

# Syntax-based Rewriting for Simultaneous Machine Translation

He He, Alvin Grissom II, John Morgan,  
Jordan Boyd-Graber and Hal Daumé III



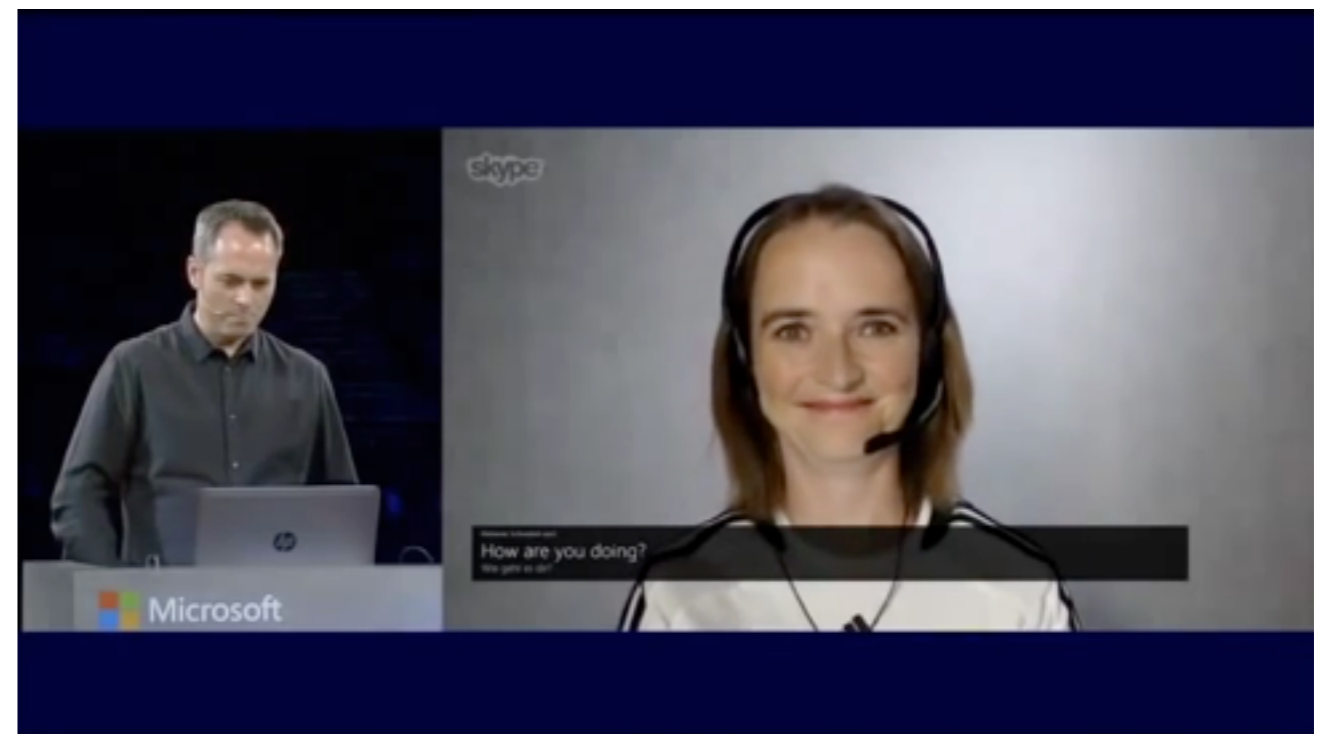
UNIVERSITY OF  
MARYLAND



University of Colorado  
Boulder

# Simultaneous Machine Translation

- Translate segments instead of sentences
- Trade-off between
  - Translation **quality**
  - Translation **delay**
- Intelligent segmentation
  - Grissom II et al., 2014 EMNLP
  - Oda et al., 2014 ACL
  - Fujita et al., 2014 Interspeech
- Challenges
  - Divergent word order
  - SOV vs. SVO



Skype Translator demo from WPC 2014

# Example I

我々は  
We-TOP

# Example I

我々は  
We-TOP

We

# Example 1

我々は 政府の  
We-TOP government-GEN

We

*... of the government*

# Example 1

我々は 政府の 構造 や 組織を  
We-TOP government-GEN structure and composition-ACC

We



# Example 1

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

We



# Example 1

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

We should change the structure and composition of the government



# Example I

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

我々は || 政府の構造や組織を変更すべき だ  
We should change the structure and composition of the government

# Example I

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

我々は || 政府の構造や組織を変更すべきだ  
We should change the structure and composition of the government



- Divergent word order → large segment → **delay**
- Use word orders similar to the source language

# Example II

我々は  
We-TOP

# Example II

我々は  
We-TOP

HOLD ON

We

# Example II

我々は 政府の  
We-TOP government-GEN

HOLD ON

We

*... of the government → the government's*

# Example II

我々は 政府の  
We-TOP government-GEN

**HOLD ON** the government's

We

*... of the government → the government's*

# Example II

我々は 政府の 構造 や 組織を  
We-TOP government-GEN structure and composition-ACC

**HOLD ON** the government's

We

*... of the government → the government's*

# Example II

我々は 政府の 構造 や 組織を  
We-TOP government-GEN structure and composition-ACC

**HOLD ON** the government's structure and composition

We

*... of the government → the government's*



# Example II

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

**HOLD ON** the government's structure and composition

We

*... of the government → the government's  
change ... → ... change → be changed*

# Example II

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

**HOLD ON** the government's structure and composition should be changed

We should change the structure and composition of the government

*... of the government → the government's  
change ... → ... change → be changed*

# Example II

我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

the government's structure and composition should be changed by us  
We should change the structure and composition of the government

*... of the government → the government's  
change ... → ... change → be changed*

