



Problems in Current Text Simplification Research

Wei Xu

Chris Callison-Burch

Courtney Napoles

UPenn

UPenn

JHU

What is Text Simplification

What is Text Simplification

INPUT *Applesauce is a puree made of apples.*

What is Text Simplification

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

What is Text Simplification

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

What is Text Simplification

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

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

What is Text Simplification



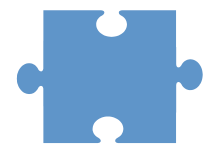
INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

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

What is Text Simplification



paraphrasing



deletion

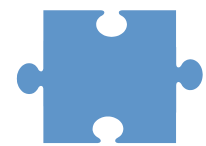
INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

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

What is Text Simplification



paraphrasing



deletion



splitting

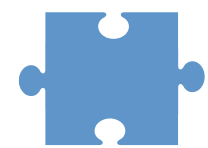
INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

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

What is Text Simplification



paraphrasing



deletion



splitting

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*



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

Goal of Text Simplification

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

Goal of Text Simplification

grammaticality

INPUT *Applesauce is a puree made of apples.*

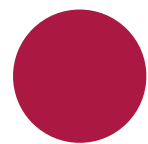
OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

Goal of Text Simplification



grammaticality



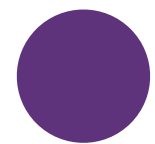
**meaning
preservation**

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

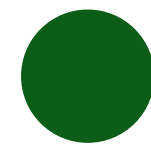
Goal of Text Simplification



grammaticality



**meaning
preservation**



simplicity

INPUT *Applesauce is a puree made of apples.*

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

Goal of Text Simplification

 **grammaticality**

 **meaning
preservation**

 **simplicity**

INPUT *Applesauce is a puree made of apples.*

Human Evaluation

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

Goal of Text Simplification

 **grammaticality**

 **meaning
preservation**

 **simplicity**

INPUT *Applesauce is a puree made of apples.*

Human Evaluation

OUT-1 *Applesauce is a soft paste.*

OUT-2 *Applesauce is a paste. It is made of apples.*

(no reliable automatic evaluation yet)

Brief History of Sentence Simplification

Brief History of Sentence Simplification

rule-based

- 1997 Chandrasekar & Srinivas
- 1999 Dras (PhD thesis)
- 2000 Carroll, Minnen, Pearce, Canning, Devlin
- 2002 Canning (PhD thesis)
- 2004 Siddharthan (PhD thesis)

Brief History of Sentence Simplification

rule-based

Simple English
WIKIPEDIA



- 1997 Chandrasekar & Srinivas
- 1999 Dras (PhD thesis)
- 2000 Carroll, Minnen, Pearce, Canning, Devlin
- 2002 Canning (PhD thesis)
- 2004 Siddharthan (PhD thesis)
- 2010 Zhu, Bernhard, Gurevych**

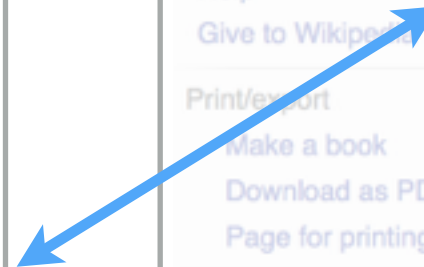
Parallel Wikipedia Corpus



The screenshot shows the standard Wikipedia article for "Apple sauce". The title is "Apple sauce" and the subtitle is "From Wikipedia, the free encyclopedia". The article text is partially visible, with a blue box highlighting the first sentence: "Apple sauce or applesauce is a purée made of apples. It can be made with peeled and/or unpeeled and allspice). Flavorings or sweeteners such as sugar or honey are also commonly added. Apple sauce is inexpensive and is widely used in the United Kingdom, North America and some European countries.^[1]"



The screenshot shows the Simple English Wikipedia article for "Applesauce". The title is "Applesauce" and the subtitle is "From Wikipedia, the free encyclopedia". The article text is partially visible, with a blue box highlighting the first sentence: "Applesauce (or apple sauce) is a sauce that is made from stewed and mashed apples. Peeled or cinnamon can be used. Sugar or high fructose corn syrup is sometimes added to the applesauce to sweeten it. Applesauce can be fine or coarse textured, and can include large pieces of apple. It is easy to make at home, and it is also sold already made in supermarkets or as a snack for children and people who have problems with eating solid food. Parents often feed it because teeth are not needed to eat it. It is sometimes used to fight diarrhoea, since it is high in dietary fibre."



Brief History of Sentence Simplification

rule-based

Simple English
WIKIPEDIA



- 1997 Chandrasekar & Srinivas
- 1999 Dras (PhD thesis)
- 2000 Carroll, Minnen, Pearce, Canning, Devlin
- 2002 Canning (PhD thesis)
- 2004 Siddharthan (PhD thesis)
- 2010 Zhu, Bernhard, Gurevych**

Brief History of Sentence Simplification

rule-based

Simple English
WIKIPEDIA



**machine
translation**

- 1997 Chandrasekar & Srinivas
- 1999 Dras (PhD thesis)
- 2000 Carroll, Minnen, Pearce, Canning, Devlin
- 2002 Canning (PhD thesis)
- 2004 Siddharthan (PhD thesis)
- 2010 **Zhu, Bernhard, Gurevych****
- 2011 Woodsend & Lapata
- 2011 Coster & Kauchak
- 2012 Wubben, van den Bosch, Krahmer
- 2014 Narayan & Gardent
- 2014 Siddharthan (Survey)
- 2014 Angrosh, Nomoto, Siddharthan
- 2014 Narayan (PhD thesis)

Brief History of Sentence Simplification

rule-based

Simple English
WIKIPEDIA



**machine
translation**

- 1997 Chandrasekar & Srinivas
- 1999 Dras (PhD thesis)
- 2000 Carroll, Minnen, Pearce, Canning, Devlin
- 2002 Canning (PhD thesis)
- 2004 Siddharthan (PhD thesis)
- 2010 Zhu, Bernhard, Gurevych**
- 2011 Woodsend & Lapata
- 2011 Coster & Kauchak
- 2012 Wubben, van den Bosch, Krahmer
- 2014 Narayan & Gardent
- 2014 Siddharthan (Survey)
- 2014 Angrosh, Nomoto, Siddharthan
- 2014 Narayan (PhD thesis)
- Now Xu, Callison-Burch, Napoles (Opinion)**

Problems in Simplification Research

- State-of-the-art evaluation is suboptimal. But we have been doing this in the past 5 years*.
- Simple Wikipedia data dominated in the past 5 years. But its quality was taken for granted. It limits the scope of research.

* (Angrosh et al. 2014) tried comprehension quiz

Why this is important?

Breakthrough on Sea

wind direction



a straight path
upwind ❌



a zigzag path
upwind ✅

- better understanding
- better **review**
- more diverse research
- better data and evaluation
- better model

Why this is important?

