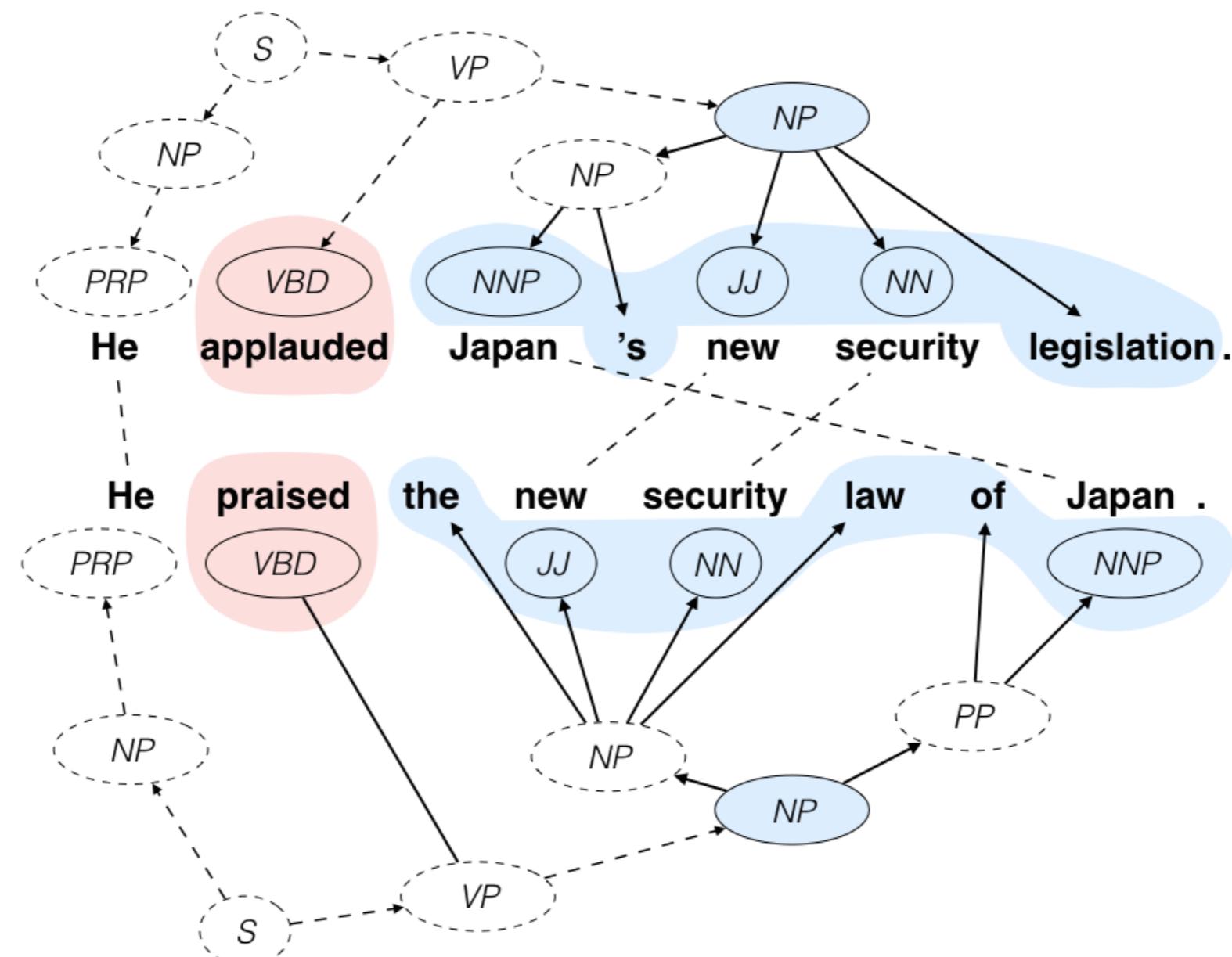


Optimizing Statistical Machine Translation for **Text Simplification**

Wei Xu	UPenn → Ohio State University
Courtney Napolis	JHU
Ellie Pavlich	UPenn
Quanze Chen	UPenn → University of Washington
Chris Callison-Burch	UPenn

Text Simplification

(A Monolingual Machine Translation Problem)



- ★ for children, disabled, non-native speakers ...
- ★ for other NLP tasks (MT, summarization ...)

Our Paper Last Year (TACL '15)



EMNLP 2015
CONFERENCE ON EMPIRICAL METHODS
IN NATURAL LANGUAGE PROCESSING
LISBON

Problems in Current Text Simplification Research

Wei Xu UPenn
Chris Callison-Burch UPenn
Courtney Napoles JHU

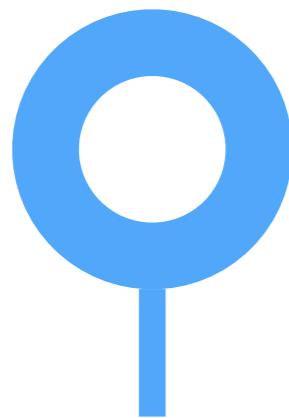
Follow @cocoweixu

TACL paper @ EMNLP Sep-20-2015



Problems in Current Text Simplification Research (TACL '15)

Data



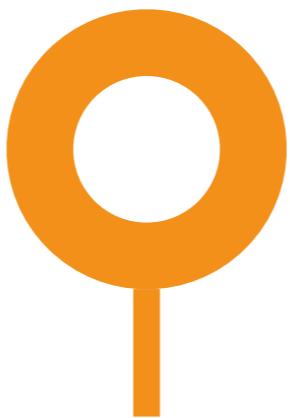
Simple Wikipedia
was not that simple



The Newsela Corpus!

very high quality
freely available

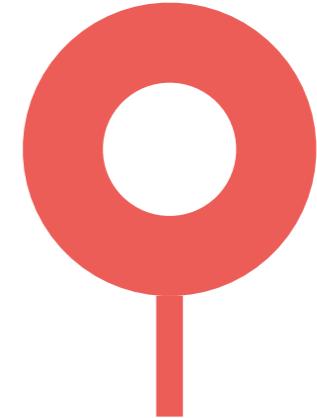
Evaluation



no reliable
automatic metric

(often use Moses as a black-box)

System



not much
paraphrasing

(often use Moses as a black-box)

This Paper and This Talk

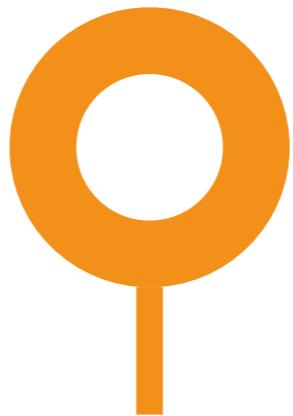
(TACL '16)

Data



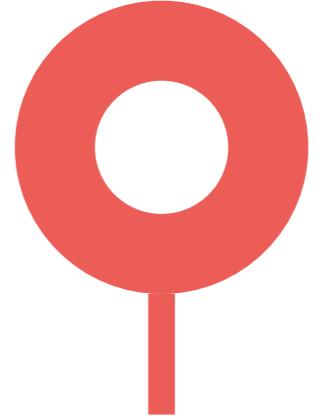
Simple Wikipedia
was not that simple

Evaluation



no reliable
automatic metric

System



not much
paraphrasing

This Paper and This Talk

(TACL '16)

Data

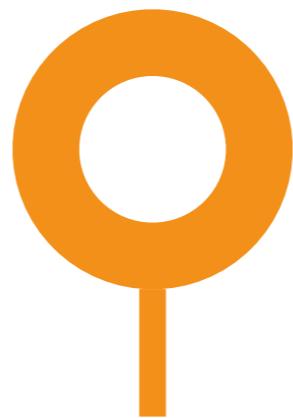


Simple Wikipedia
was not that simple



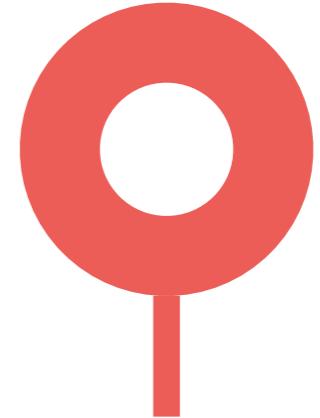
Crowdsourcing
3.5k sentences
8 references

Evaluation



no reliable
automatic metric

System



not much
paraphrasing

This Paper and This Talk

(TACL '16)

Data



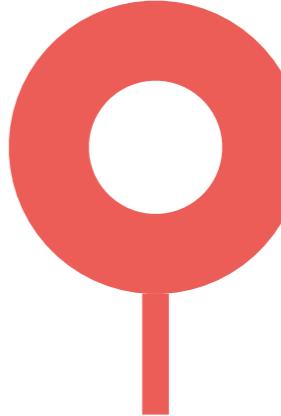
**Simple Wikipedia
was not that simple**