# Example II

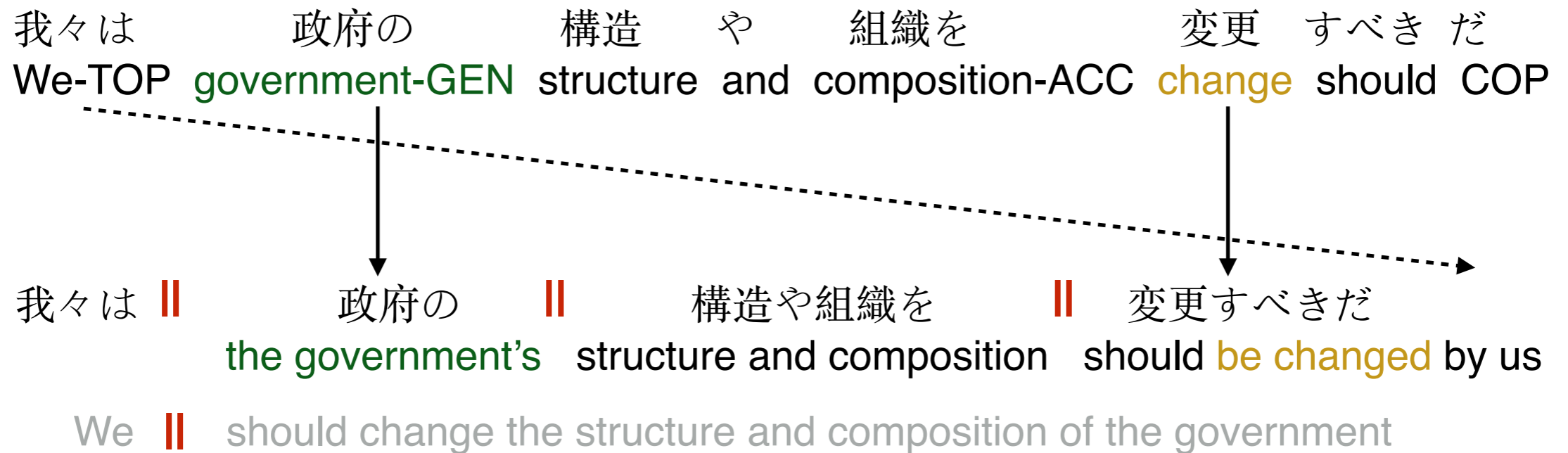
我々は 政府の 構造 や 組織を 変更 すべき だ  
We-TOP government-GEN structure and composition-ACC change should COP

我々は || 政府の || 構造や組織を || 変更すべきだ  
the government's structure and composition should be changed by us

We || should change the structure and composition of the government

*... of the government → the government's  
change ... → ... change → be changed*