Breakthrough on Sea

wind direction



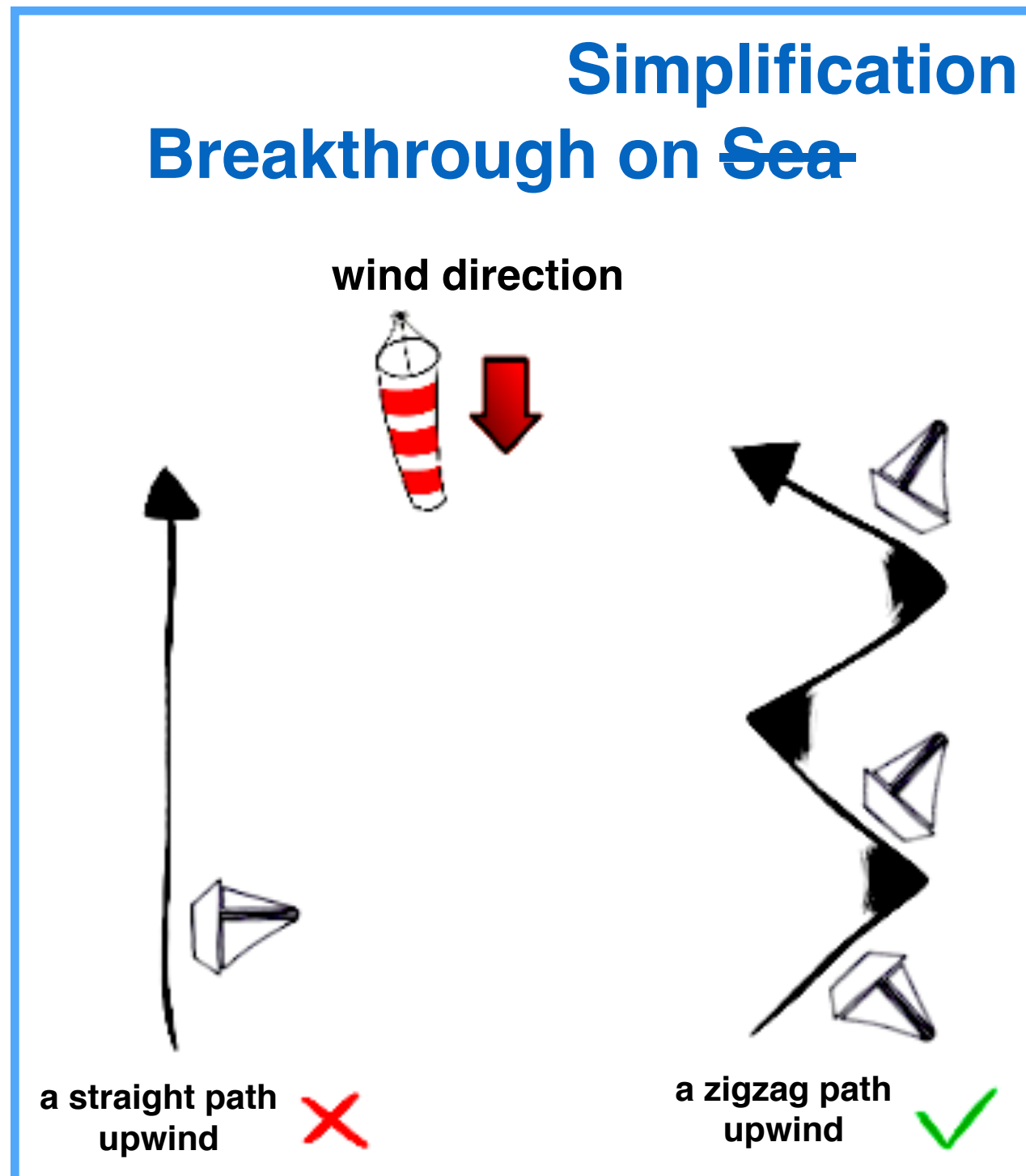
a straight path
upwind ❌



a zigzag path
upwind ✅

- better understanding
- better **review**
- more diverse research
- better data and evaluation
- better model

Why this is important?



- better understanding
- better **review**
- more diverse research
- better data and evaluation
- better model

“Recently, there have been several attempts at addressing the text simplification task as a monolingual translation problem ... However, they did not try to seek reasons for the success or the failure of their systems.”

—— Štajner, Béchara, Saggion (2015)

“Recently, there have been several attempts at addressing the text simplification task as a monolingual translation problem ... However, they did not try to seek reasons for the success or the failure of their systems.”

—— Štajner, Béchara, Saggion (2015)

WHY DID THIS HAPPEN?

- ① “state-of-the-art” competition
- ② not easy to do

Opinion #1

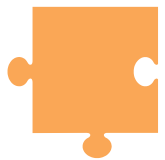
**Current evaluation doesn't
tell us what's going on.**

System Comparability

sub-systems



paraphrasing

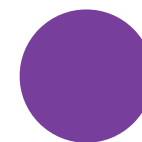


deletion

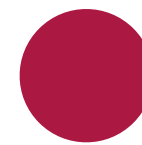


splitting

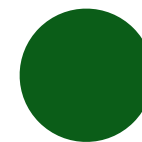
evaluation criteria



grammaticality



**meaning
preservation**



simplicity

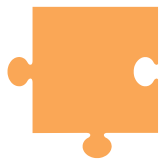
not easy to measure

System Comparability

sub-systems



paraphrasing



deletion

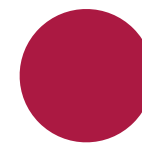


splitting

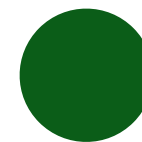
evaluation criteria



grammaticality



**meaning
preservation**



simplicity



not easy to measure

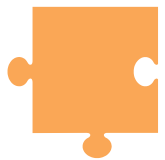
We need more controlled evaluation:

System Comparability

sub-systems



paraphrasing

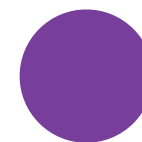


deletion

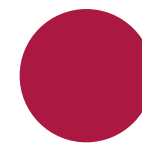


splitting

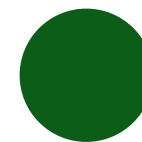
evaluation criteria



grammaticality



**meaning
preservation**



simplicity

not easy to measure

We need more controlled evaluation:

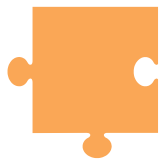
- evaluate sub-tasks separately

System Comparability

sub-systems



paraphrasing

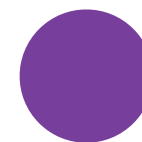


deletion

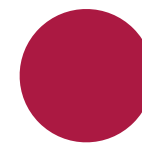


splitting

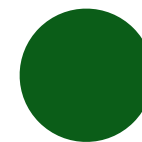
evaluation criteria



grammaticality



**meaning
preservation**



simplicity

not easy to measure

We need more controlled evaluation:

- evaluate sub-tasks separately
- target specific audience (e.g. 10-12 year old)

Opinion #2

Simple Wikipedia is
not that simple



Simple English
WIKIPEDIA

Page

Talk

Read

Change

More ▼

Search

Photolithography

From Wikipedia, the free encyclopedia

[Main page](#)

[Photolithography](#) is the combination of [photography](#) and [lithography](#)

Microphotolithography is the use of photolithography to transfer geometric shapes on a [photomask](#) to the surface of a semiconductor wafer for making [integrated circuits](#).

“Specific questions that need addressing are :

... we need to better understand the quality of Simple English Wikipedia, a resource that has been used to train many SMT based simplification systems...”

—— Advait Siddharthan (2014 Survey)

“Specific questions that need addressing are :

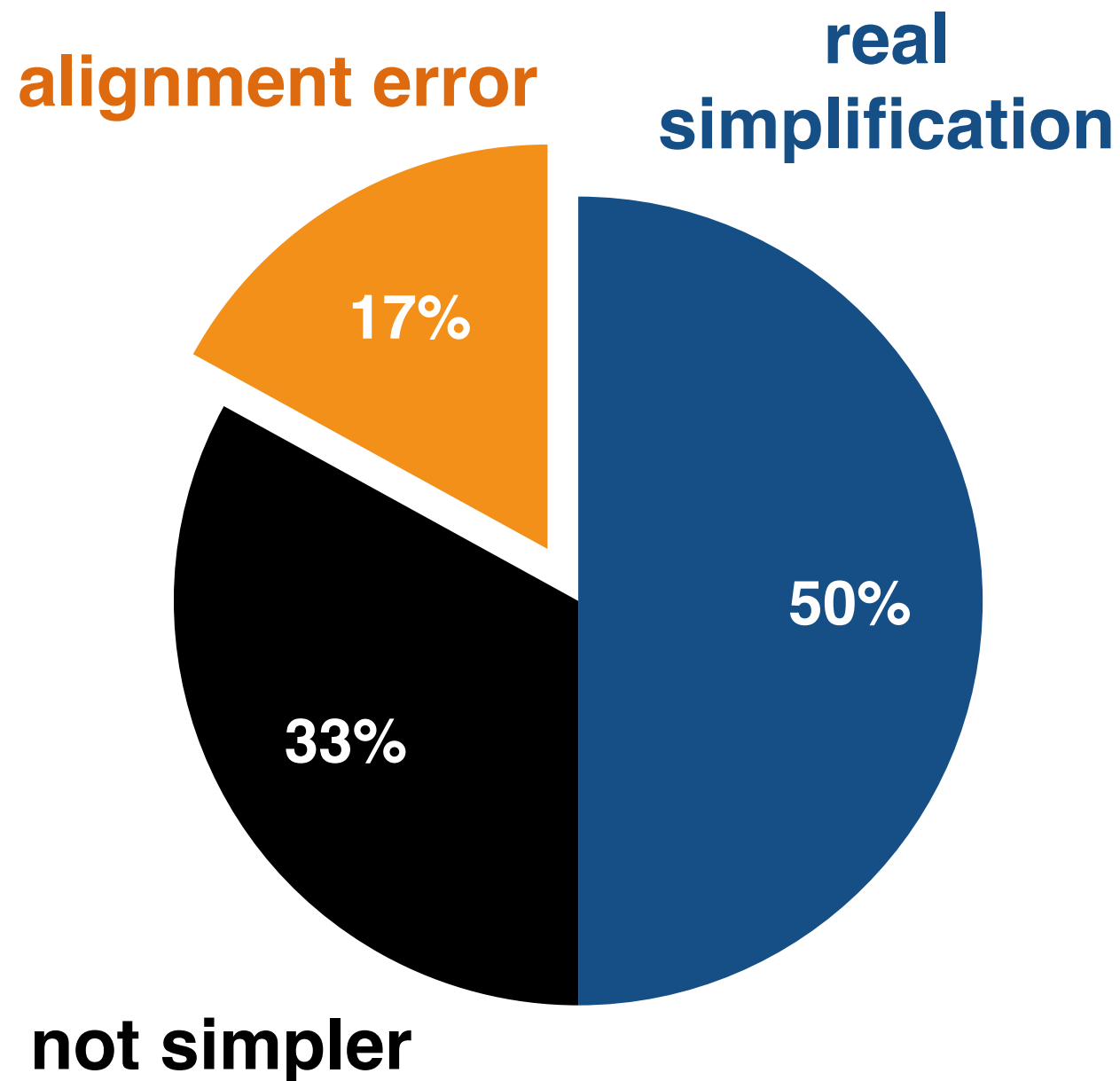
... we need to better understand the quality of Simple English Wikipedia, a resource that has been used to train many SMT based simplification systems...”

—— Advait Siddharthan (2014 Survey)

WHAT'S NEW?

We quantitatively and systematically answer this quest.

