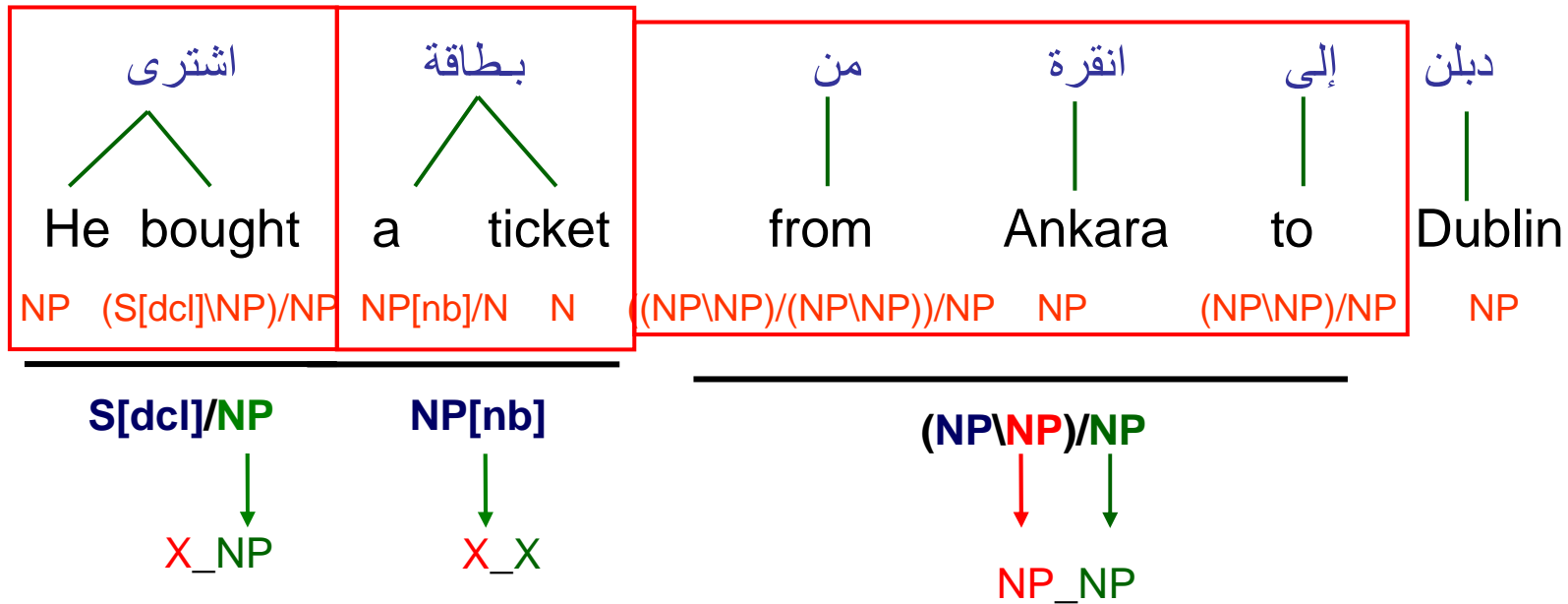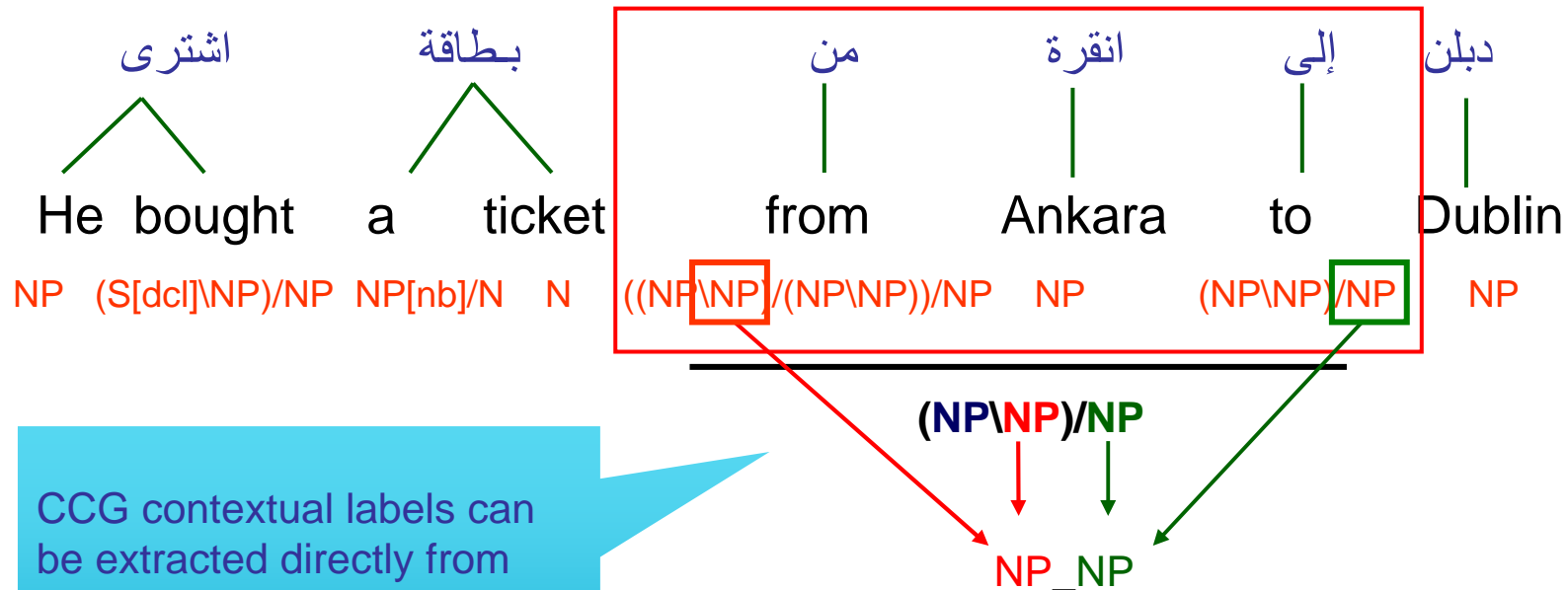- However, this comes at the expense of the *accuracy of the syntactic* labels.