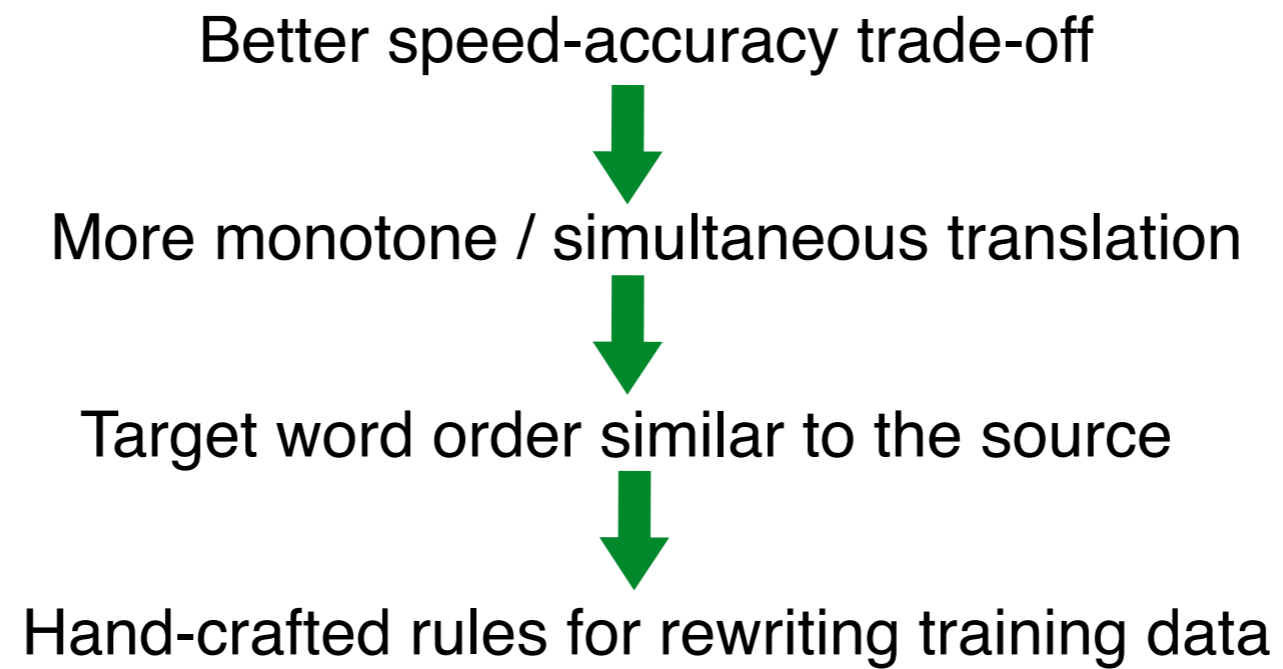
# Example II



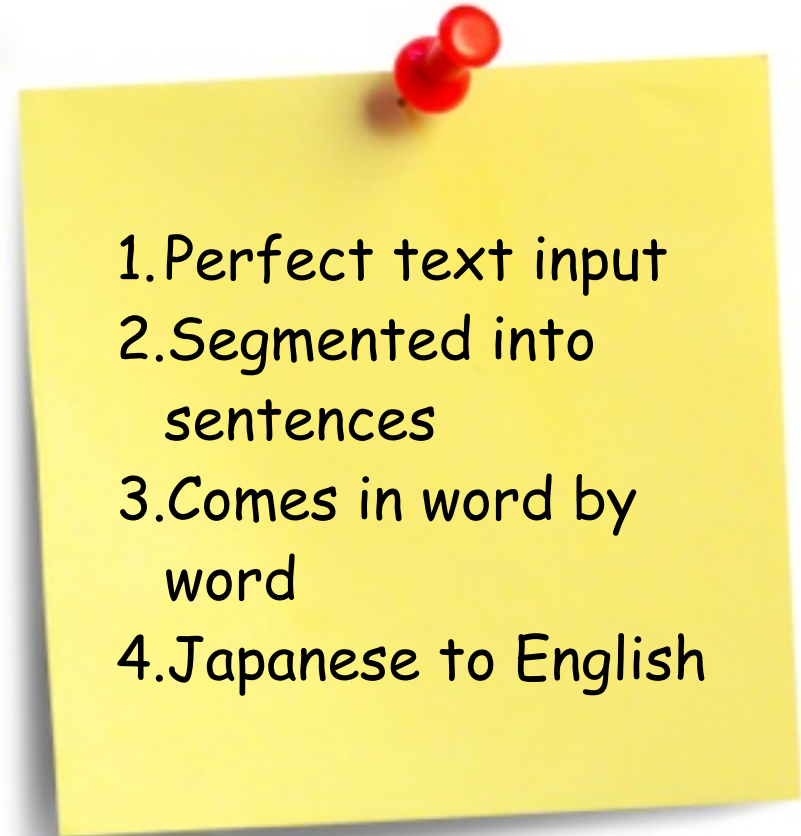
*... of the government → the government's*  
*change ... → ... change → be changed*

- Similar word order → smaller segment → reduce **delay**
- Rewrite target / reference sentence during training

# Roadmap



# The Problem of Delay Reduction

- 
1. Perfect text input
  2. Segmented into sentences
  3. Comes in word by word
  4. Japanese to English

- Rewrite the training target translation to  
min translation delay  
s.t. obey the grammar  
keep the original meaning
- Delay
  - # words to observe between two (segment) translations
- Data-driven approach
  - Lack of parallel translation-interpretation data
  - Flaw of human interpretation
- Rule-based approach
  - Linguistically inspired
  - Syntactical structure (parse trees)

# Transformation Rules: Voice

We should change the structure and composition of the government

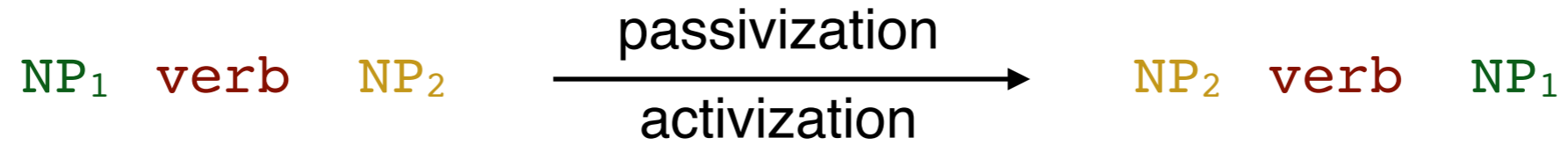


# Transformation Rules: Voice

the structure and composition of the government should be changed by us

The talk was denied by the boycott group spokesman

# Transformation Rules: Voice



the structure and composition of the government should be changed by us

the boycott group spokesman denied

the talk

- Useful when  $\text{len}(\text{NP}_2) > \text{len}(\text{NP}_1)$
- Accommodate verbs coming at the end

# Transformation Rules: Verb+Clause

They announced that the president will restructure the division

# Transformation Rules: Verb+Clause

NP<sub>1</sub> verb SBAR → SBAR NP<sub>1</sub> verb

the president will restructure the division , they announced

- Quotative: *say, announce, conclude ...*
- (Anti-)Factive: *know, realize, observe, doubt, deny ...*
- Belief: *think, believe, suspect ...*

# Transformation Rules: Noun Phrases

the structure and composition of the government

# Transformation Rules: Noun Phrases

$NP_1$  *of*  $NP_2$   $\longrightarrow$   $NP_2$  *'s / '*  $NP_1$

the government *'s* the structure and composition

# Transformation Rules: Noun Phrases

*It is* ADJP SBAR  $\longrightarrow$  SBAR *is* ADJP

*It is* important to remain watchful

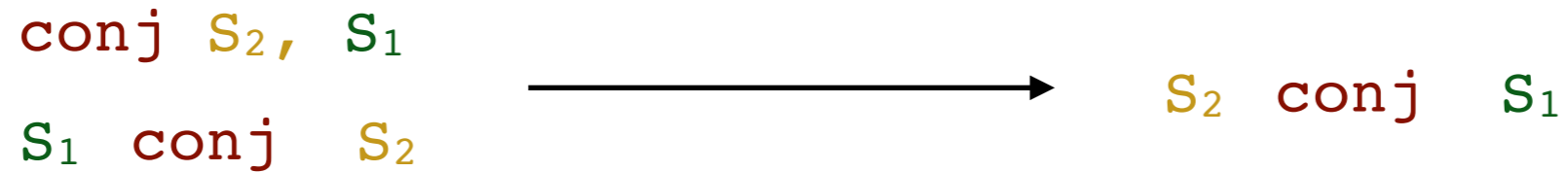
that we remain watchful *is* important

# Transformation Rules: Conjunction Clauses

We should march because winter is coming



# Transformation Rules: Conjunction Clauses



winter is coming ; because of this , We should march

It was cold outside, even though this is the case, they went to swim

- Clausal: *because (of), in order to*
- Contrastive: *despite, even though, although*
- Conditional: *even (if), as a result (of)*
- Misc: *according to*