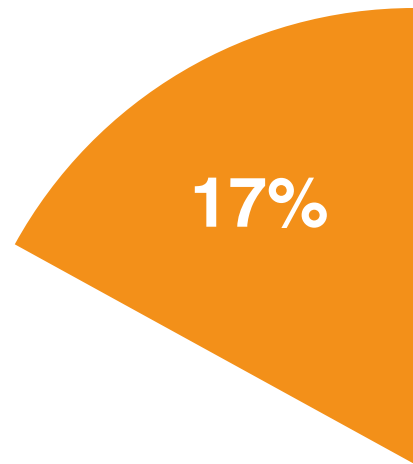
Quality of Parallel Wikipedia Corpus*



Inaccuracy in Parallel Wikipedia Corpus*

alignment error

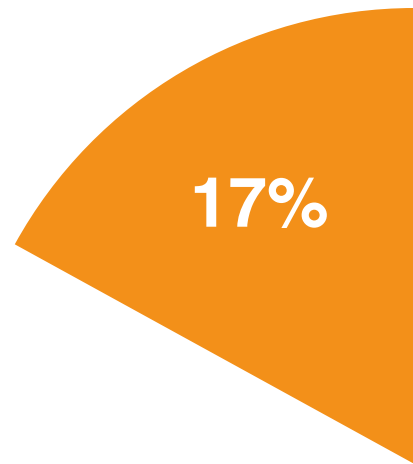
(two sentences have different meaning)



Inaccuracy in Parallel Wikipedia Corpus*

alignment error

(two sentences have different meaning)

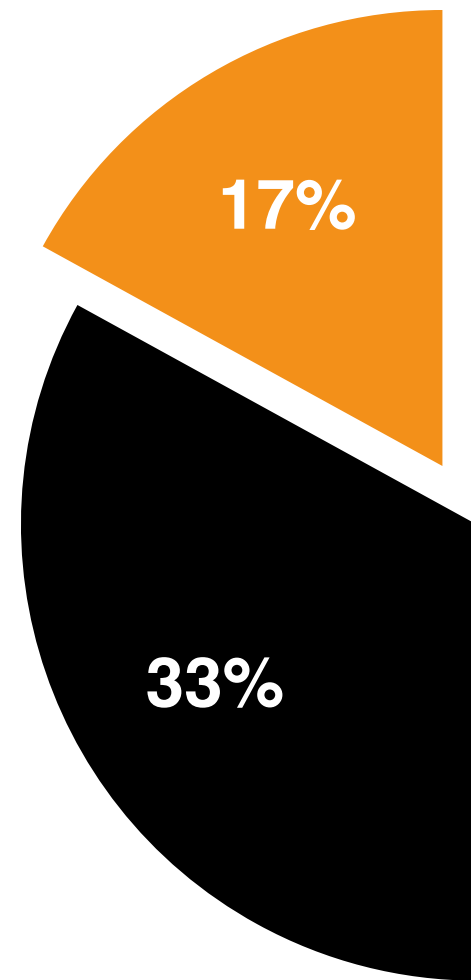


Best automatic sentence alignment gets about **0.7 F1** score (Hwang et al. 2015)

Inadequacy in Parallel Wikipedia Corpus*

alignment error

(two sentences have different meaning)



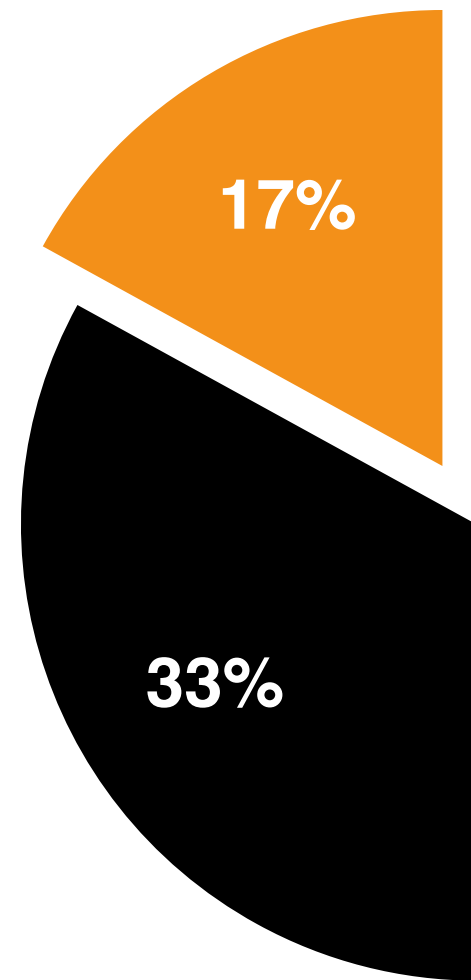
Best automatic sentence alignment gets about **0.7 F1** score (Hwang et al. 2015)

not simpler

Inadequacy in Parallel Wikipedia Corpus*

alignment error

(two sentences have different meaning)



17%

33%

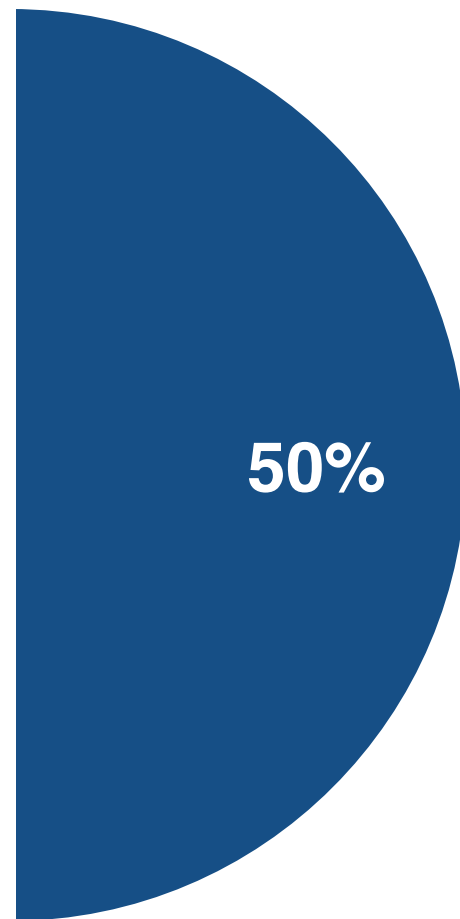
Best automatic sentence alignment gets about **0.7 F1** score (Hwang et al. 2015)

Sentences can have similar meaning but not simplification

not simpler

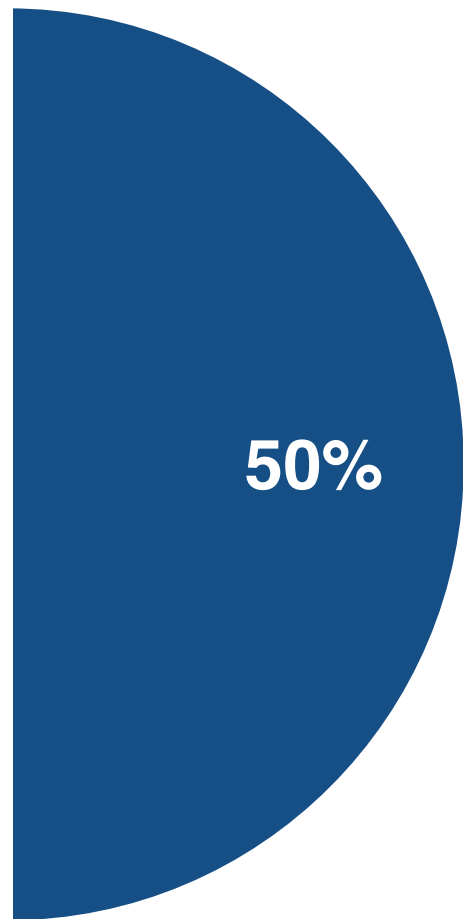
Inadequacy in Parallel Wikipedia Corpus*

**real
simplification** (aligned and simpler)



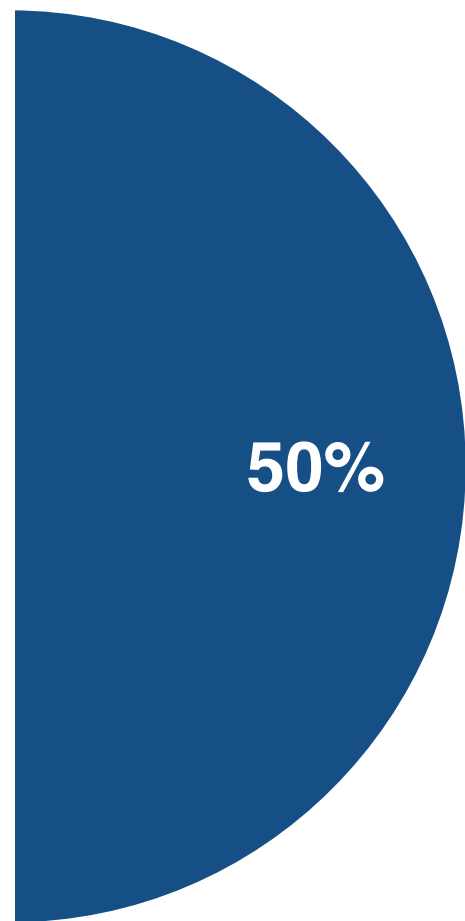
Inadequacy in Parallel Wikipedia Corpus*

