

Revisiting non-English Text Simplification

A Unified Multilingual Benchmark

Michael J. Ryan, Tarek Naous, Wei Xu

Georgia Tech

TL;DR

We released a collection of **27 datasets** for text simplification, spanning **12 diverse languages**, containing **1.7 million sentence pairs**.

We further showed that this resource is useful in training and evaluating automatic text simplification models **especially for languages with fewer resources**.

Code and Data

Georgia Tech

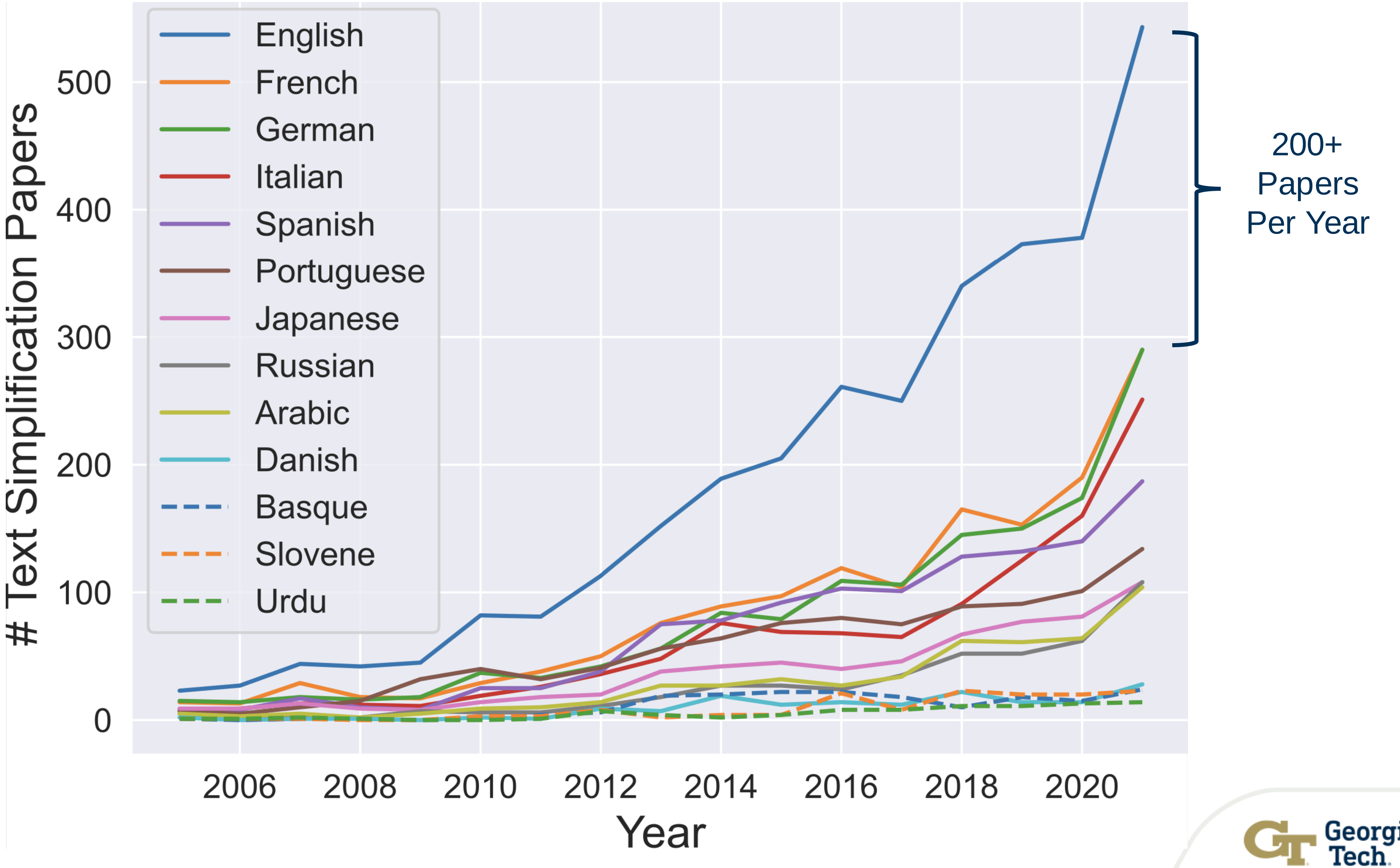
Background

Automatic Text Simplification (What)

"We hold these truths to be self-evident, that all men are created equal..."