# Rewriting Evaluation

Delay reduction

delay =  
# words / segment

We love the new world

我々は 新しい世界を 愛して

We new world the love

1

4

delay = 2.5 words / seg

The new world is loved by us

我々は 新しい世界を 愛して

We new world the love

2

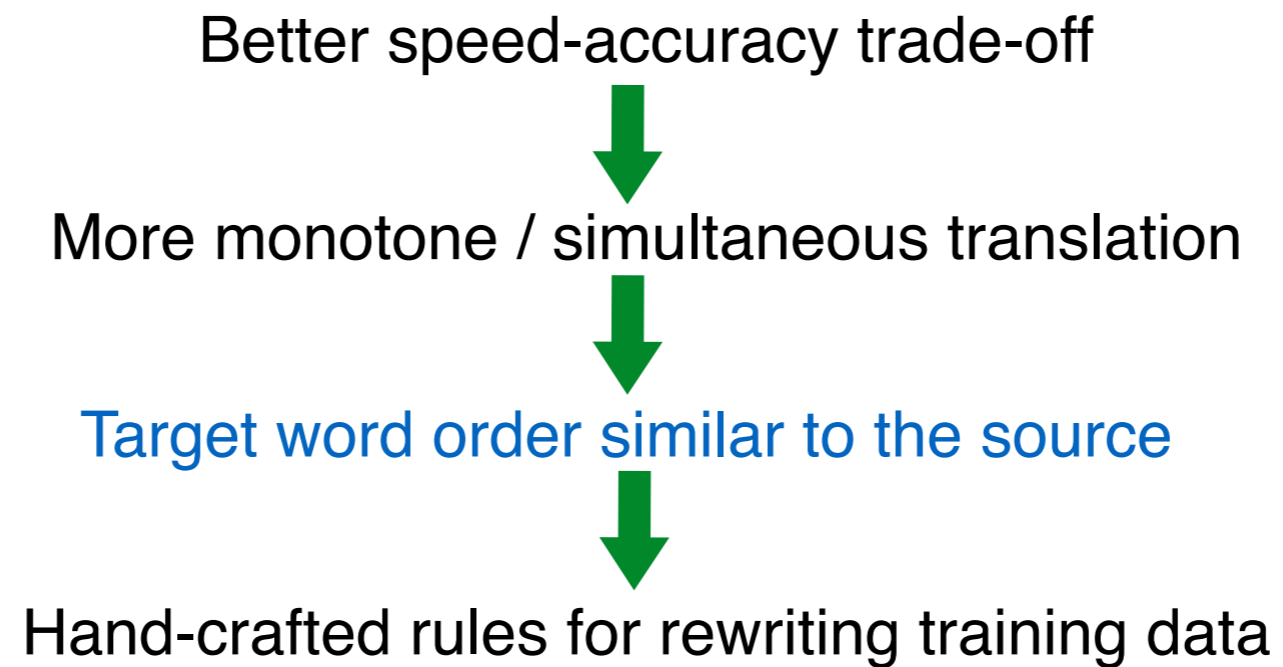
1

2

delay = 1.7 words / seg



# Roadmap



# Example Rewritten Sentences

この措置は、台湾の企業や投資家に、台湾ドルと他の外国通貨の間の為替オプション取引の機会を**提供する**ことになるという。

**The move will offer** options trading for Taiwan companies and investors in exchanges between the Taiwan dollar and other foreign currencies.

*Options trading for Taiwan companies and investors **will be offered by the move** in exchanges between the Taiwan dollar and other foreign currencies.*

さらに同議長は、商品価格はかつてほどにはインフレ指標として有効ではなくなった、と**述べた**。

**Greenspan also said** that commodity prices were not as good an indicator of inflation as they once were.

*Commodity prices were not as good an indicator of inflation as they once were, **Greenspan also said.***

# Example Rewritten Sentences

Lee said the central bank believes exchange rates should be left to the market, although he expressed concern about sharp fluctuations.

# Example Rewritten Sentences

**Lee said** the central bank believes exchange rates should be left to the market, although he expressed concern about sharp fluctuations.

*The central bank believes exchange rates should be left to the market, although he expressed concern about sharp fluctuations, **Lee said**.*

# Example Rewritten Sentences

**Lee said** the central bank believes exchange rates should be left to the market, although he expressed concern about sharp fluctuations.

*The central bank believes exchange rates should be left to the market, **although** he expressed concern about sharp fluctuations, **Lee said**.*

*He expressed concern about sharp fluctuations, **although this is the case**, the central bank believes exchange rates should be left to the market, **Lee said**.*

# Example Rewritten Sentences

**Lee said** the central bank believes exchange rates should be left to the market, although he expressed concern about sharp fluctuations.

*The central bank believes exchange rates should be left to the market, although he expressed concern about sharp fluctuations, Lee said.*

*He expressed concern about sharp fluctuations, although this is the case, the central bank believes exchange rates should be left to the market, Lee said.*

*Concern was expressed by him about sharp fluctuations, although this is the case, exchange rates should be left to the market, the central bank believes, Lee said.*



# Error Analysis

- Word usage error

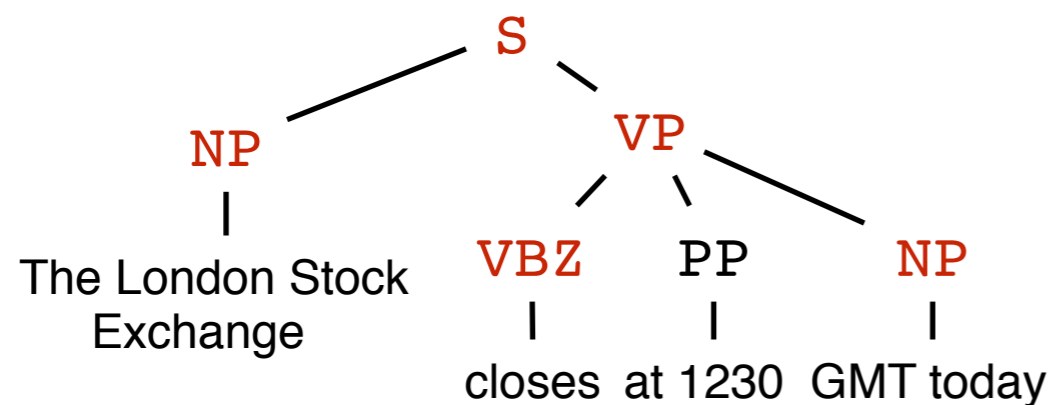
ASEAN currently **comprises** Indonesia , Malaysia , Vietnam , Thailand , Singapore , the Philippines and Brunei.