**real
simplification** (aligned and simpler?)



Inadequacy in Parallel Wikipedia Corpus*

**real
simplification** (aligned and simpler?)



Some sentences are simpler by only one word while the rest of sentence is still complex

Issues with Parallel Wikipedia Corpus

Issues with Parallel Wikipedia Corpus

- suboptimal for estimating “translation” probabilities

Issues with Parallel Wikipedia Corpus

- suboptimal for estimating “translation” probabilities
- suboptimal for developing automatic metrics

Issues with Parallel Wikipedia Corpus

- suboptimal for estimating “translation” probabilities
- suboptimal for developing automatic metrics
- suboptimal for tuning MT system

Issues with Parallel Wikipedia Corpus

- suboptimal for estimating “translation” probabilities
- suboptimal for developing automatic metrics
- suboptimal for tuning MT system
- unsuitable for document-level simplification

Opinion #3

New data **can** help

Newsela Dataset

NEWSELA

WAR & PEACE SCIENCE KIDS MONEY LAW HEALTH

SCIENCE 1738 SHARE

Archaeologist may have found remains of ancient Egyptian Queen Nefertiti

By Robert Gebelhoff, Washington Post. 08.17.15

Grade 5 Word Count



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

NEWSELA

WAR & PEACE SCIENCE KIDS MONEY LAW HEALTH

SCIENCE 1738 SHARE

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

By Washington Post, adapted by Newsela staff. 08.17.15

Grade 5 Word Count



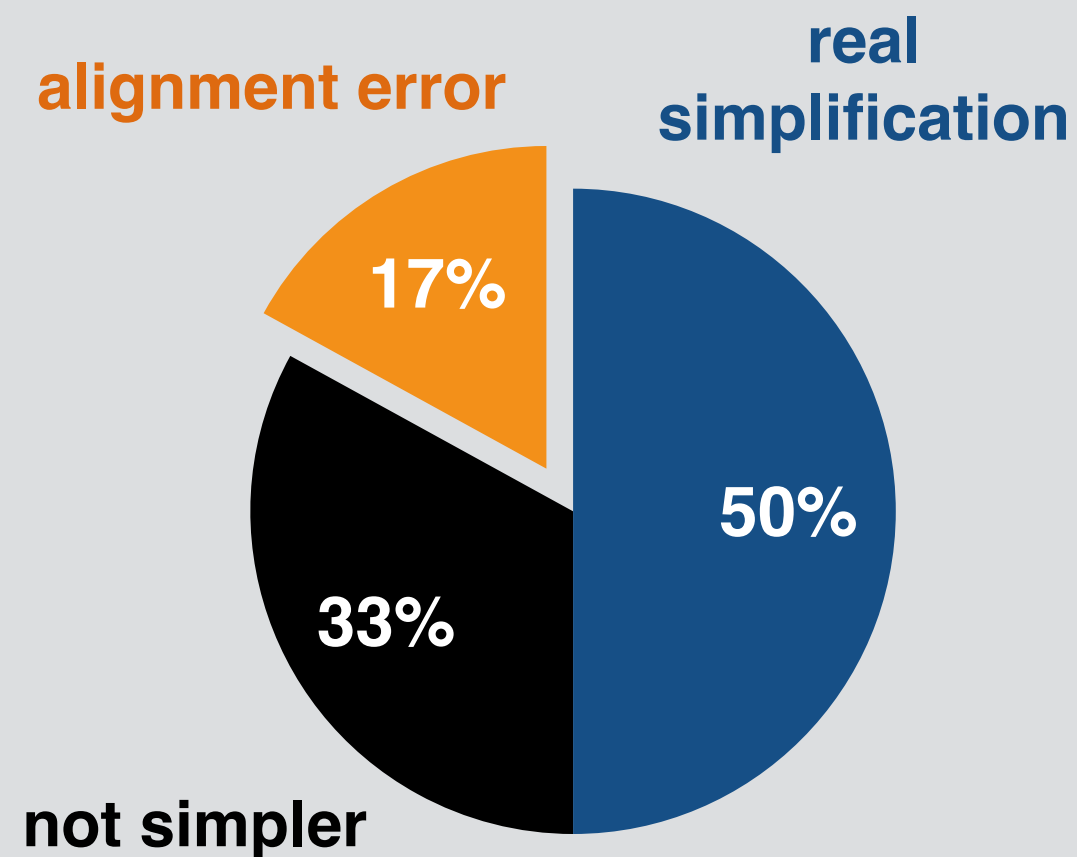
The 3,330-year-old bust of Nefertiti sits in an exhibition in the Kulturforum in Berlin, Germany, March 1, 2005. Photo: AP/Herbert Knosowski

The ancient Egyptian Queen Nefertiti has long been at the center of a mystery.

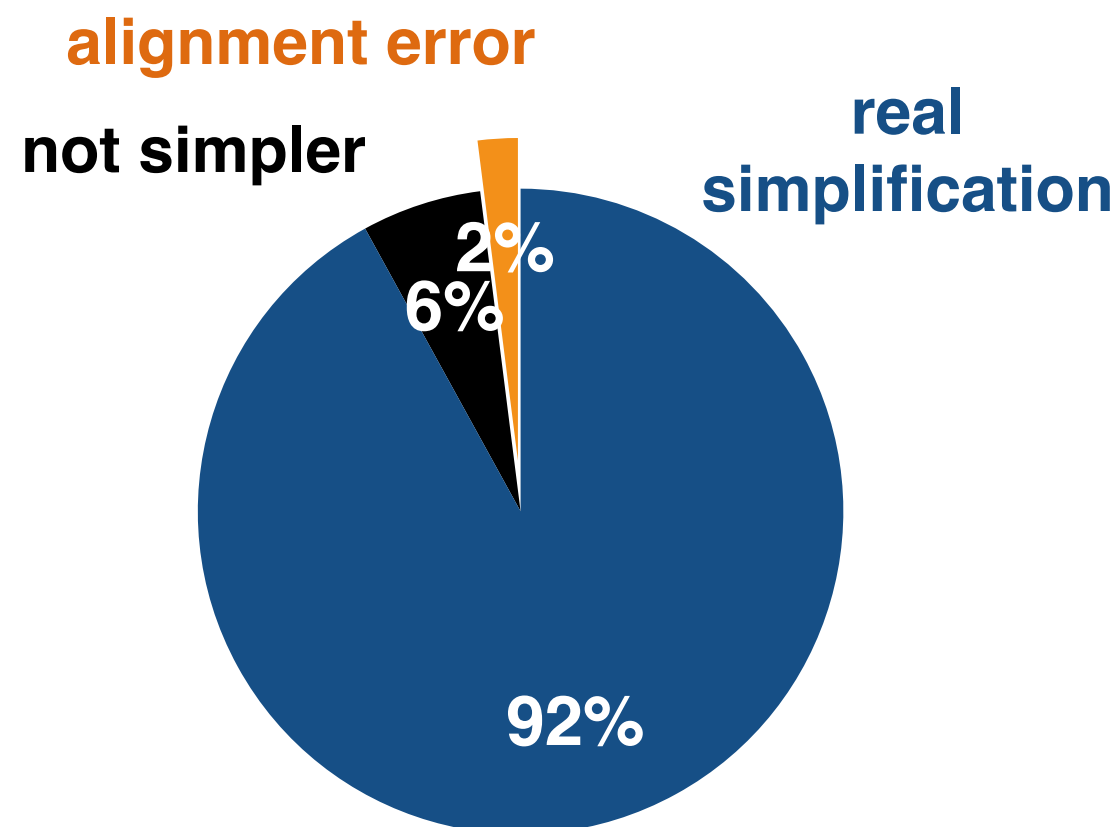
For years, archaeologists have wondered where her tomb might be hidden. Nefertiti belonged to the family line of the famous King Tutankhamun, better known as King Tut. Indeed, some believe she was Tut's mother. While the other royals in her line are

every article at **5 levels** of simplification
written by trained editors, comes with comprehension quizzes