Evaluation



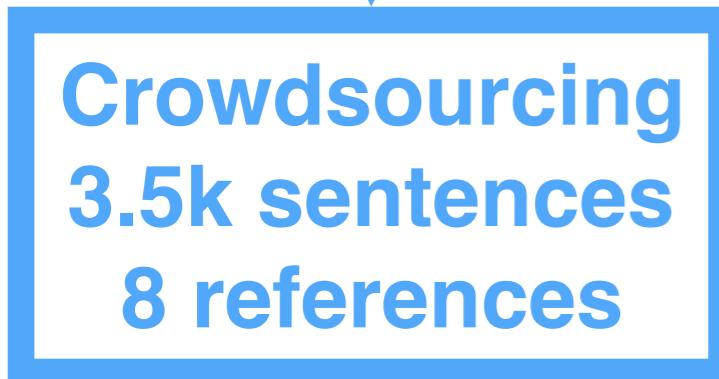
**no reliable
automatic metric**

System



**not much
paraphrasing**

↓



**Crowdsourcing
3.5k sentences
8 references**

↓



**SARI
the 1st metric
that works**

This Paper and This Talk

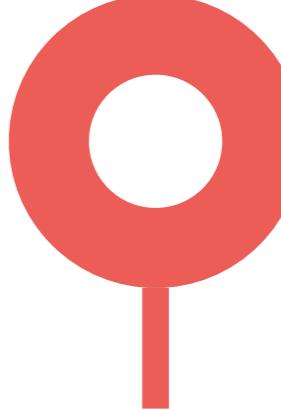
(TACL '16)

Data

**Simple Wikipedia
was not that simple**

Evaluation

**no reliable
automatic metric**

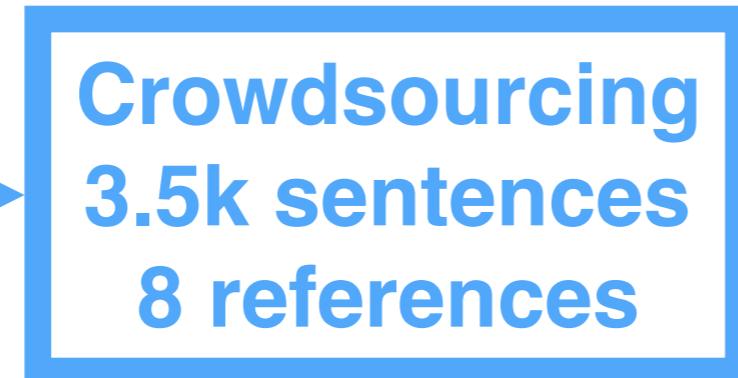
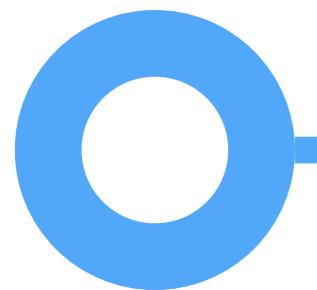
System

**not much
paraphrasing**


**Crowdsourcing
3.5k sentences
8 references**

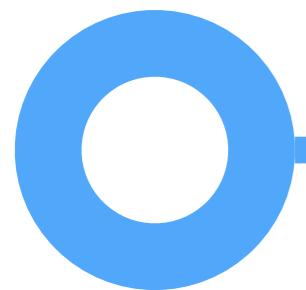

**SARI
the 1st metric
that works**


**state-of-the-art
paraphrasing
for simplicity**

Data



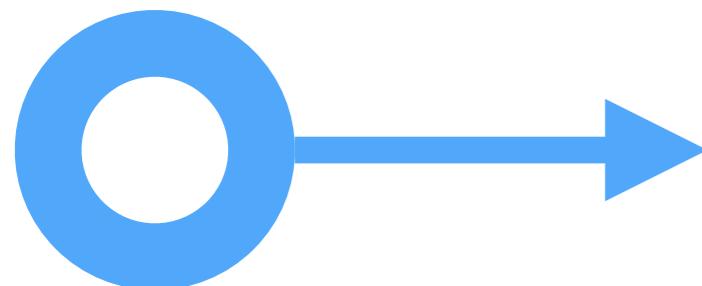
Data



Crowdsourcing
3.5k sentences
8 references

- ★ **paraphrasing!** (little deletion; no splitting)
- ★ simple quality control (see our NAACL 2015 paper)
- ★ freely available

Data



Crowdsourcing
3.5k sentences
8 references

- ★ **paraphrasing!** (little deletion; no splitting)
- ★ simple quality control (see our NAACL 2015 paper)
- ★ freely available

Wiki

Jeddah is the **principal gateway** to Mecca, **Islam's holiest city**, which **able-bodied** Muslims **are required to** visit at least once in their **lifetime**.

Data



Crowdsourcing
3.5k sentences
8 references

- ★ **paraphrasing!** (little deletion; no splitting)
- ★ simple quality control (see our NAACL 2015 paper)
- ★ freely available

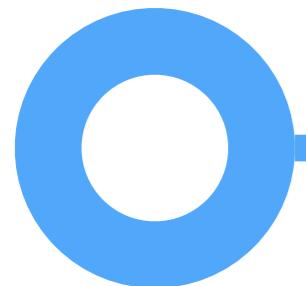
Wiki

Jeddah is the **principal gateway** to Mecca, **Islam's holiest city**, which **able-bodied** Muslims **are required to** visit at least once in their **lifetime**.

Turk #1

Jeddah is the **main entrance** to Mecca, **the holiest city in Islam**, which **all healthy** Muslims **need to** visit at least once in their **life**.

Data



Crowdsourcing
3.5k sentences
8 references

- ★ **paraphrasing!** (little deletion; no splitting)
- ★ simple quality control (see our NAACL 2015 paper)
- ★ freely available

Wiki

Jeddah is the **principal gateway** to Mecca, **Islam's holiest city**, which **able-bodied** Muslims **are required to** visit at least once in their **lifetime**.

Turk #1

Jeddah is the **main entrance** to Mecca, **the holiest city in Islam**, which **all healthy** Muslims **need to** visit at least once in their **life**.

Our System

Jeddah is the **main** gateway to Mecca, Islam's holiest city, which **sound** Muslims **have to** visit at least once in their **life**.