*Indonesia , Malaysia , Vietnam , Thailand , Singapore , the Philippines and Brunei currently **is comprised by** ASEAN.*

- Parsing error

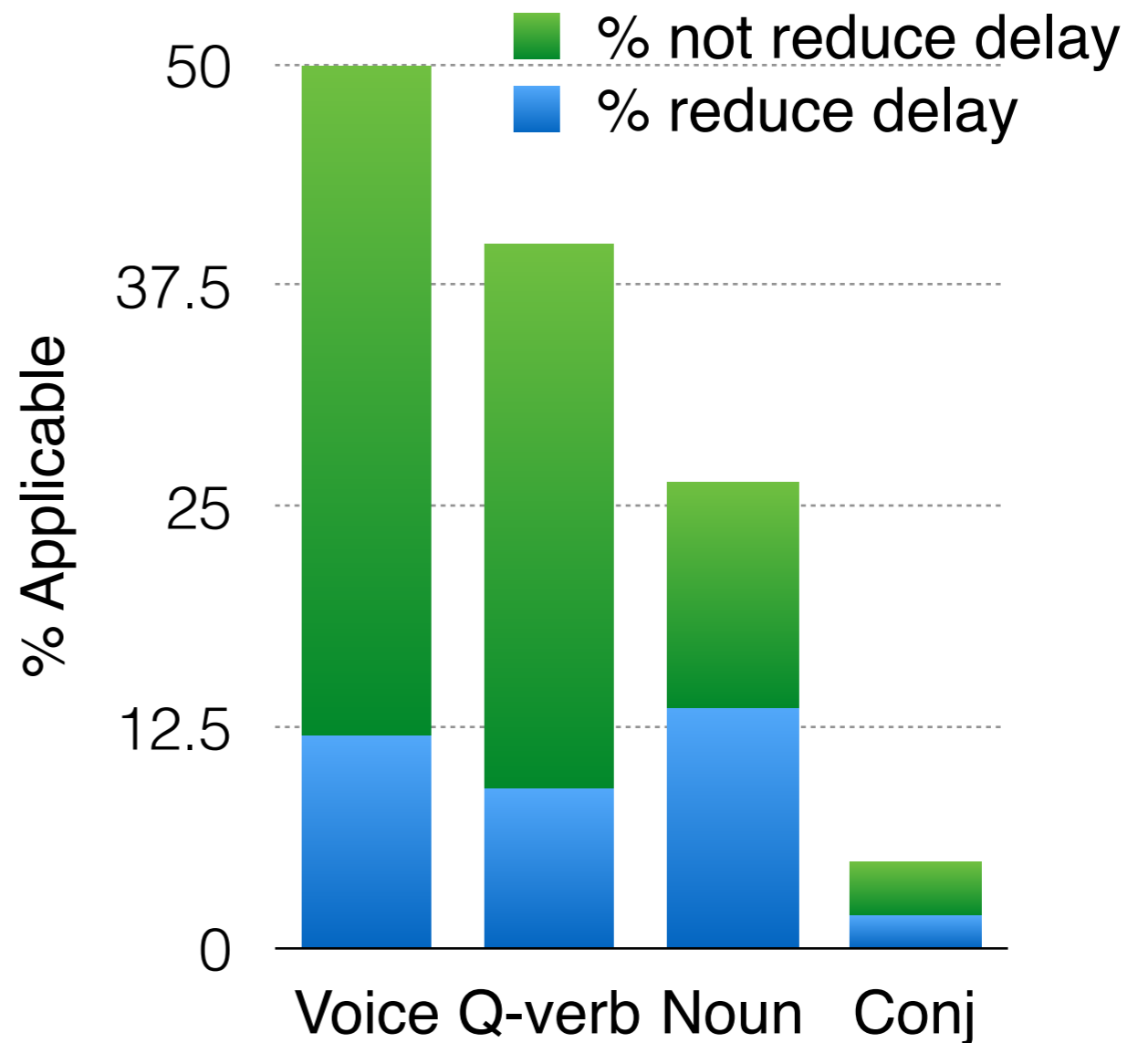
The London Stock Exchange closes at 1230 GMT today.