Wikipedia*



Newsela

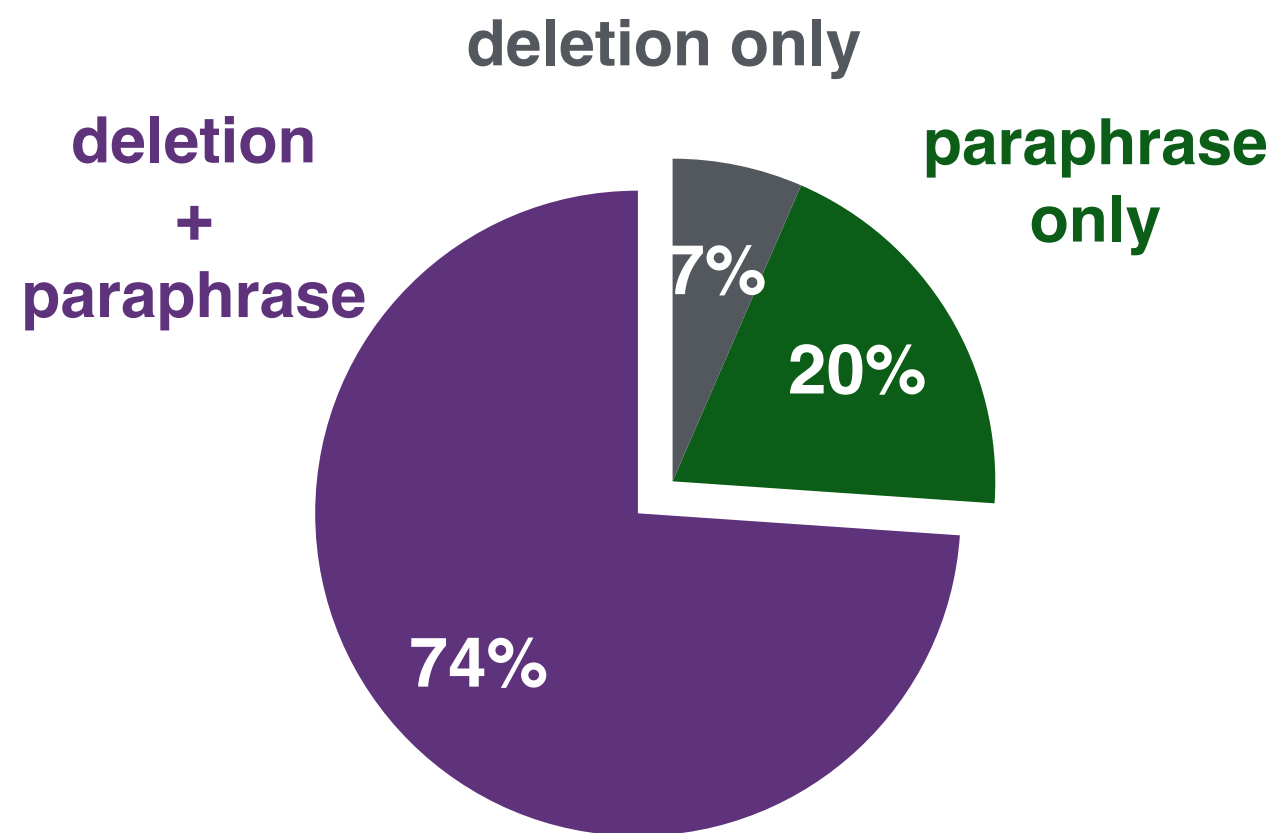
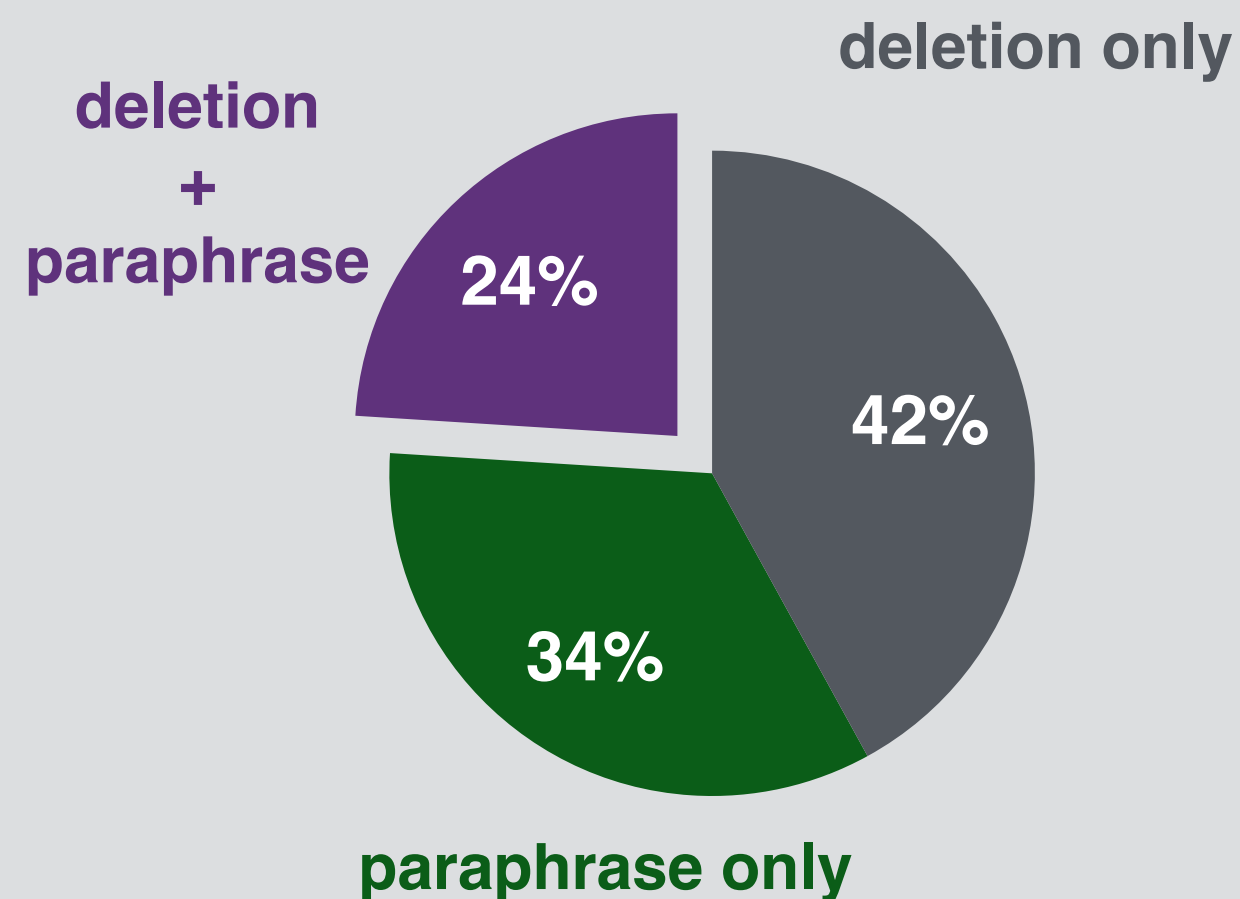


manual inspection of aligned sentence pairs

Wikipedia*

Newsela

Good simplification needs more paraphrasing.

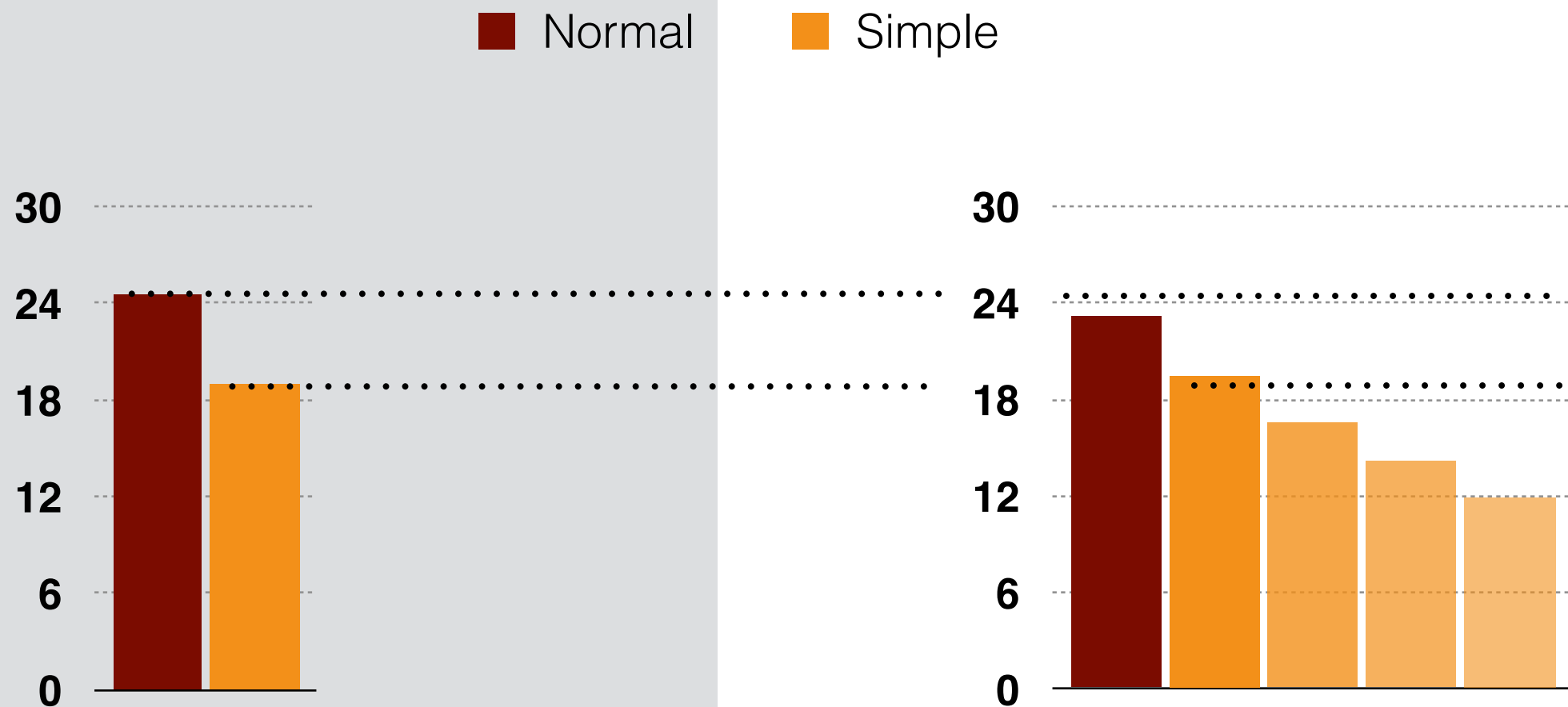


degree of paraphrasing

Wikipedia*

Newsela

Good simplification could be much shorter.