Label Accuracy

Flexibility of syntactic constraints
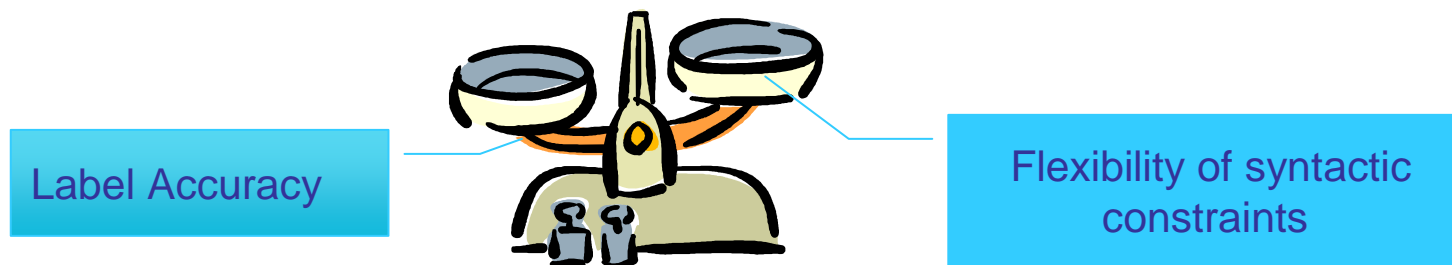
CCG Contextual Labels

$$C = (\ \backslash L1)/L2$$

L1 _ L2

Resulting category: R
Left Argument category: L1
Right Argument category: L2

اشترى
He bought
NP (S[dcl]\NP)/NP

بطاقة
a ticket
NP[nb]/N   N

من
from
(NP\NP)/(NP\NP))/NP

انقرة
Ankara
NP

إلى
to
(NP\NP)/NP

دبلن
Dublin
NP

S[dcl]/NP
X_NP

NP[nb]
X_X

(NP\NP)/NP
NP_NP

CCG Contextual Label Extraction

اشترى   بطاقة   من   انقرة   إلى   دبلن

He   bought   a   ticket   from   Ankara   to   Dublin

NP   (S[dcl]\NP)/NP   NP[nb]/N   N   ((NP\NP)/(NP\NP))/NP   NP   (NP\NP)/NP   NP

(NP\NP)/NP

NP_NP

CCG contextual labels can be extracted directly from CCG supertags without the need to parse the phrase

Feature-removed CCG Labels

Experiments

- Language pairs: Arabic—English and Chinese—English
- Data used: from the news and travelling speech expressions domains (IWSLT 2010 evaluation campaign).