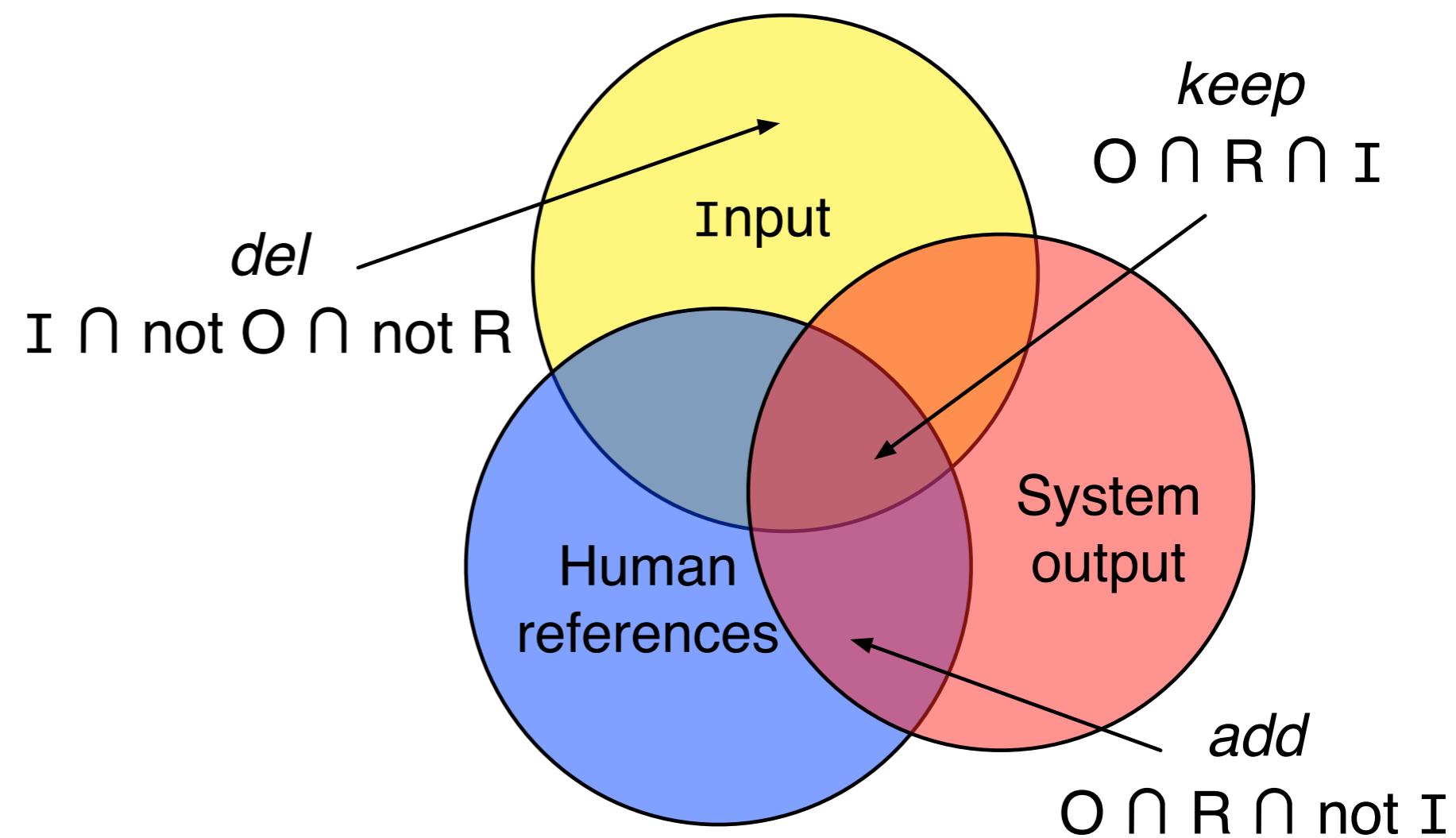
Evaluation



Evaluation



SARI
the 1st metric
that works

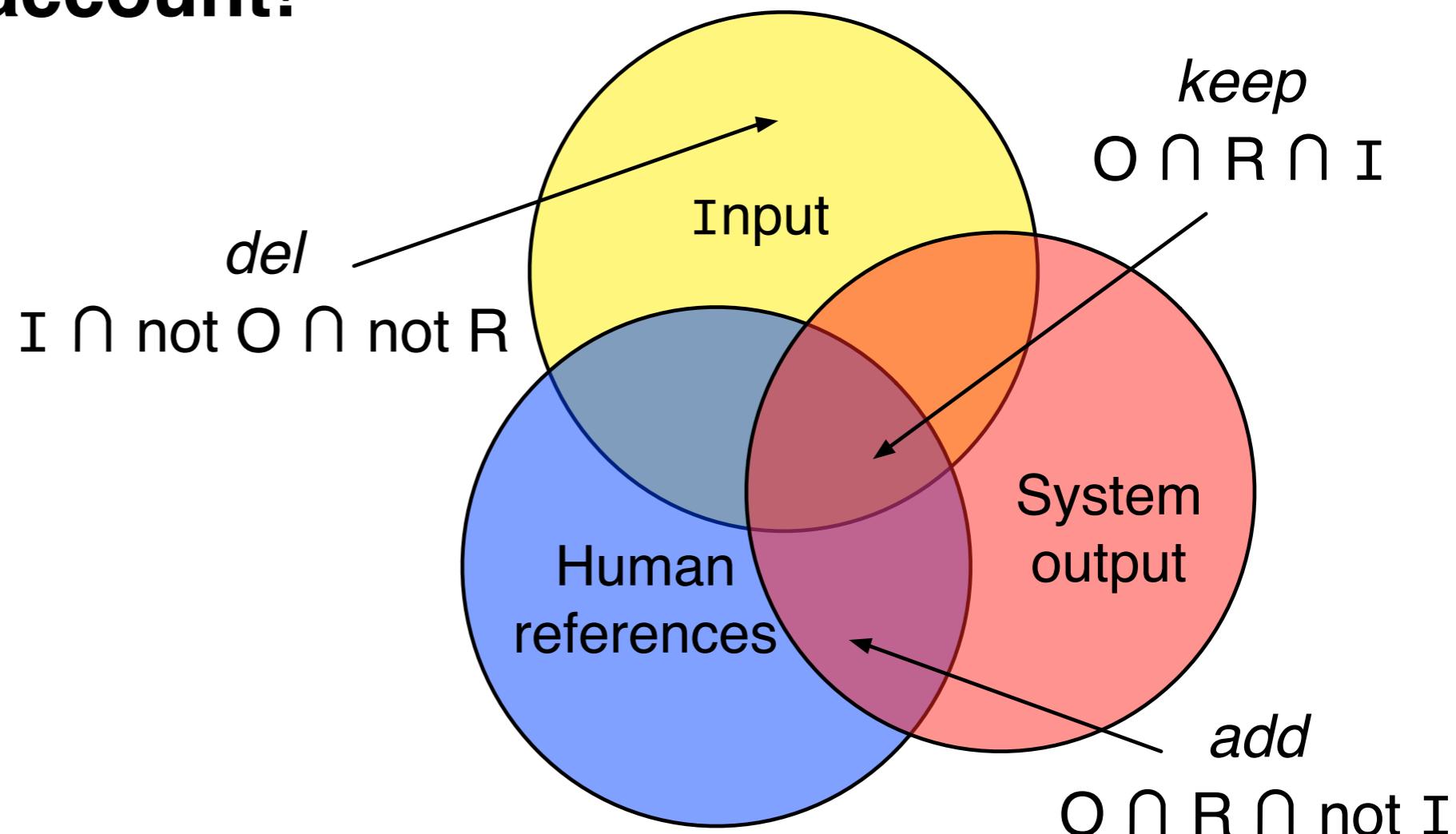


Evaluation



SARI
the 1st metric
that works

- ★ light-weighted and tunable
- ★ SARI — “the BLEU for simplification”
- ★ **take input into account!**
- ★ use F-measure

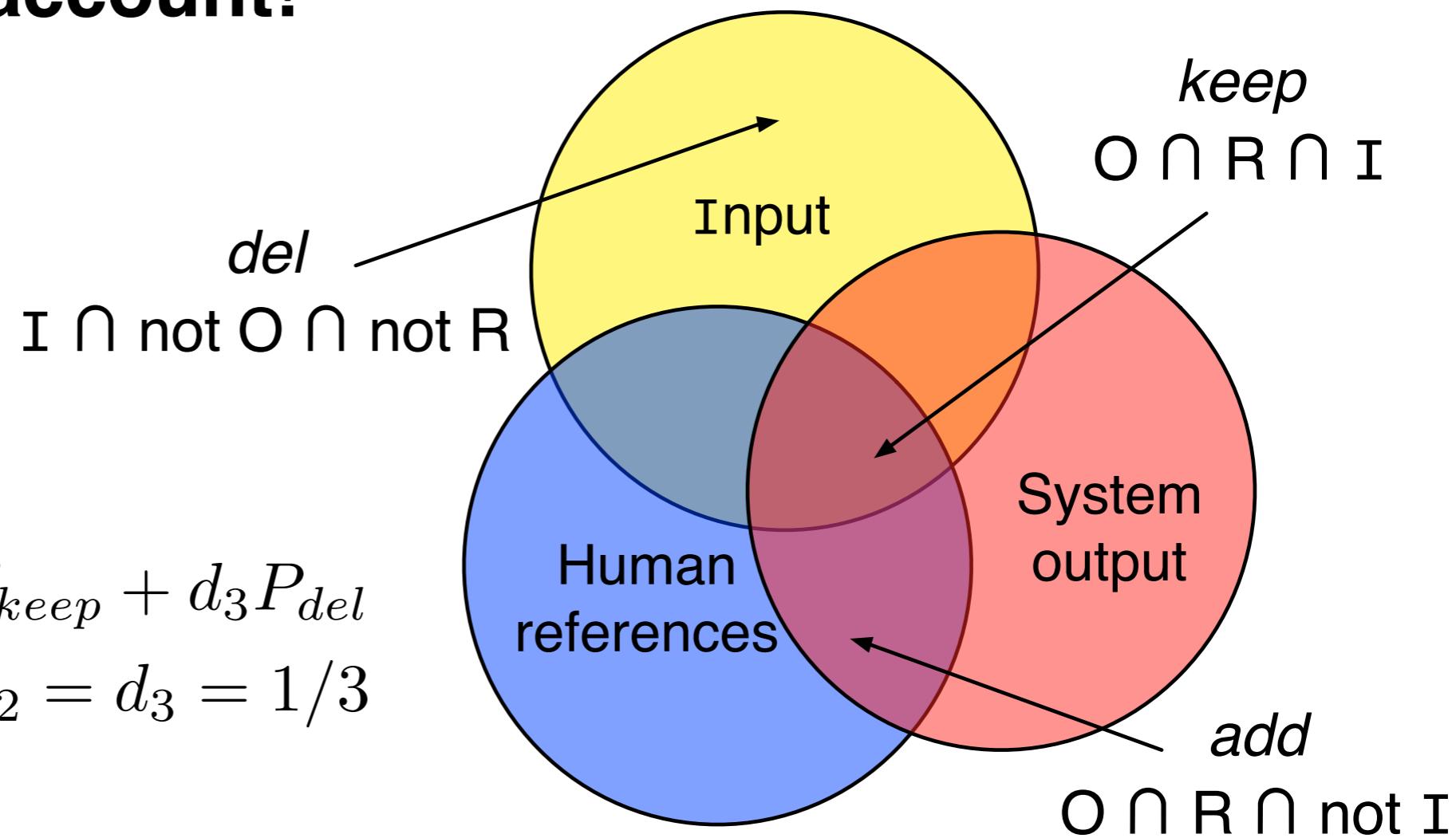


Evaluation



SARI
the 1st metric
that works

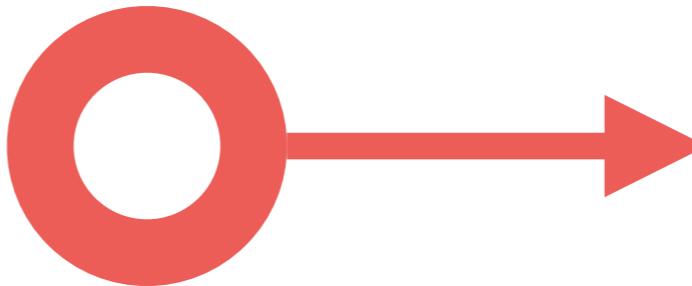
- ★ light-weighted and tunable
- ★ SARI — “the BLEU for simplification”
- ★ **take input into account!**
- ★ use F-measure