*GMT today is closed at 1230 by the London Stock Exchange.*

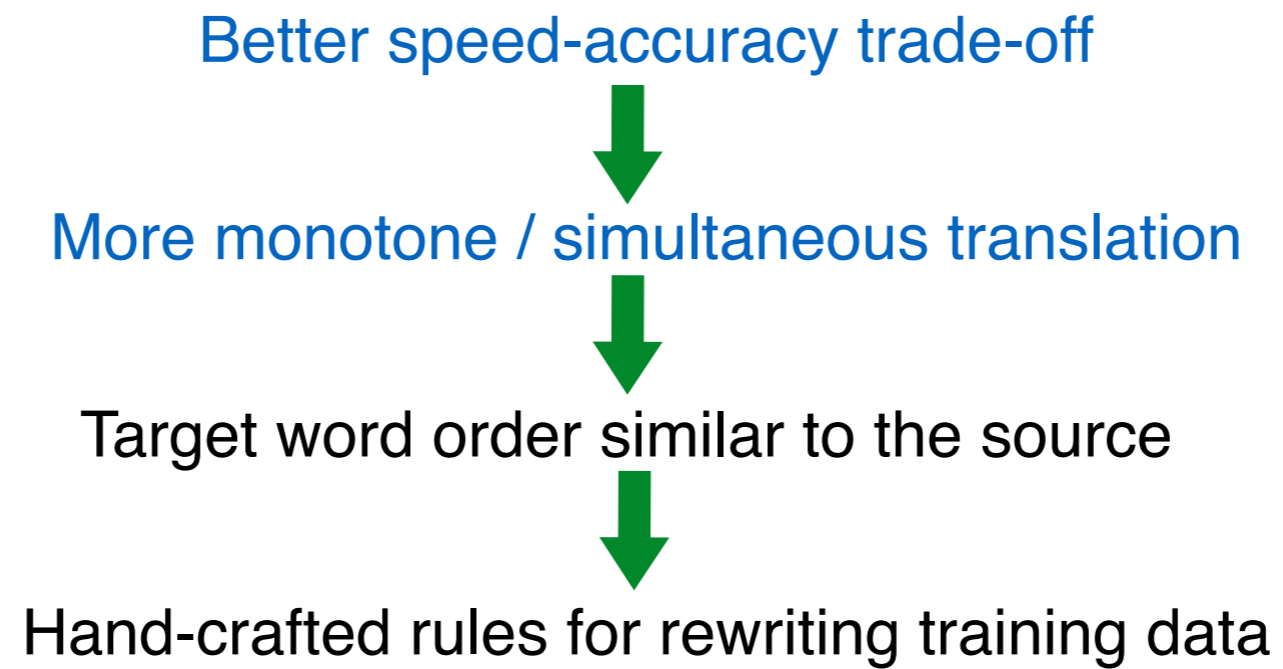


# Statistics of Rewritten Sentences

- Dataset
  - Reuters Ja-En corpus of news articles
- Most widely applicable
  - Voice changing
- Most effective
  - Possessive case, *It is* + clause
- 95.5% words have the same semantic role labels in the corresponding rewritten sentences

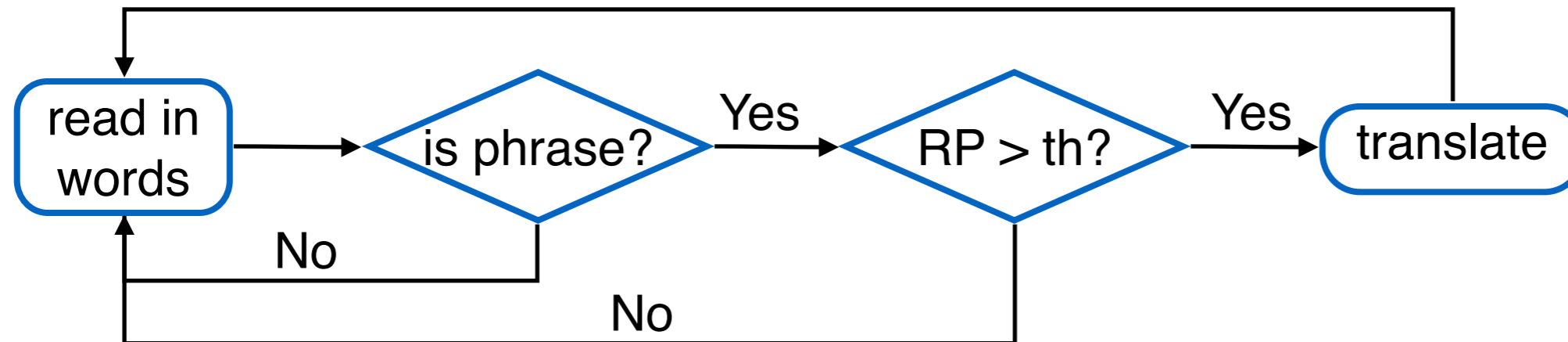


# Roadmap



# Machine Translation Settings

- Baseline MT system
  - Phrase-based model trained by Moses
  - Trigram language model, GIZA++ for word alignment, MIRA for tuning
- Segmentation (Fujita et al., 2013)



- Pr(translation of two consecutive phrases being monotone)
- Threshold controls how aggressive the segmentation is