	News	IWSLT
AE (ar-en)	48065	21484
CE (zh-en)	51044	63234

- Baseline Systems:
  - The PB baseline system: built using the Moses PB Decoder.
  - The HPB baseline system: built using the Moses Chart Decoder.
- CCG-based Systems:
  - CCG Context: uses CCG contextual labels.
  - CCG: uses CCG supertag labels.
  - CCG (s): uses feature-removed CCG supertag labels.
  - CCG Context (s): uses feature-removed CCG contextual labels.

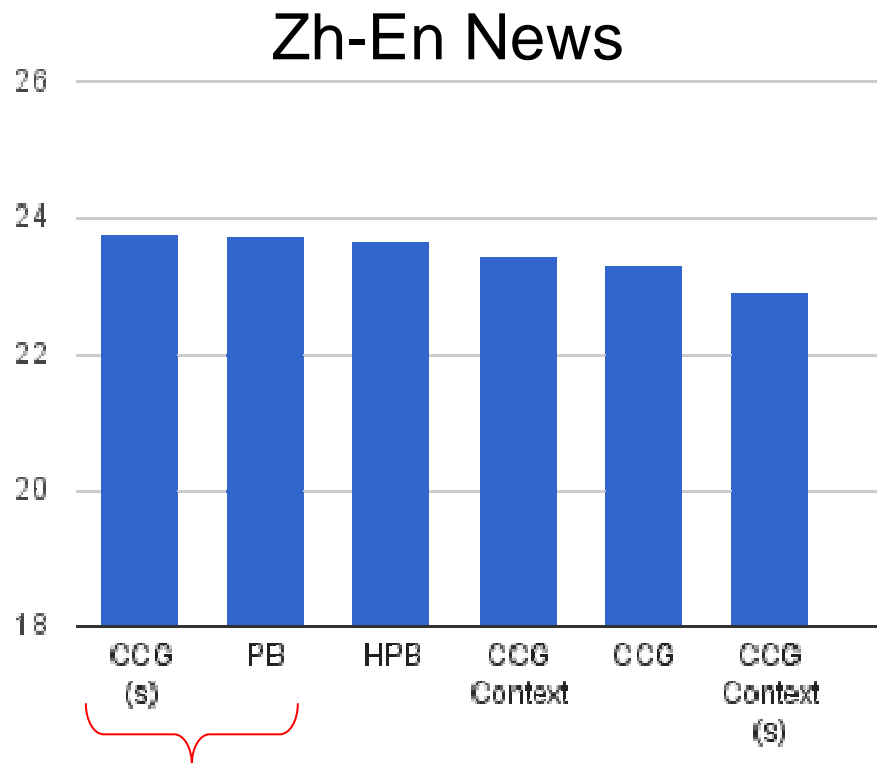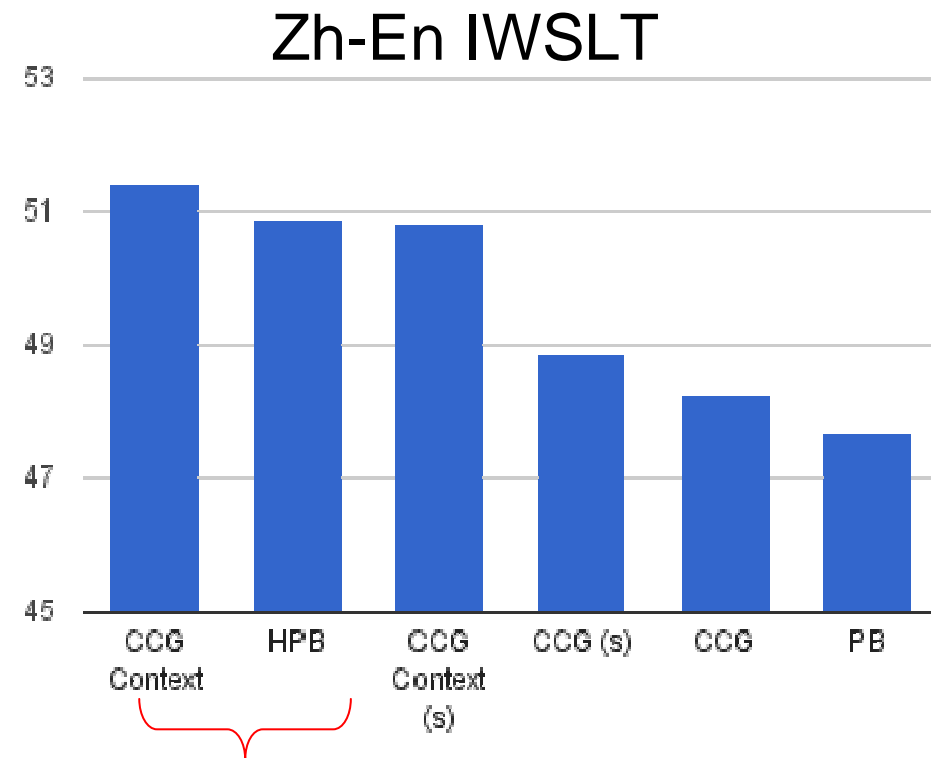# BLEU Scores for Arabic—English Experiments