$$\text{SARI} = d_1 F_{add} + d_2 F_{keep} + d_3 P_{del}$$

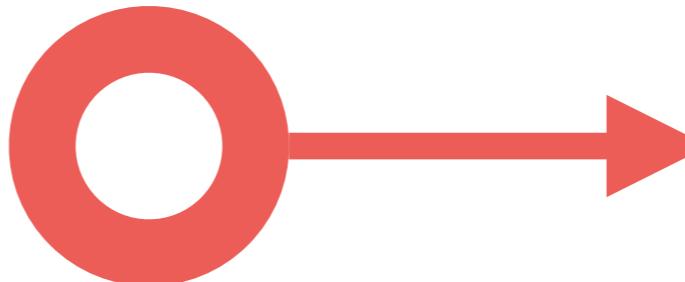
$$d_1 = d_2 = d_3 = 1/3$$

System



state-of-the-art
paraphrasing
for simplicity

System

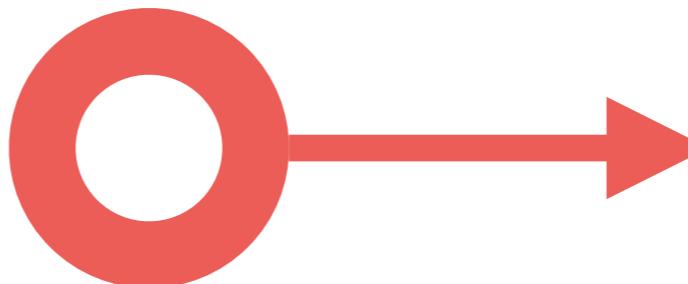


state-of-the-art
paraphrasing
for simplicity

★ In-depth adaptation of statistical MT (Joshua)

- tuning data (small parallel corpus)
- tunable metric (PRO pairwise ranking optimization)
- PPDB (**no large parallel corpus needed!**)
- simplification-specific features

System



state-of-the-art
paraphrasing
for simplicity

★ In-depth adaptation of statistical MT (Joshua)

- tuning data (small parallel corpus)
- tunable metric (PRO pairwise ranking optimization)
- PPDB (**no large parallel corpus needed!**)
- simplification-specific features

Lexical

applauded → praised

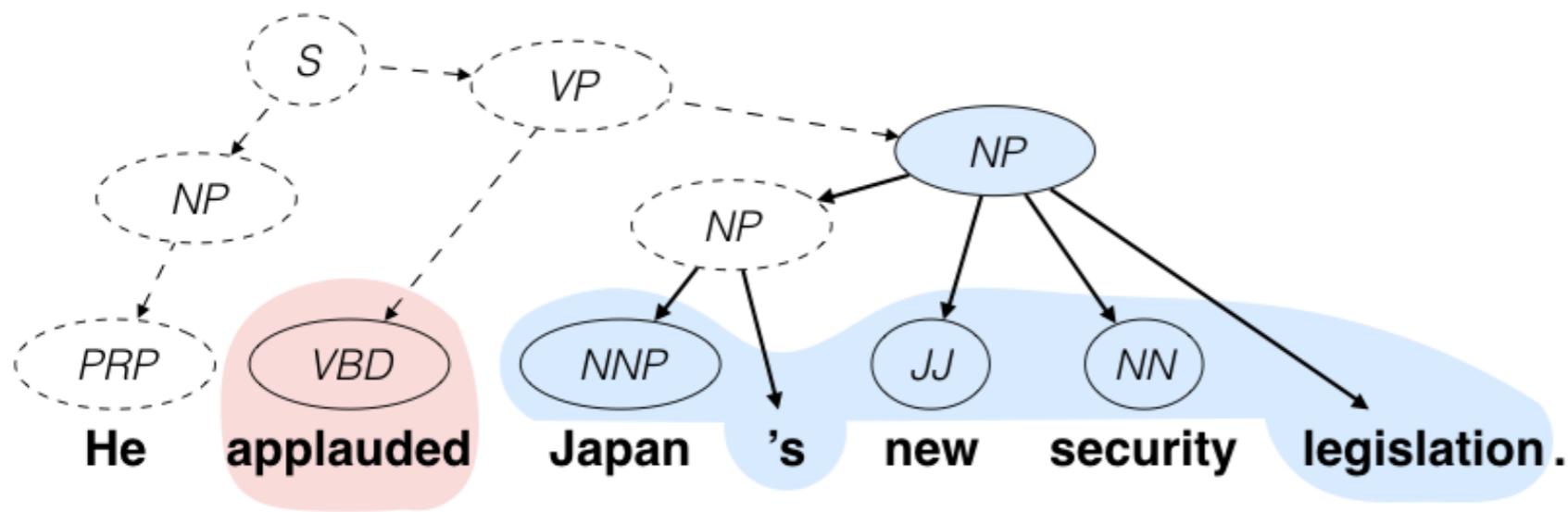
Phrasal

generally acknowledged → widely accepted

Syntactic

NNP's JJ legislation → the JJ law of NNP

Input:



PPDB (4+ million rules)

Lexical

applauded → praised

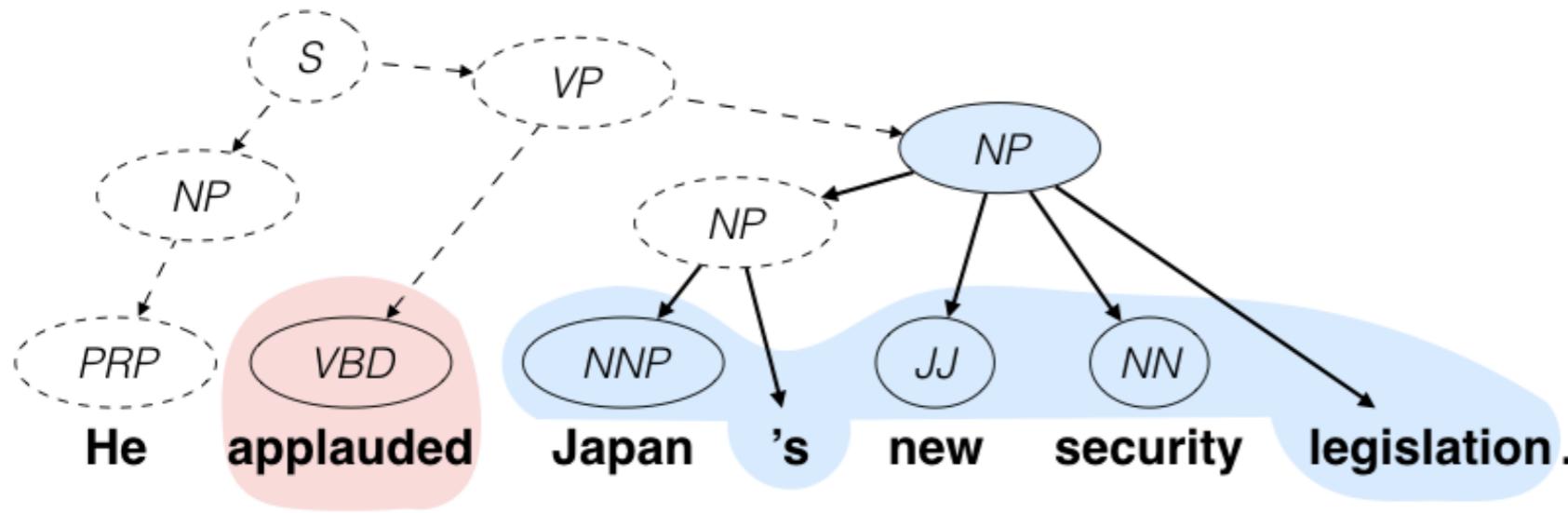
Phrasal

generally acknowledged → widely accepted

Syntactic

NNP's JJ legislation → the JJ law of NNP

Input:



PPDB (4+ million rules)

Lexical

applauded → praised

Phrasal

generally acknowledged → widely accepted

Syntactic

NNP's JJ legislation → the JJ law of NNP

Candidate Simplifications:

He saluted Japan 's new security law .