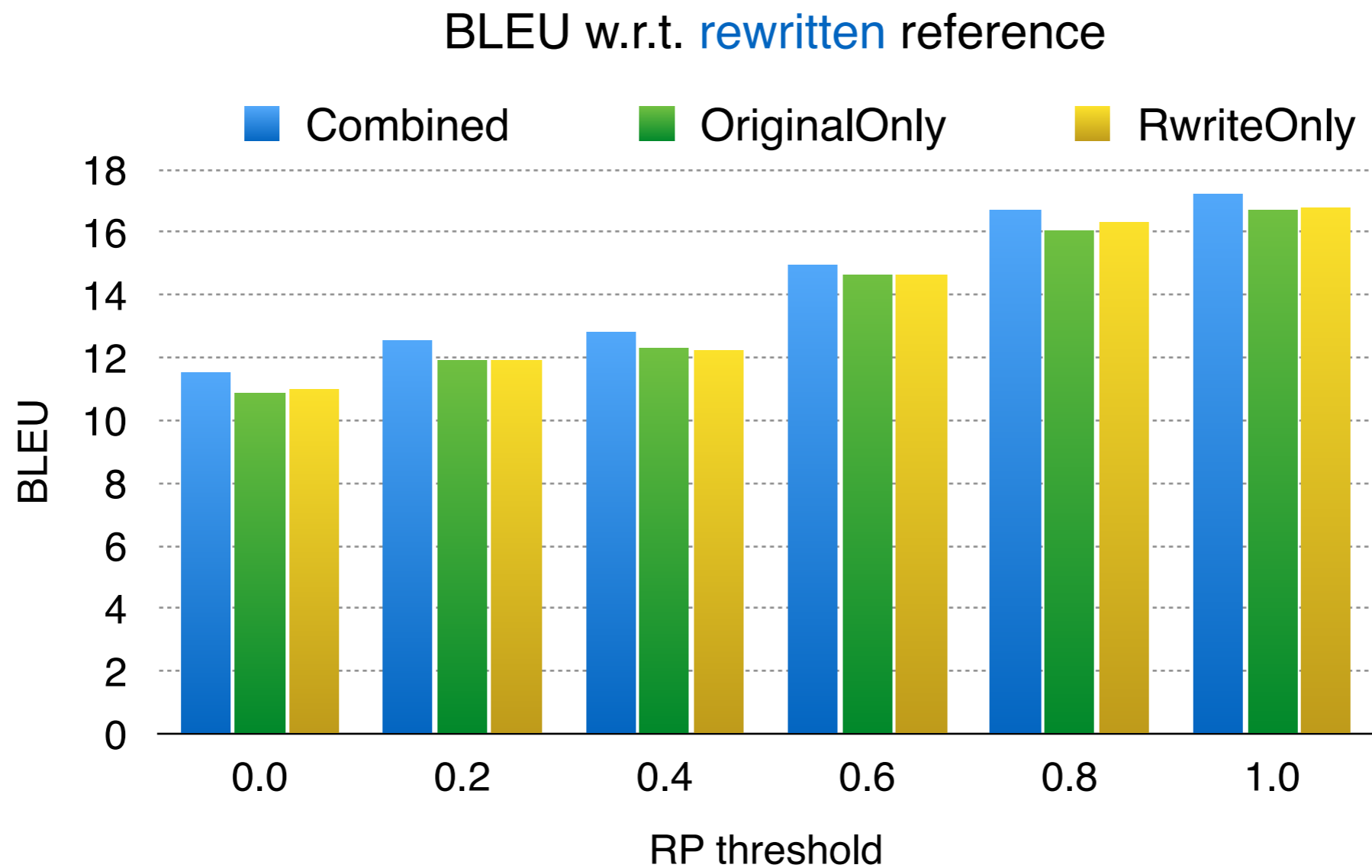


# Machine Translation Models

- OriginalOnly
  - Only use original reference translations
- RewriteOnly
  - Only use rewritten reference translations
- Combined [helps alleviate errors introduced by rewriting](#)
  - Interpolate language models of OriginalOnly and RewriteOnly
  - Combine translation models of OriginalOnly and RewriteOnly by fill-up (Bisazza et al., 2011)
  - Tune on rewritten reference translations
- RP threshold
  - Controls length of segments
  - 0.0 (phrase), 0.2, 0.4, 0.6, 0.8, 1.0 (batch)