Ar-En News

Ar-En IWSLT

Only 20k
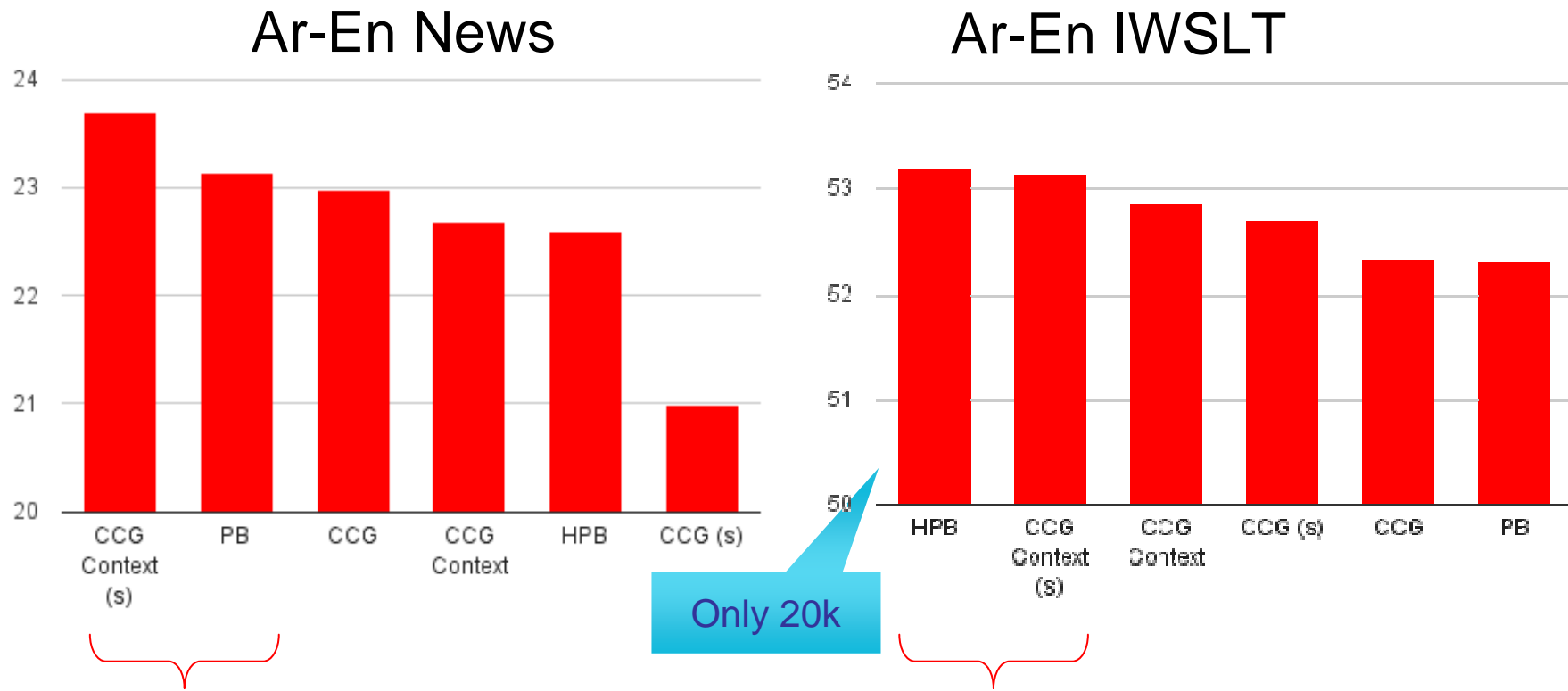
0.56 BLEU points

0.06 BLEU points

Statistically Significant
at p-level=0.05

BLEU Scores for Chinese—English Experiments

Label Sparsity
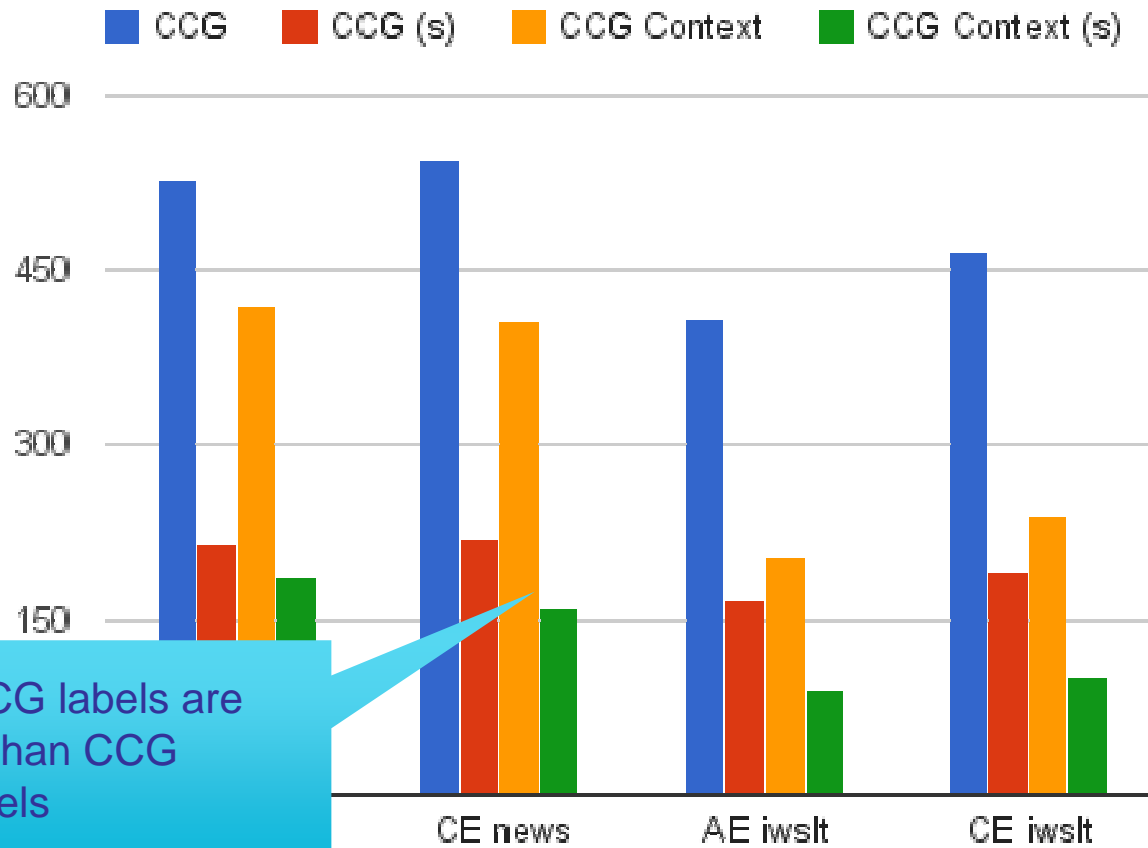

Number of different syntactic labels

Conclusions

- CCG label simplification demonstrated to be promising
  - At least one of the systems which use simplified CCG labels achieved better BLEU score than the CCG supertags HPB baseline.
  - Simplified CCG label systems were the best performing systems on all but AE IWSLT experiment.
- In comparison with CCG supertag labels, CCG contextual labels demonstrated to be:
  - less sparse
  - easier to extract than CCG supertags
- Simplification schemes did not show consistent improvement over baseline systems on a specific language pair or corpus type.

Future Work

- Conducting a thorough evaluation of *CCG label simplification schemes* using *larger training corpora* and on *more language pairs.*

- Examining the effect of *source language segmentation* on the performance of CCG-based systems.

- Using *system combination* on CCG-based systems to obtain a better performing system.

- Conducting a *manual analysis* on selected sentences to examine the effect of using CCG-based labels on producing more grammatical translations.

Thanks for your Attention!
Questions??