SARI = 0.21

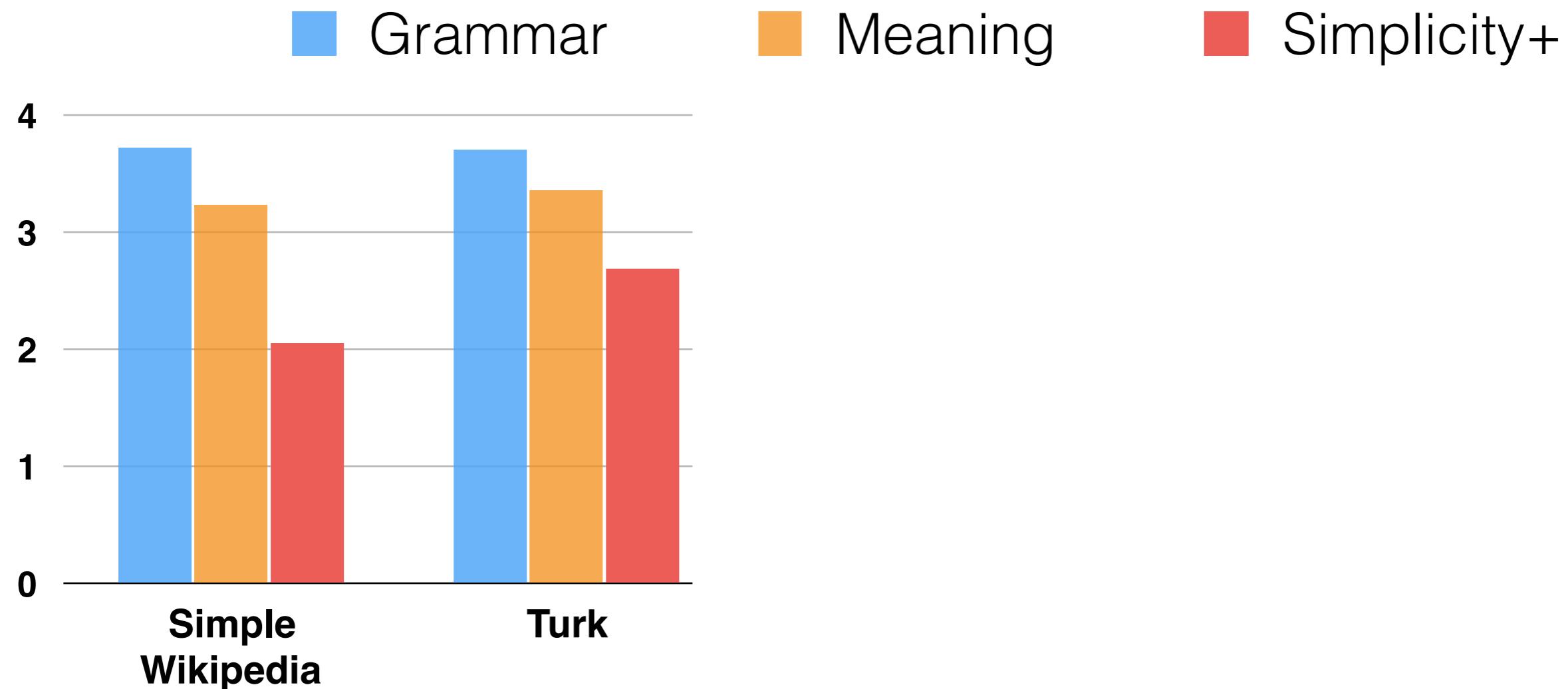
He praised the new security law of Japan .

SARI = 0.33

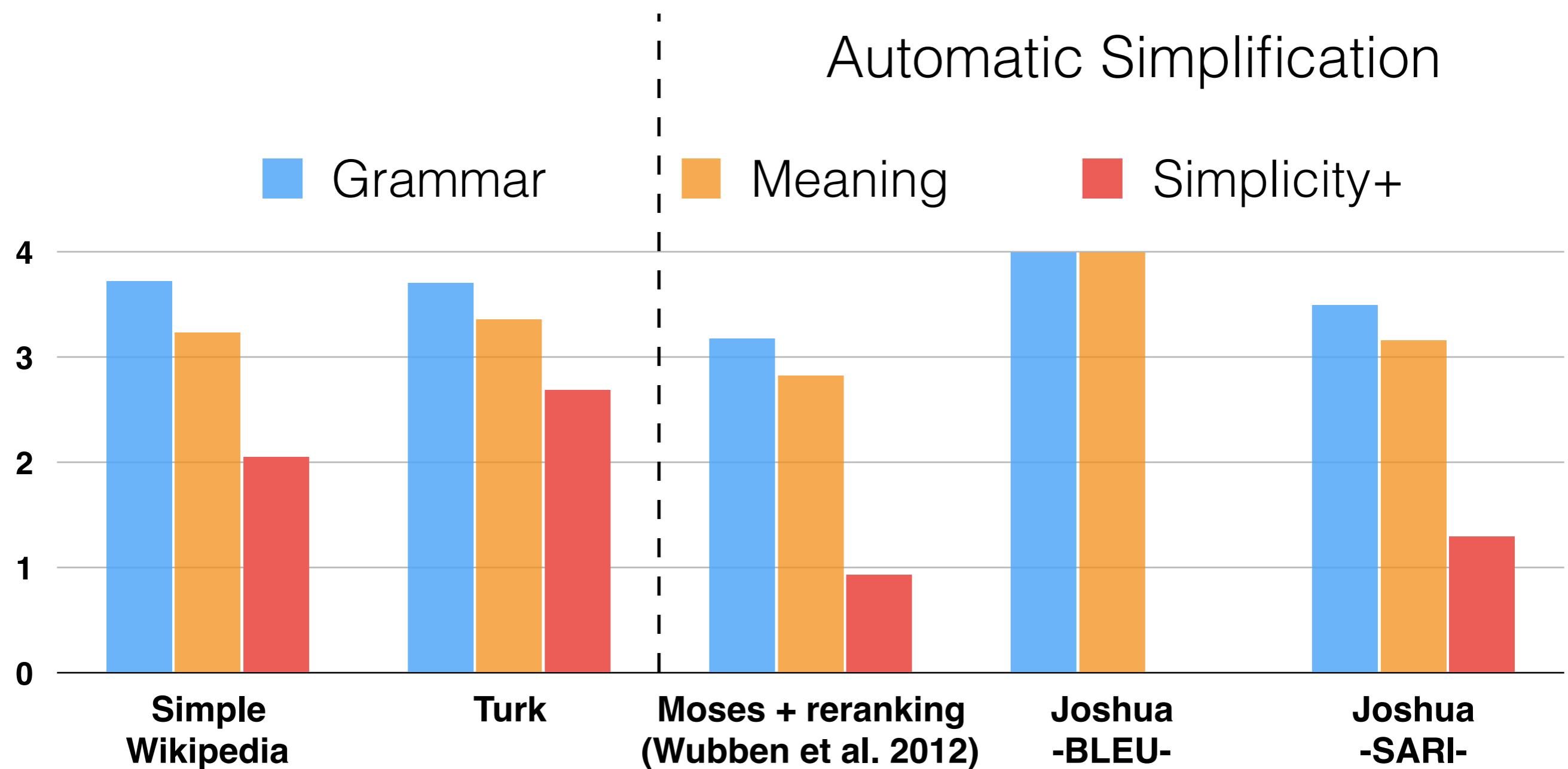
He praised Japanese new security law .

SARI = 0.15

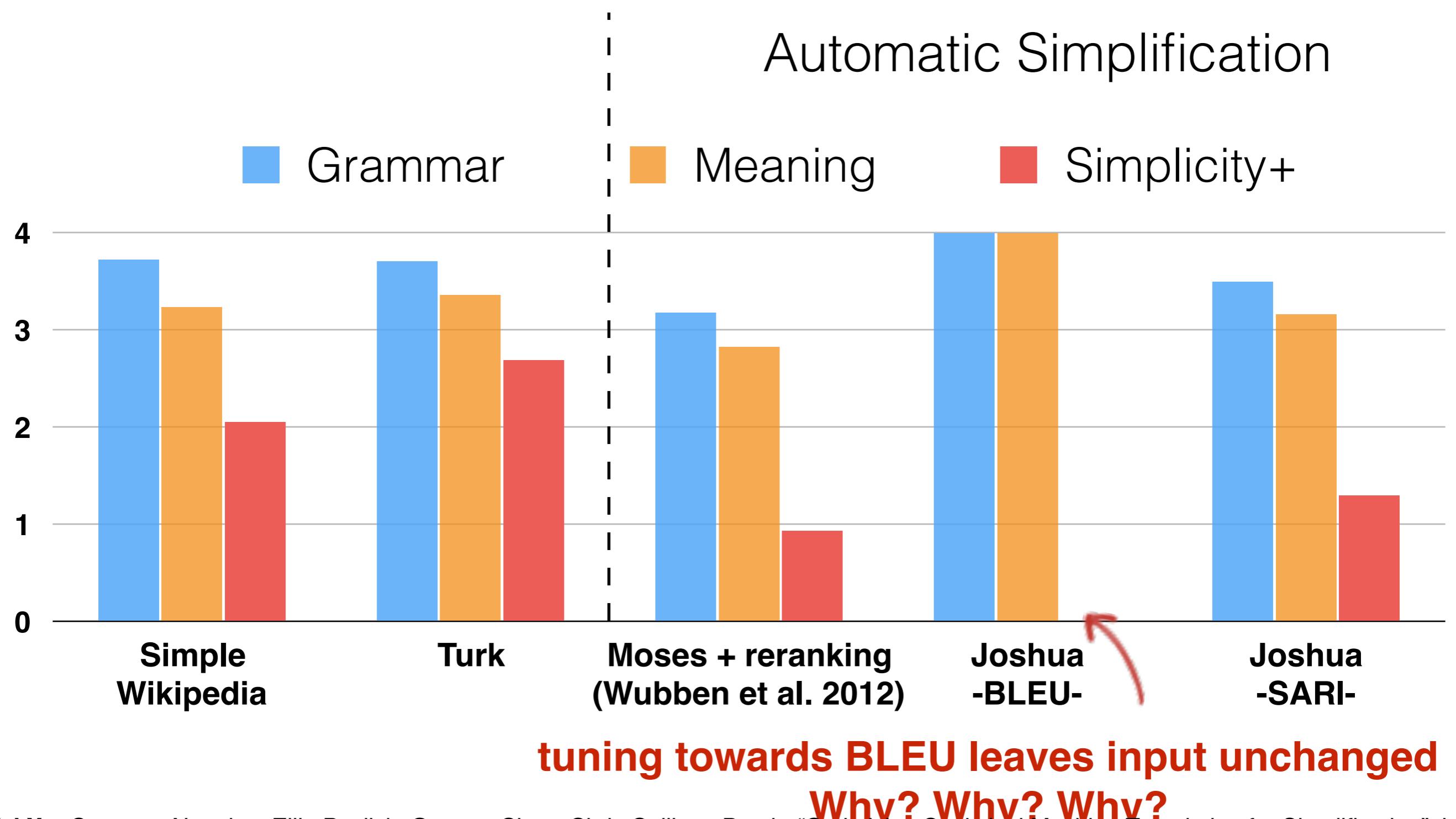
Human Evaluation



Human Evaluation



Human Evaluation



Automatic Metrics

(Spearman's rho correlation w/ human ratings)