sentence length (#words)
see syntax analysis in the paper

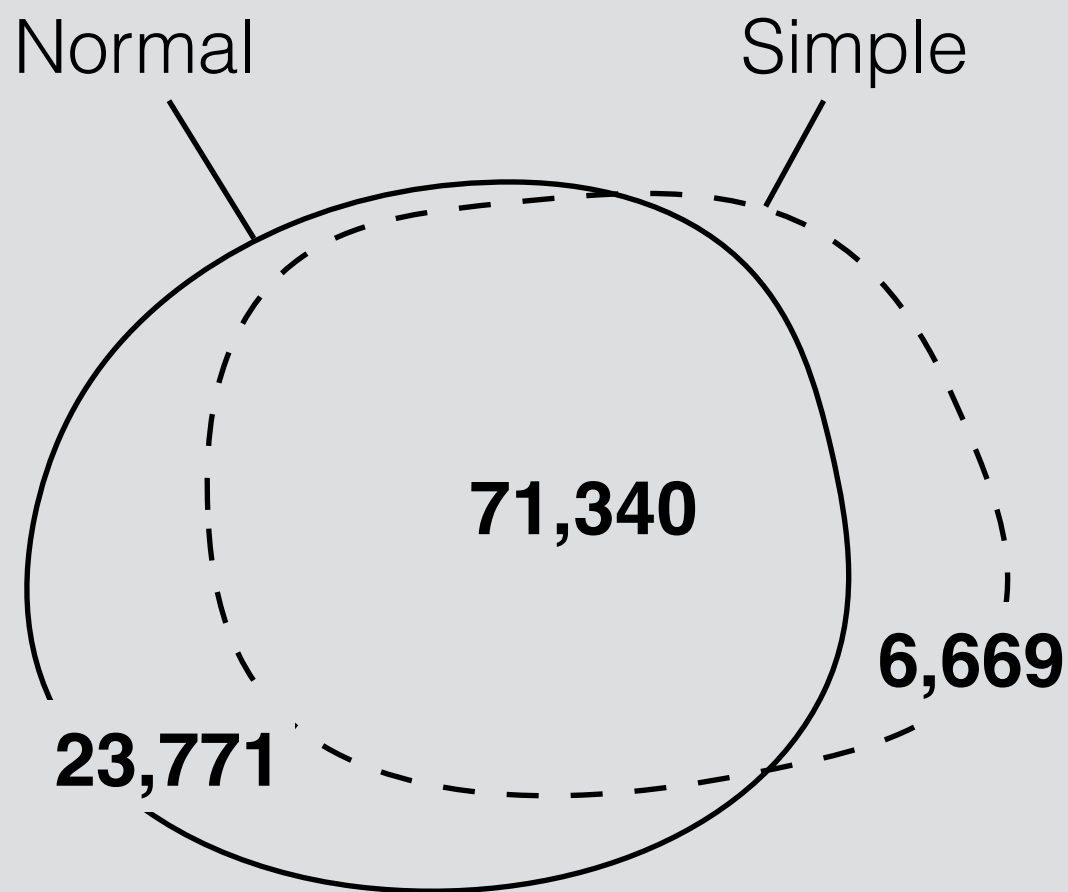
Wikipedia*

(total 2.6 million tokens)

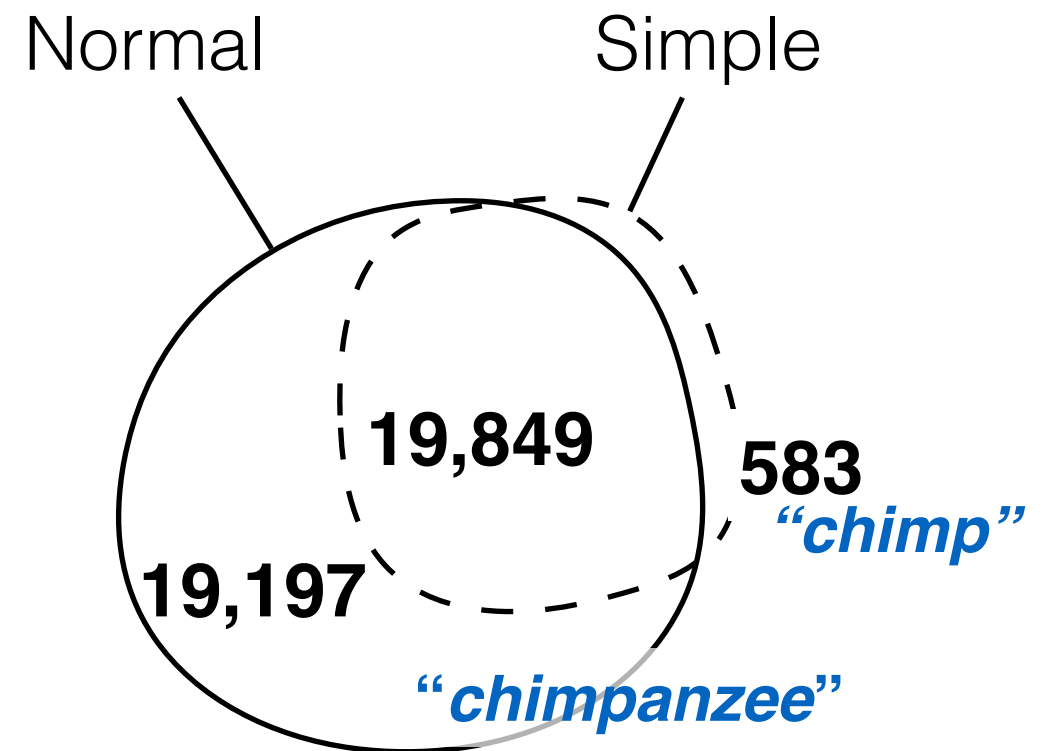
Newsela

(total 1.3 million tokens)

Good simplification uses a much smaller vocabulary.



18% reduction



48% reduction

vocabulary size (#unique words)

Wikipedia*

Newsela

Good simplification reduces certain function word usage.

commune

,

as

and

northern

northwestern

film

;

southwestern

footballer

,

and

"

—

of

which

as

percent

including

director

most significantly reduced words

(weighted log-odds-ratio analysis w/ informative Dirichlet prior)