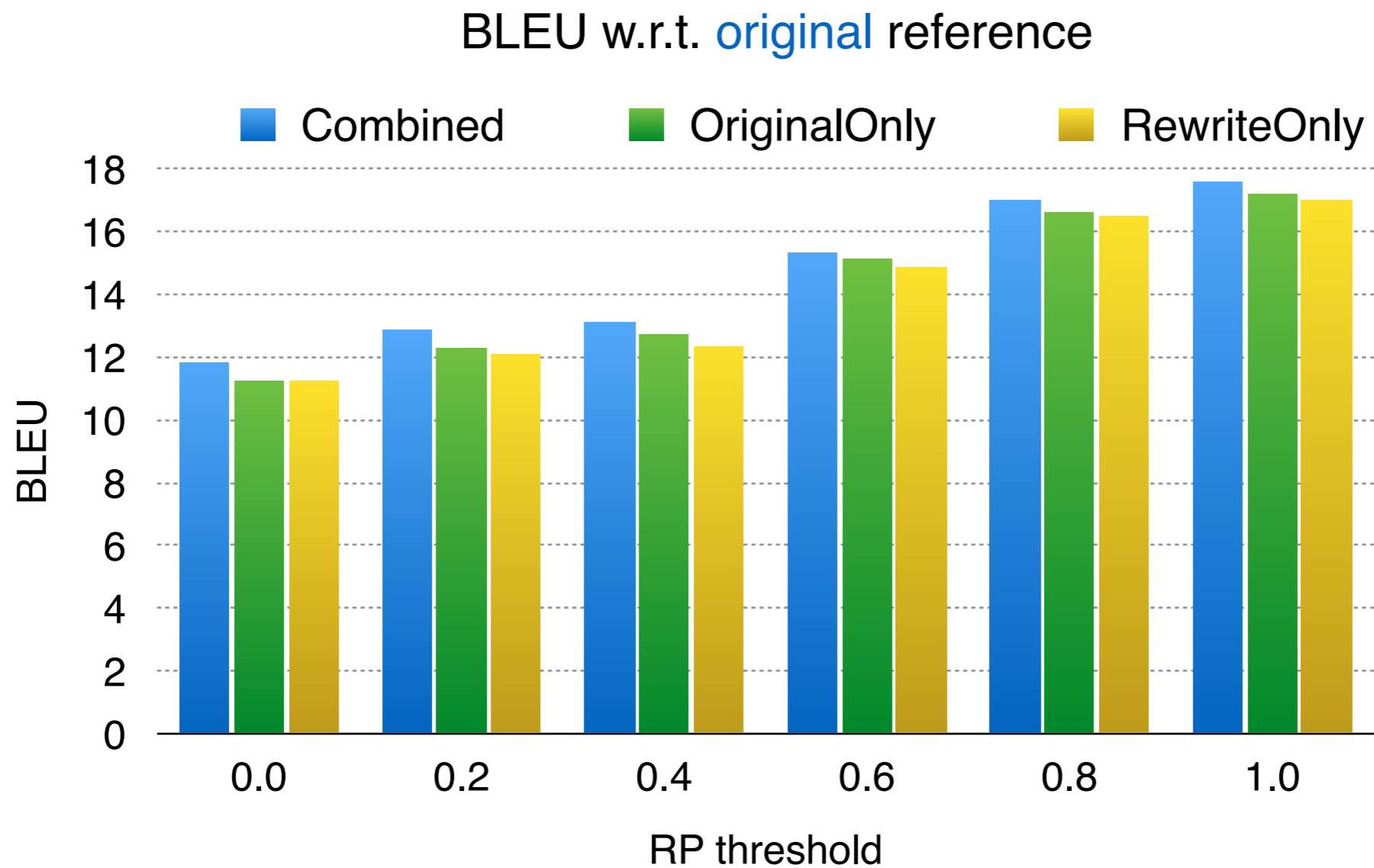
# Experiments: Machine Translation

- Better trade-off against the rewritten reference



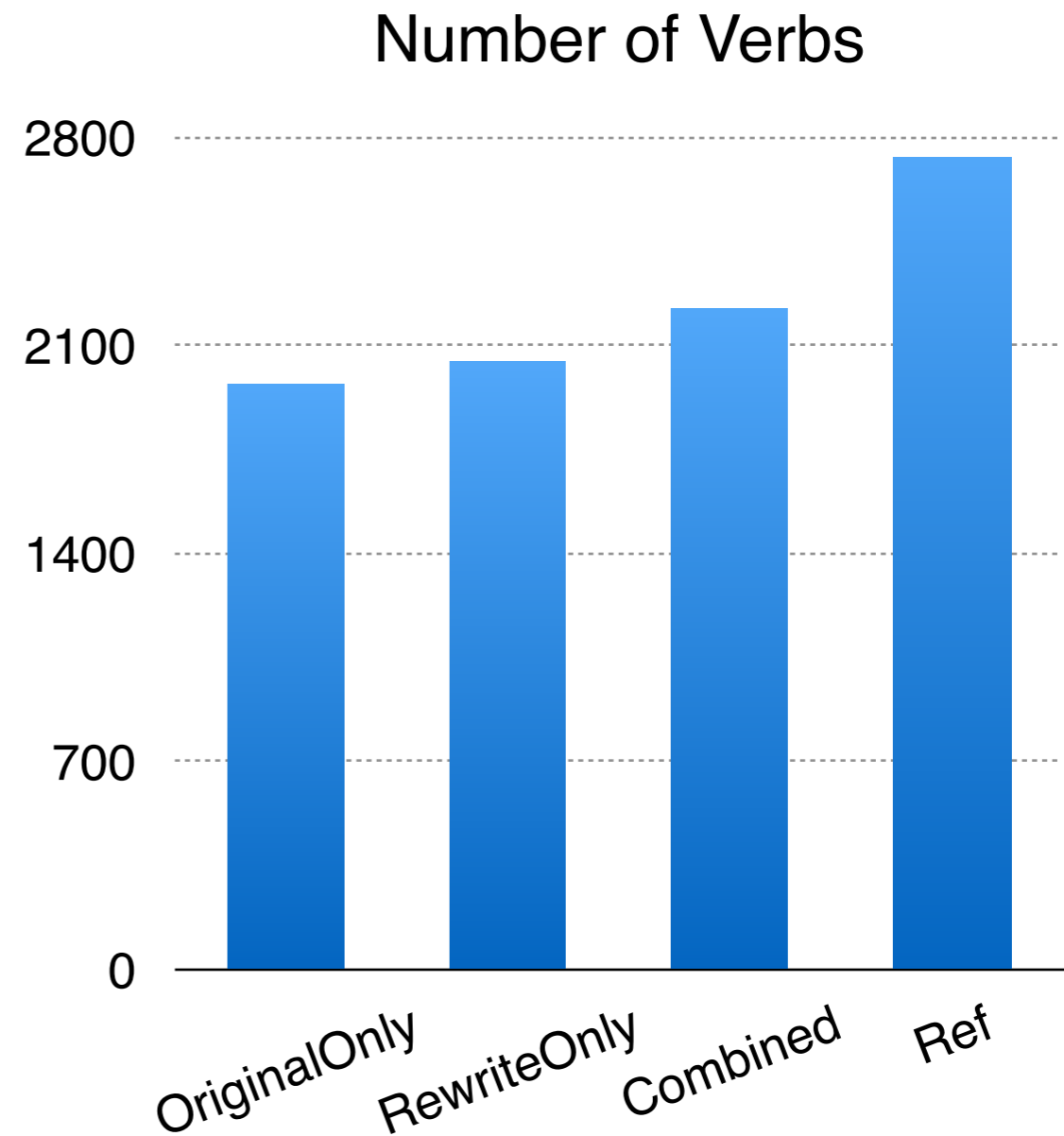
# Experiments: Machine Translation

- Better trade-off against the original reference
  - Especially with shorter segments



# Bonus: Help Keep the Verbs

- Verb dropping problem in SOV-to-SVO translation, e.g., Ja-En, De-En
  - Long range distortion
- Putting verb at the end helps
  - Similar to source-side pre-reordering





# Error Analysis

## Ref

- **He also said** that the real dangers for the euro lay in the potential for divergences in the domestic policy needs among the various participating nations of the single currency.

## OriginalOnly

- **He** also for the euro, is a real danger to the stability of the single currency in many different countries and domestic policies to be implemented. **No verb!** The need for the possibility of a difference.

## RewriteOnly

- **He also** for the euro is a real danger to the stability of the single currency in many different countries and domestic policies to be implemented. **Duplicated subject!** The need for the possibility of a difference, **he said**.

# Conclusion



Better segmentation methods: +cache

Data selection based on syntax

Learning from human interpretation

...

**Thank you!**