	Grammar	Meaning	Simplicity
BLEU	0.59	0.70	0.15
FKBLEU	0.35	0.41	0.24
SARI	0.34	0.40	0.34

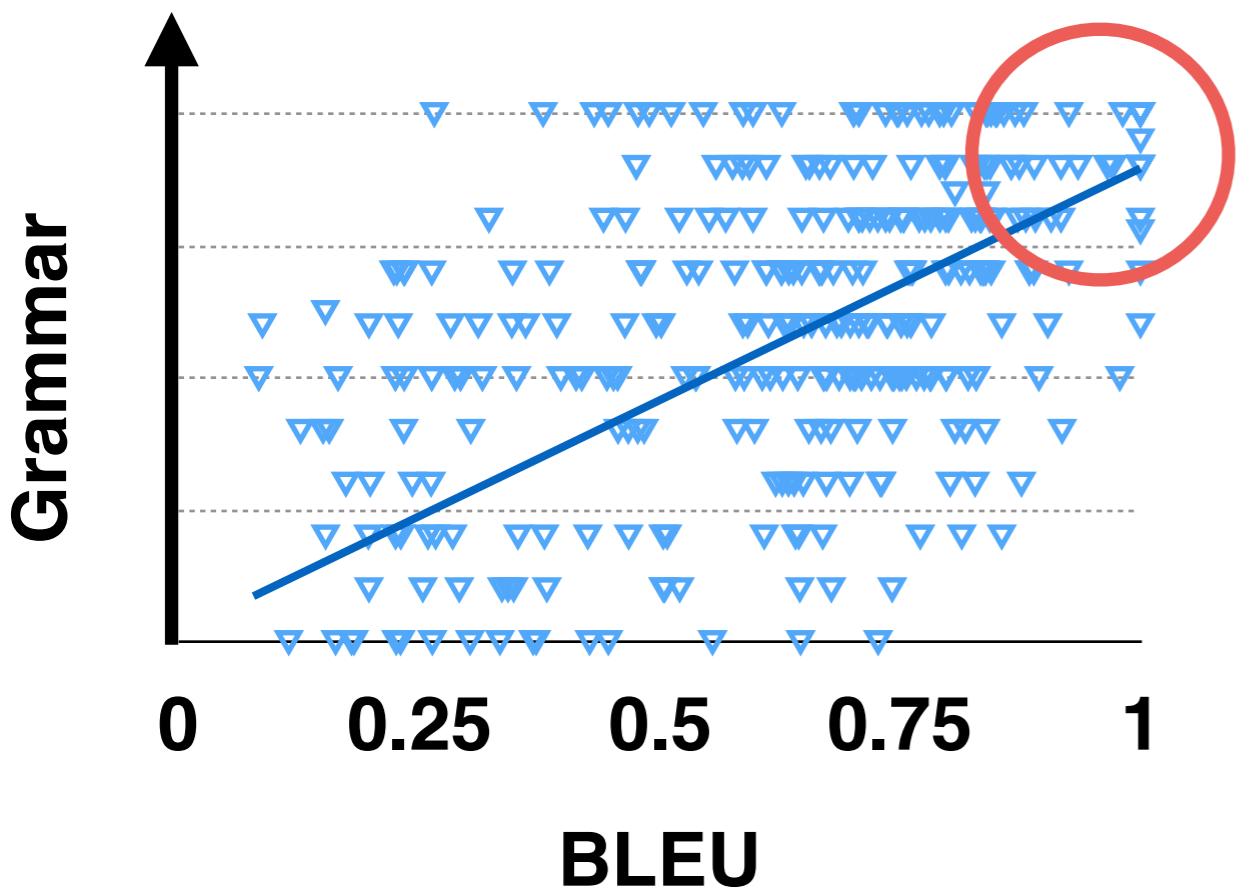
Automatic Metrics

(Spearman's rho correlation w/ human ratings)

	Grammar	Meaning	Simplicity
BLEU	0.59	0.70	0.15
FKBLEU	0.35	0.41	0.24
SARI	0.34	0.40	0.34

**this is not necessarily a bad thing
Why? Why? Why?**

BLEU correlates w/ Grammar

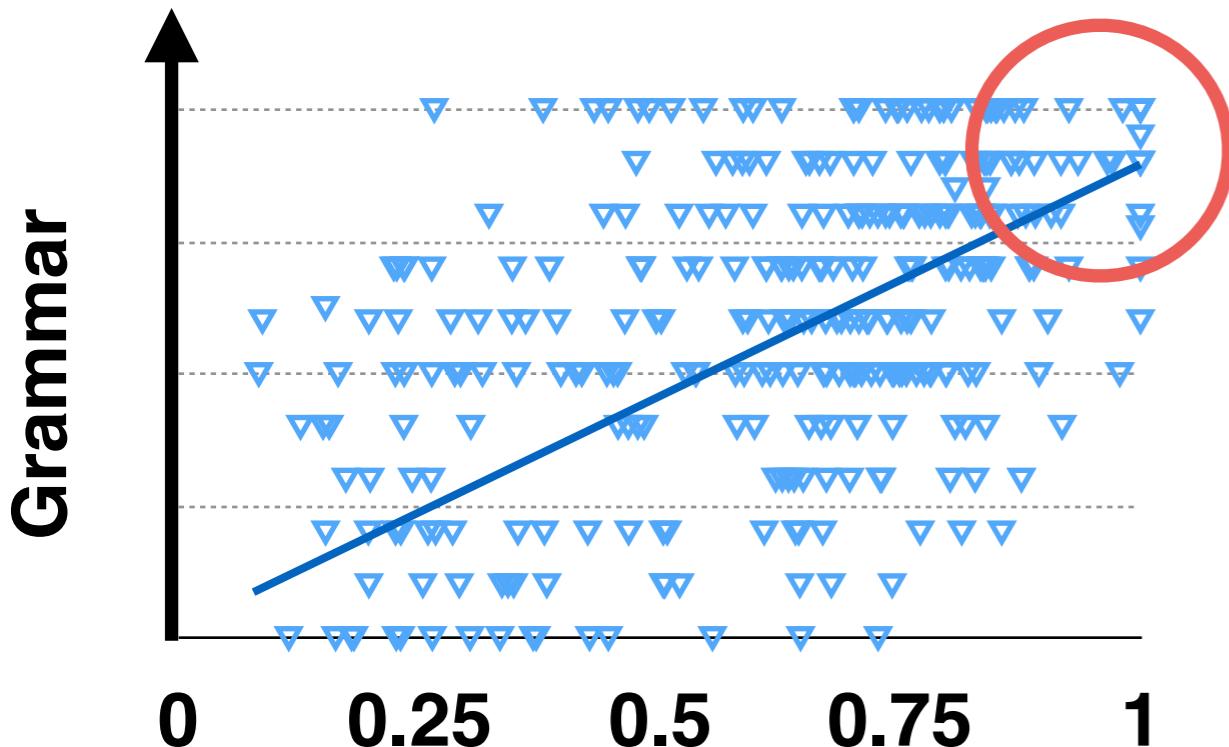


high Grammar high BLEU

If a system leaves input unchanged,
it gets
perfect Grammar
perfect Meaning
and very high BLEU score.

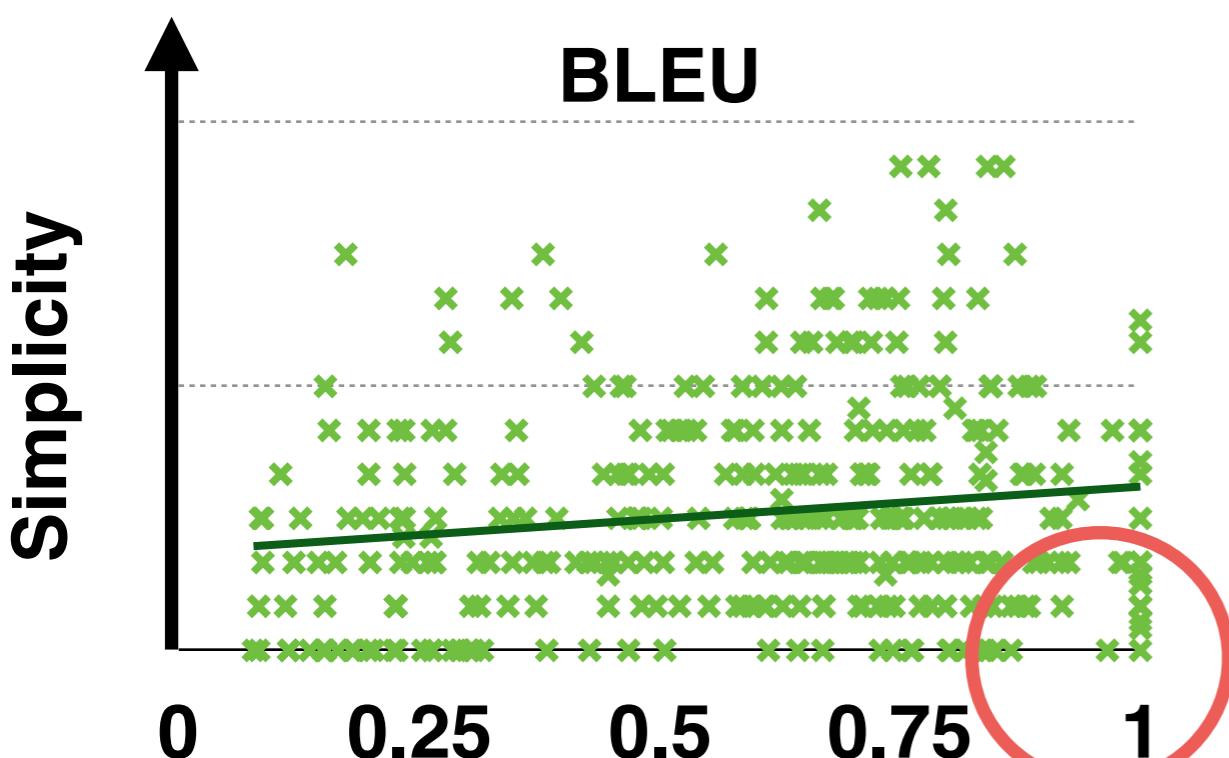
**because there are multiple references
(unlike bilingual translation)**