Wikipedia*

Newsela

- Normal
- Simple

*Postal officials recently tried to , **which** could*

*Postal officials recently tried to **That** could*

most significantly reduced words
see syntax analysis in the paper

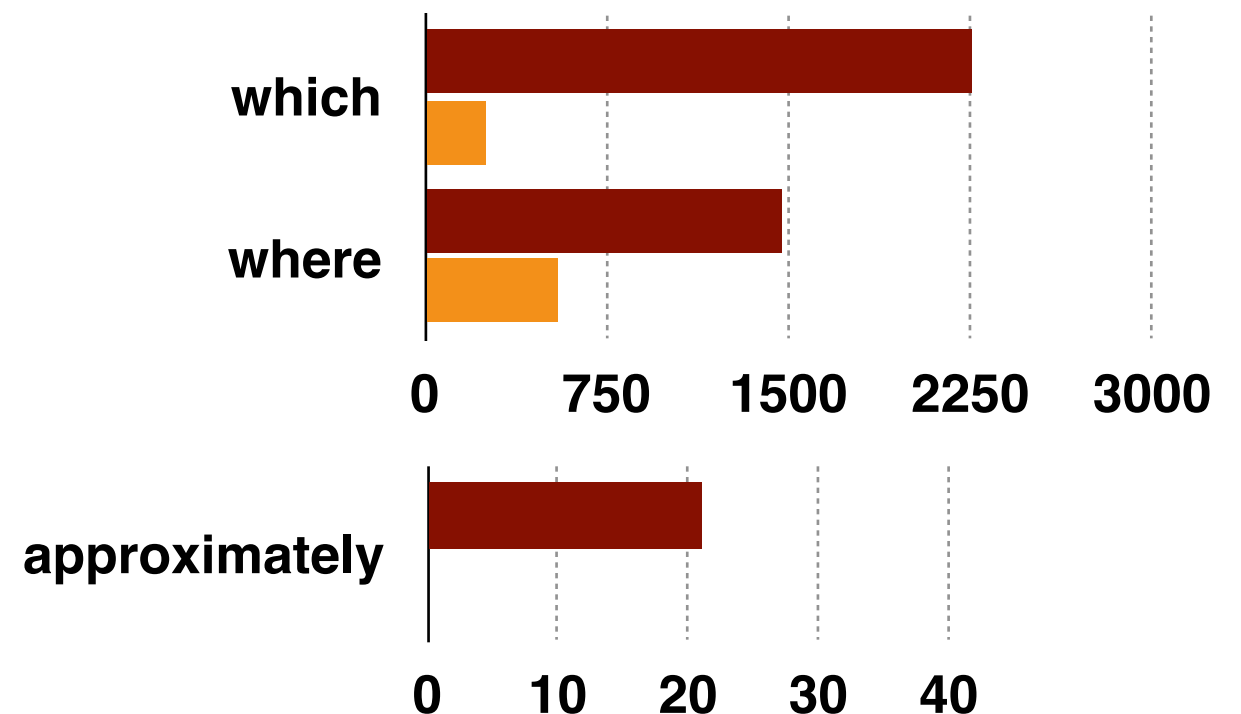
Wikipedia*

Newsela

- Normal
- Simple

*Postal officials recently tried to , **which** could*

*Postal officials recently tried to **That** could*



most significantly reduced words
see syntax analysis in the paper

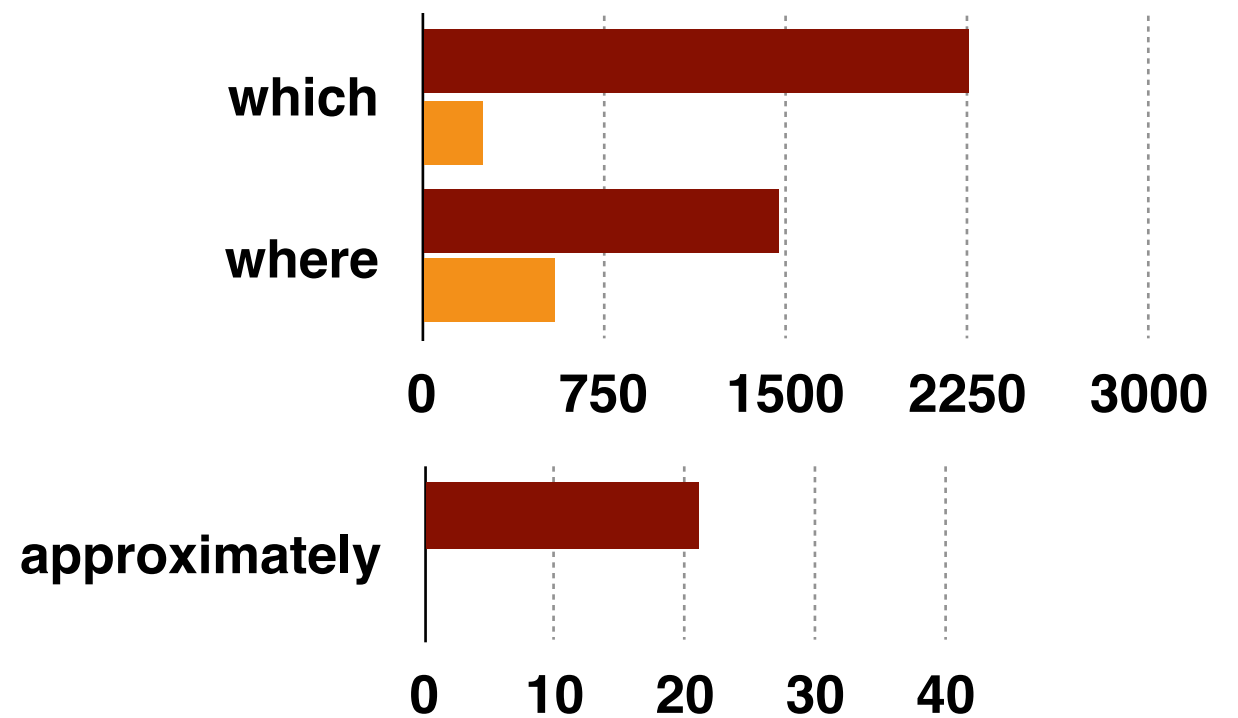
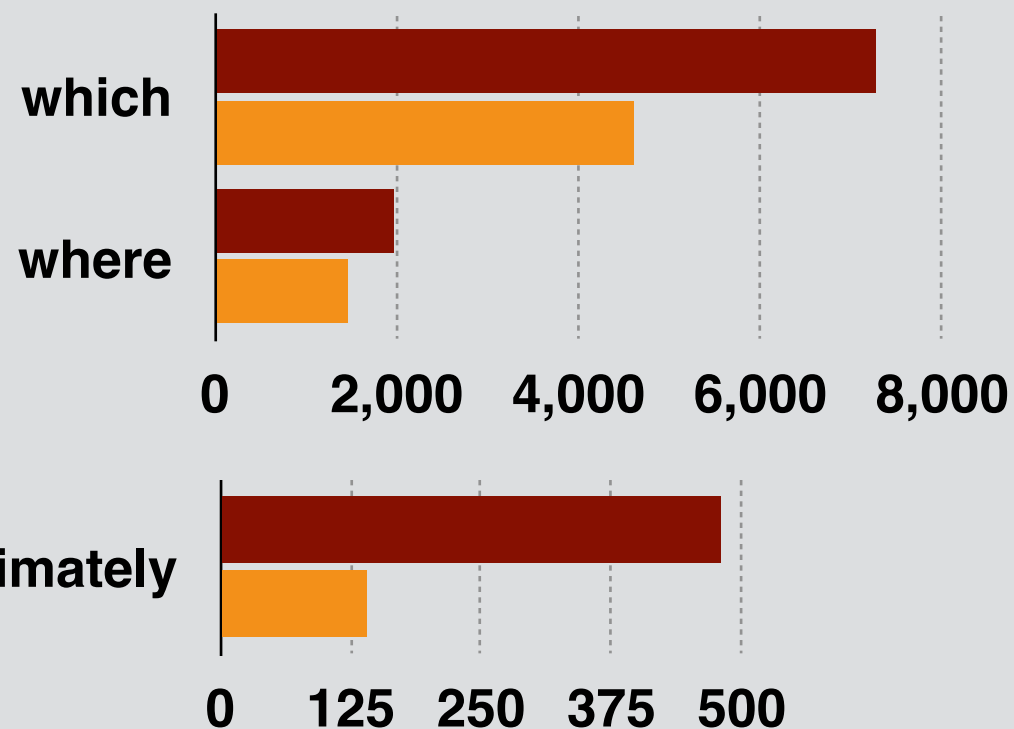
Wikipedia*

Newsela

■ Normal
■ Simple

*Postal officials recently tried to , **which** could*

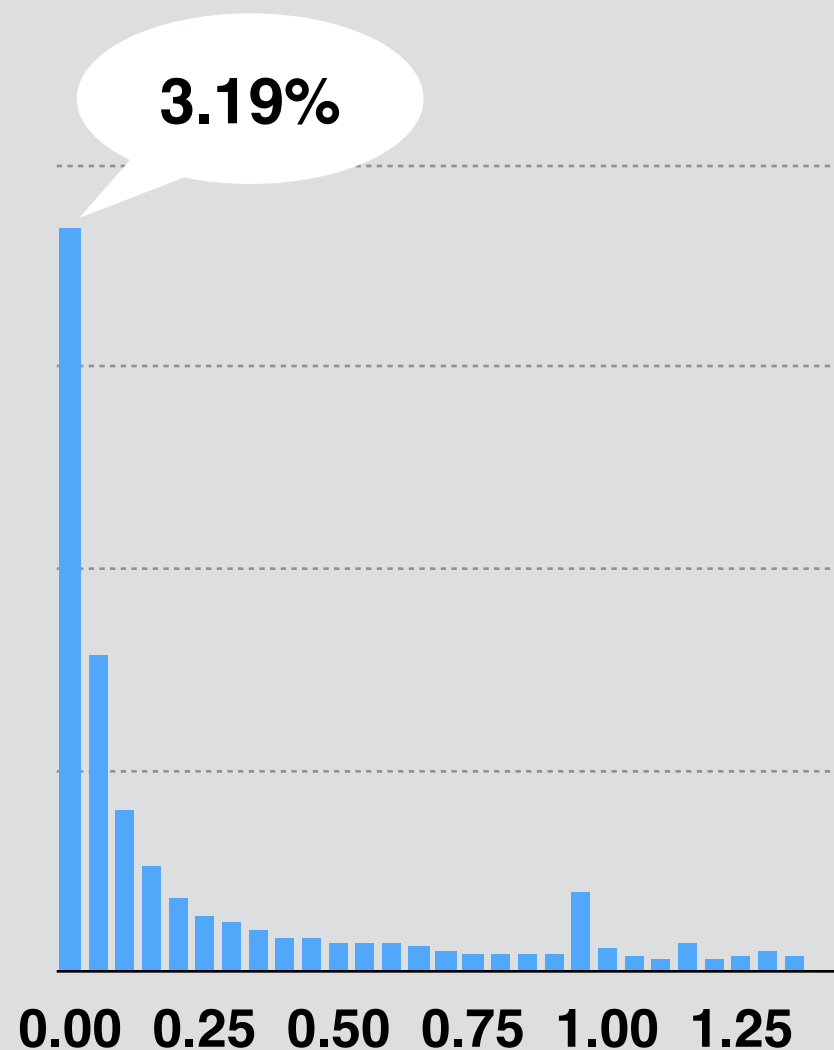
*Postal officials recently tried to **That** could*