BLEU is not for Simplification



high Grammar high BLEU

If a system leaves input unchanged,
it gets
perfect Grammar
perfect Meaning
and very high BLEU score.

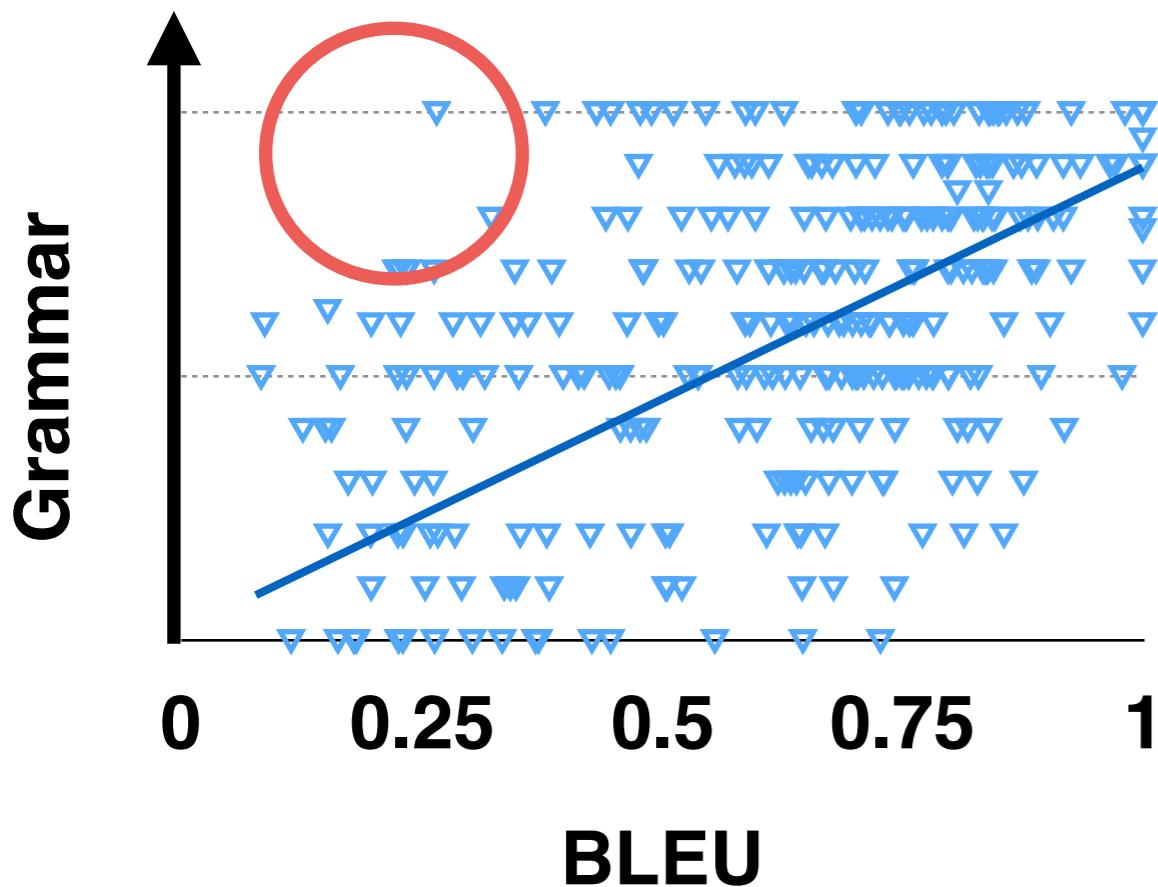


But it is not good simplification!

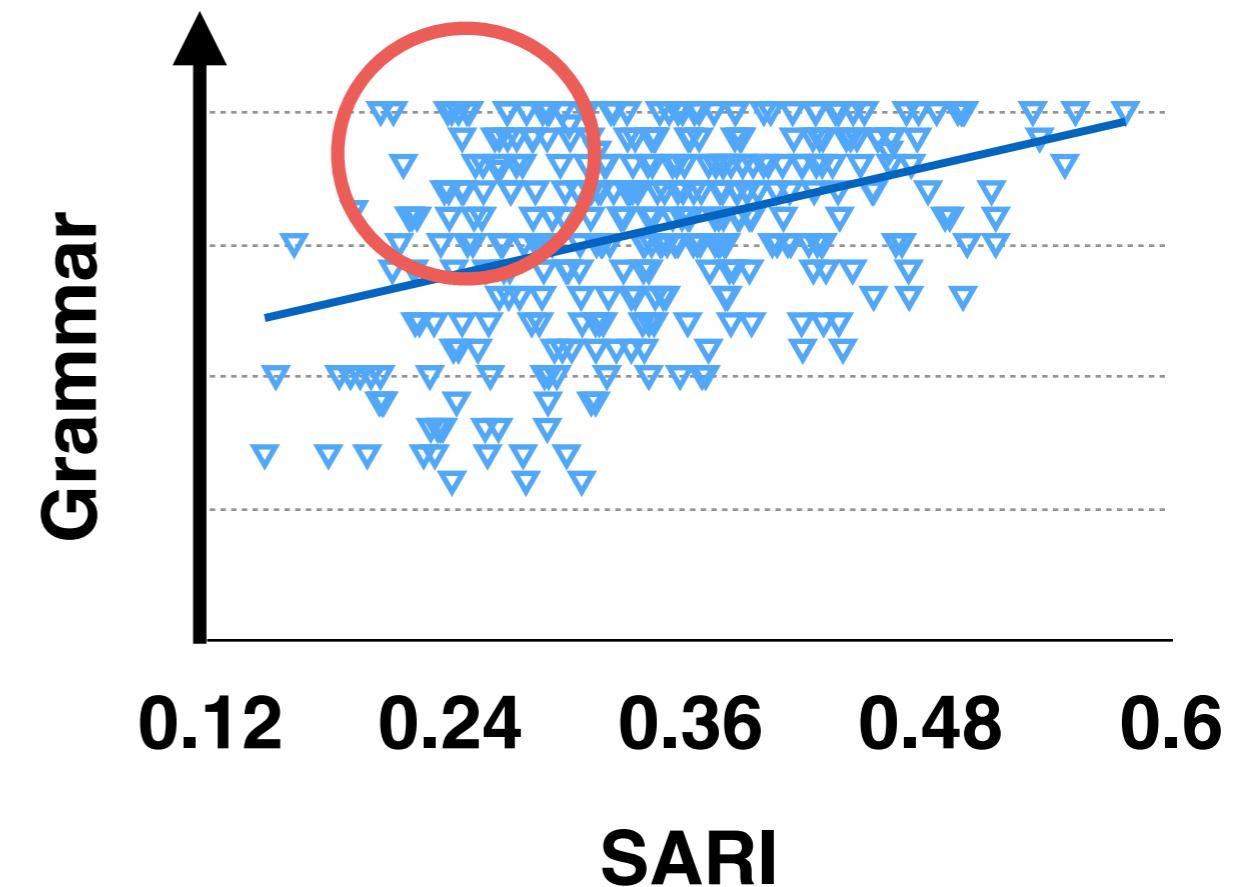
low Simplicity high BLEU

SARI is way better than BLEU

low BLEU high Grammar

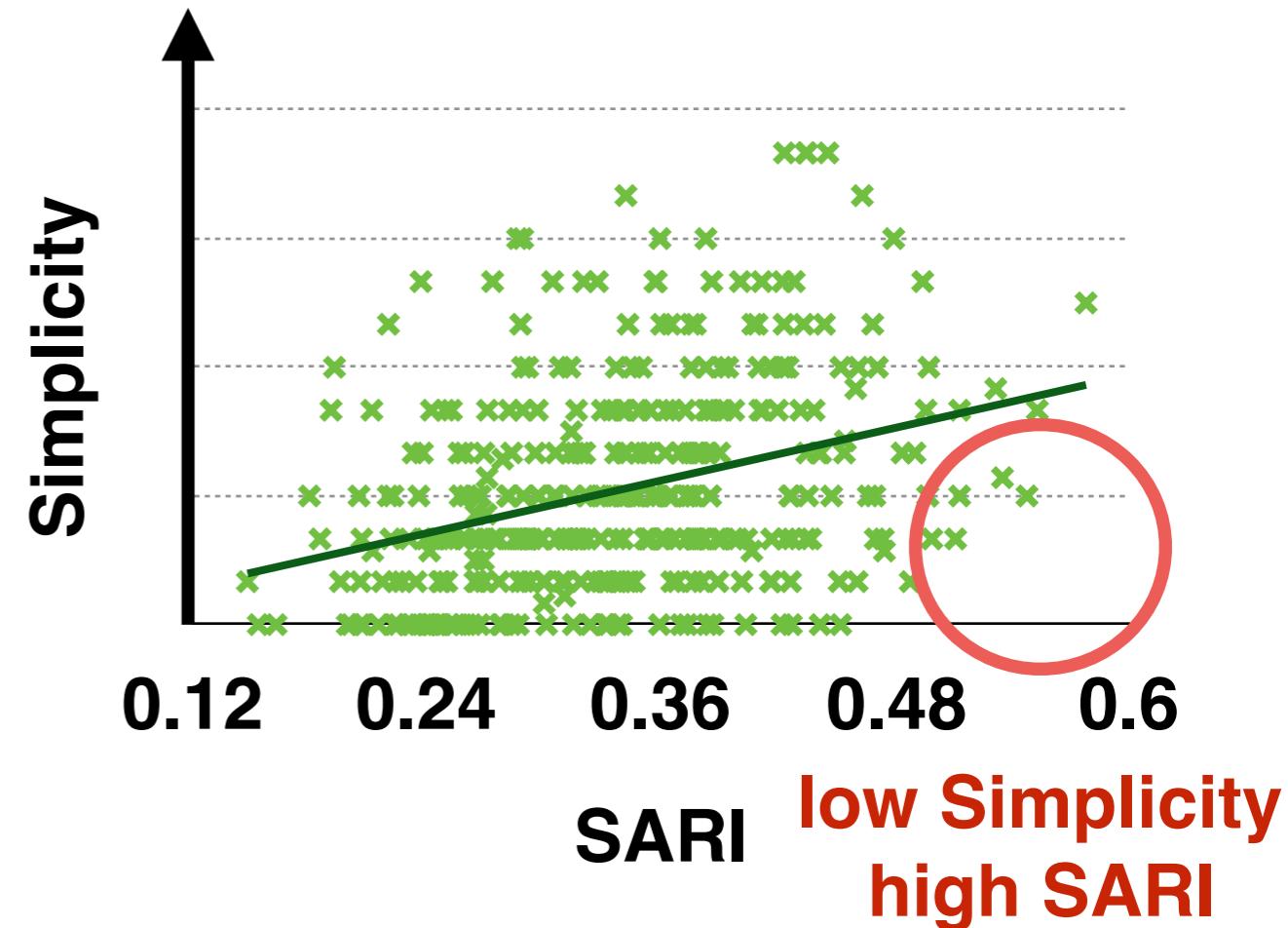
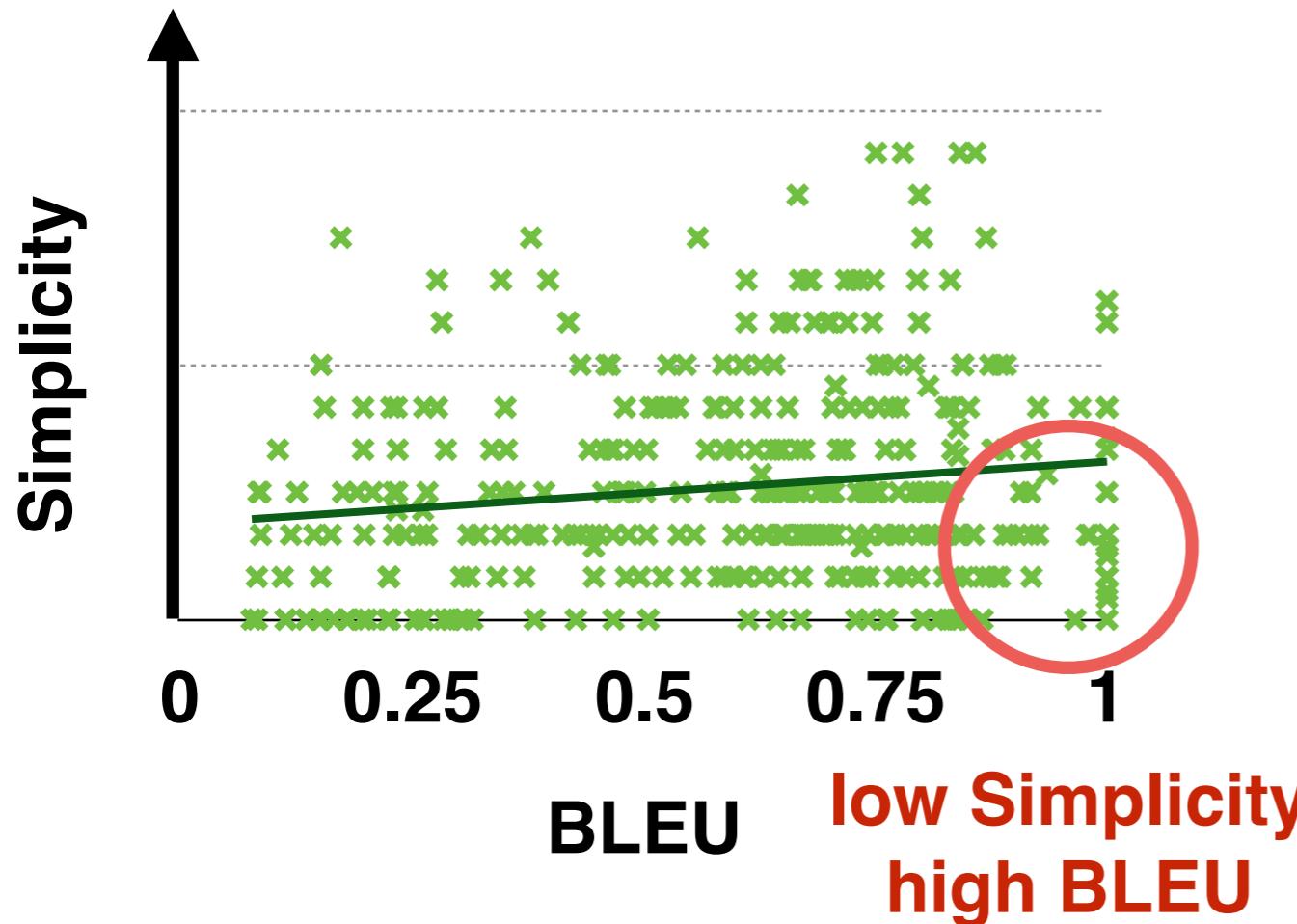


low SARI high Grammar



**high Grammar
could be bad simplification**

SARI is way better than BLEU



**low Simplicity
must be bad simplification**

Machine Translation

	Bilingual	(Paraphrase =) Monolingual
naturally available parallel text	a lot	much less
sensitive to error	less	more
objective	straightforward	sophisticated
standard metric	BLEU etc.	SARI and ?

Thank you

**Sponsor: National Science Foundation
Data and code available at <https://cocoxu.github.io/>**

Questions?

**I am looking for PhD students and post-docs.
Email me! Wei Xu @ Ohio State University**

Newsela Dataset (TACL '15)

NEWSLEA

WAR & PEACE SCIENCE KIDS MONEY HEALTH

SCIENCE 1738

Archaeologist may have found remains of ancient Egyptian Queen Nefertiti

By Robert Gebelhoff, Washington Post. 08.17.15



The 3,330-year-old bust of Nefertiti sits in an exhibition in the Kulturforum in Berlin, Germany, March 1, 2005. Photo: AP/Herbert Knosowski

Nefertiti — she's an ancient Egyptian queen and the source of a fantastic mystery regarding the iconic remnants of long-lost royalty.

For decades, archaeologists have speculated on the location of the queen's remains, the last royal mummy missing from the dynasty of the famous King Tutankhamun, better known as King Tut. But now, an archaeologist claims that he has found her

MAX
1140L
960L
720L
420L
 WRITE
 QUIZ

NEWSLEA

WAR & PEACE SCIENCE KIDS MONEY LAW HEALTH

SCIENCE 1738

Mystery of ancient Egypt solved? Tomb of queen may be hidden near King Tut

By Washington Post, adapted by Newsela staff 08.17.15



The 3,330-year-old bust of Nefertiti sits in an exhibition in the Kulturforum in Berlin, Germany, March 1, 2005. Photo: AP/Herbert Knosowski

The ancient Egyptian Queen Nefertiti has long been at the center of a mystery. For years, archaeologists have wondered where her tomb might be hidden. Nefertiti belonged to the family line of the famous King Tutankhamun, better known as King Tut. Indeed, some believe she was Tut's mother. While the other royals in her line are

1140L
960L
720L
420L
 WRITE
 QUIZ

every article at **5 levels** of simplification
total 50k+ sentences, written by trained editors, comes with comprehension quizzes