most significantly reduced words
see syntax analysis in the paper

Wikipedia*

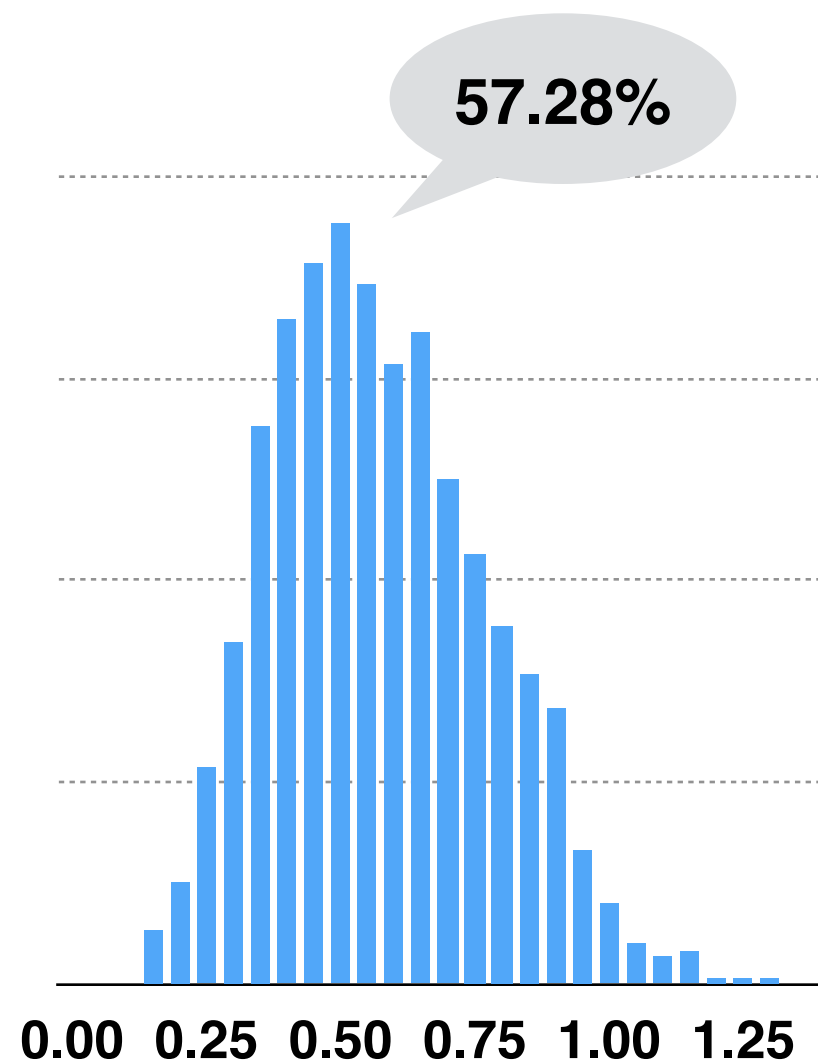
Wikipedia is not suitable for full-document simplification.



document compression ratio (simple/normal)

see discourse analysis in the paper

Newsela



Opinion #1 **Current evaluation**
doesn't tell us what's going on.

Opinion #2 **Simple Wikipedia is** **not** **that simple.**

Opinion #3 **New data** **can** **help.**

My Suggestions

My Suggestions

- **to reviewers:**

My Suggestions

- **to reviewers:**
 - be open-minded to papers that may not follow previous evaluation setup, may not outperform the “state-of-the-art” on Wikipedia

My Suggestions

- **to reviewers:**
 - be open-minded to papers that may not follow previous evaluation setup, may not outperform the “state-of-the-art” on Wikipedia
 - be sympathetic towards papers specially on data construction*, data analysis* and automatic evaluation metrics

* (Pellow & Maxine, 2014 HCOMP; Marcelo & Specia, 2014 PITR)

My Suggestions

- **to reviewers:**
 - be open-minded to papers that may not follow previous evaluation setup, may not outperform the “state-of-the-art” on Wikipedia
 - be sympathetic towards papers specially on data construction*, data analysis* and automatic evaluation metrics
 - read our paper

* (Pellow & Maxine, 2014 HCOMP; Marcelo & Specia, 2014 PITR)

My Suggestions

My Suggestions

- **to researchers:**

My Suggestions

- **to researchers:**
 - consider working on text simplification (“pre-BLEU age”)

My Suggestions

- **to researchers:**
 - consider working on text simplification (“pre-BLEU age”)
 - improve evaluation

My Suggestions

- **to researchers:**
 - consider working on text simplification (“pre-BLEU age”)
 - improve evaluation
 - make your system replicable

My Suggestions

- **to researchers:**
 - consider working on text simplification (“pre-BLEU age”)
 - improve evaluation
 - make your system replicable
 - read our paper

Thank you

Questions?
Opinions?

Sponsor: NSF

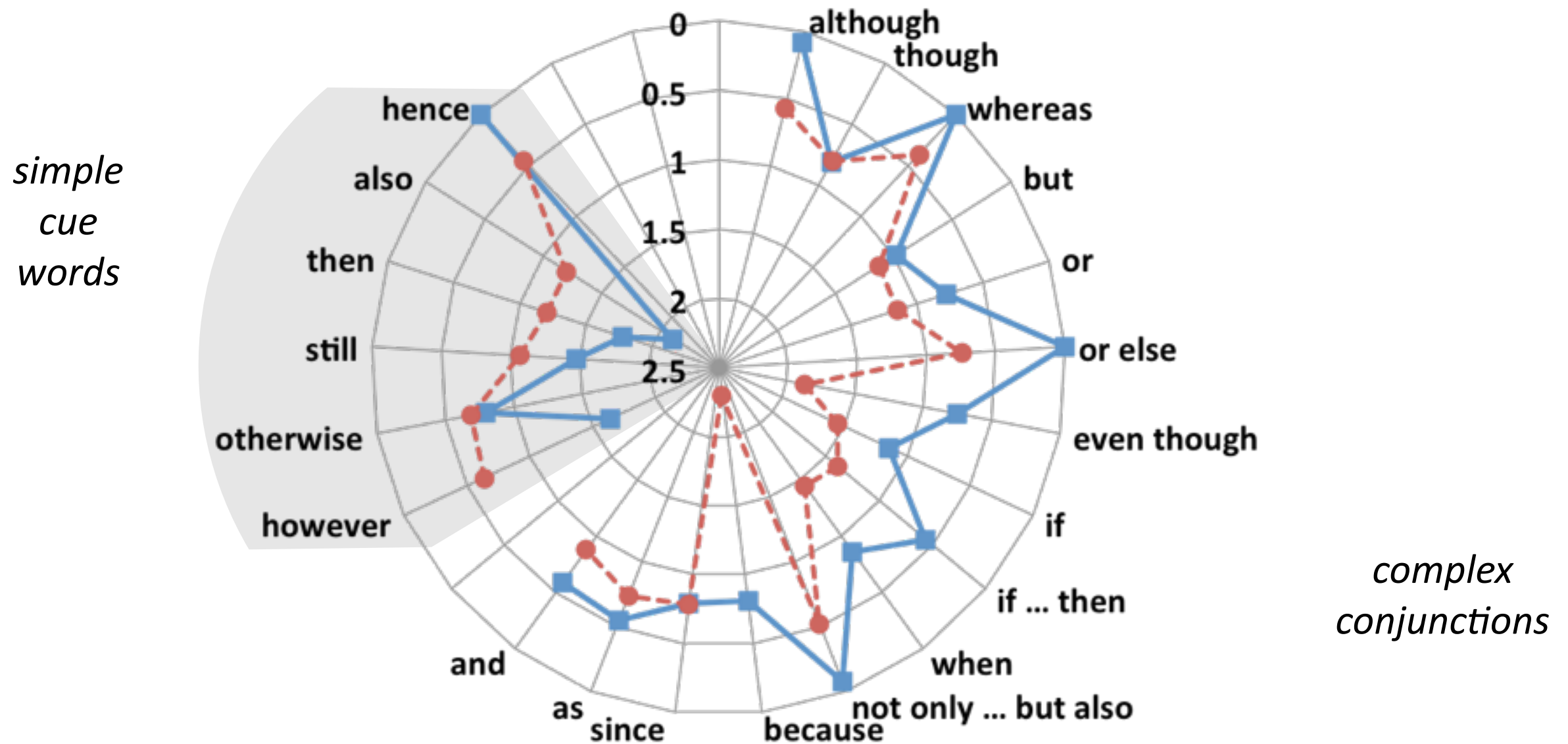
Newsela data are available at <https://newsela.com/data/>

Back Up

Wikipedia*



Newsela



change of discourse connectives (odds-ratio)

Reasons of Quality Issues

in Parallel Wikipedia Corpus

- The Simple Wikipedia was created by volunteers with no specific objective;
- Articles in Simple Wikipedia do not necessarily map Normal Wikipedia;
- As an encyclopedia, Wikipedia contains extremely difficulty words and sentences.

Newsela Dataset

Original	Slightly more fourth-graders nationwide are reading proficiently compared with a decade ago, but only a third of them are now reading well, according to a new report.
Simple-1	Fourth-graders in most states are better readers than they were a decade ago. But only a third of them actually are able to read well, according to a new report.
Simple-2	Fourth-graders in most states are better readers than they were a decade ago. But only a third of them actually are able to read well, according to a new report.
Simple-3	Most fourth-graders are better readers than they were 10 years ago. But few of them can actually read well.
Simple-4	Fourth-graders are better readers than 10 years ago. But few of them read well.

Newsela Dataset

1,130 news articles

Time: 2013 January ~ 2015 March

Source: Chicago Tribune, Seattle Times,
LA Times, The Baltimore Sun

Original: 56k sentences — Simple: 64k sentences