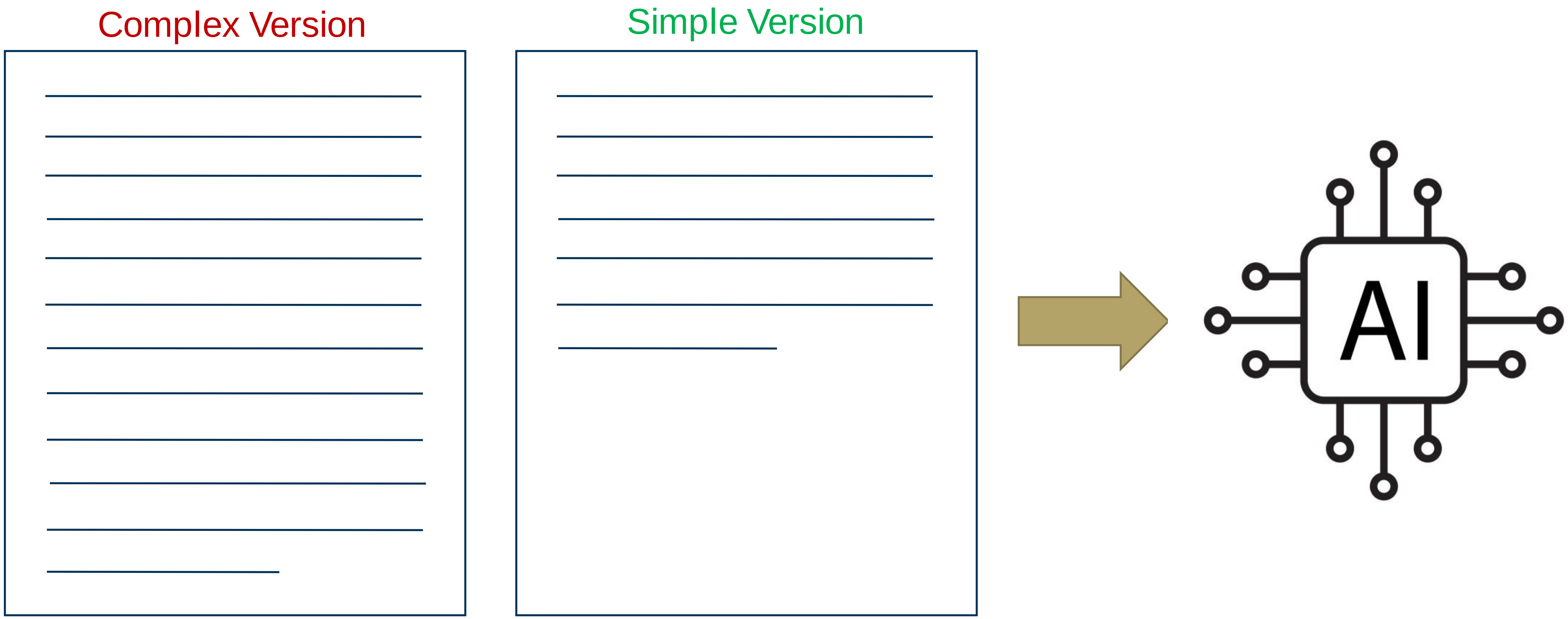
"We believe that it is obvious that all people are born with equal rights..."

Automatic Text Simplification (Why)

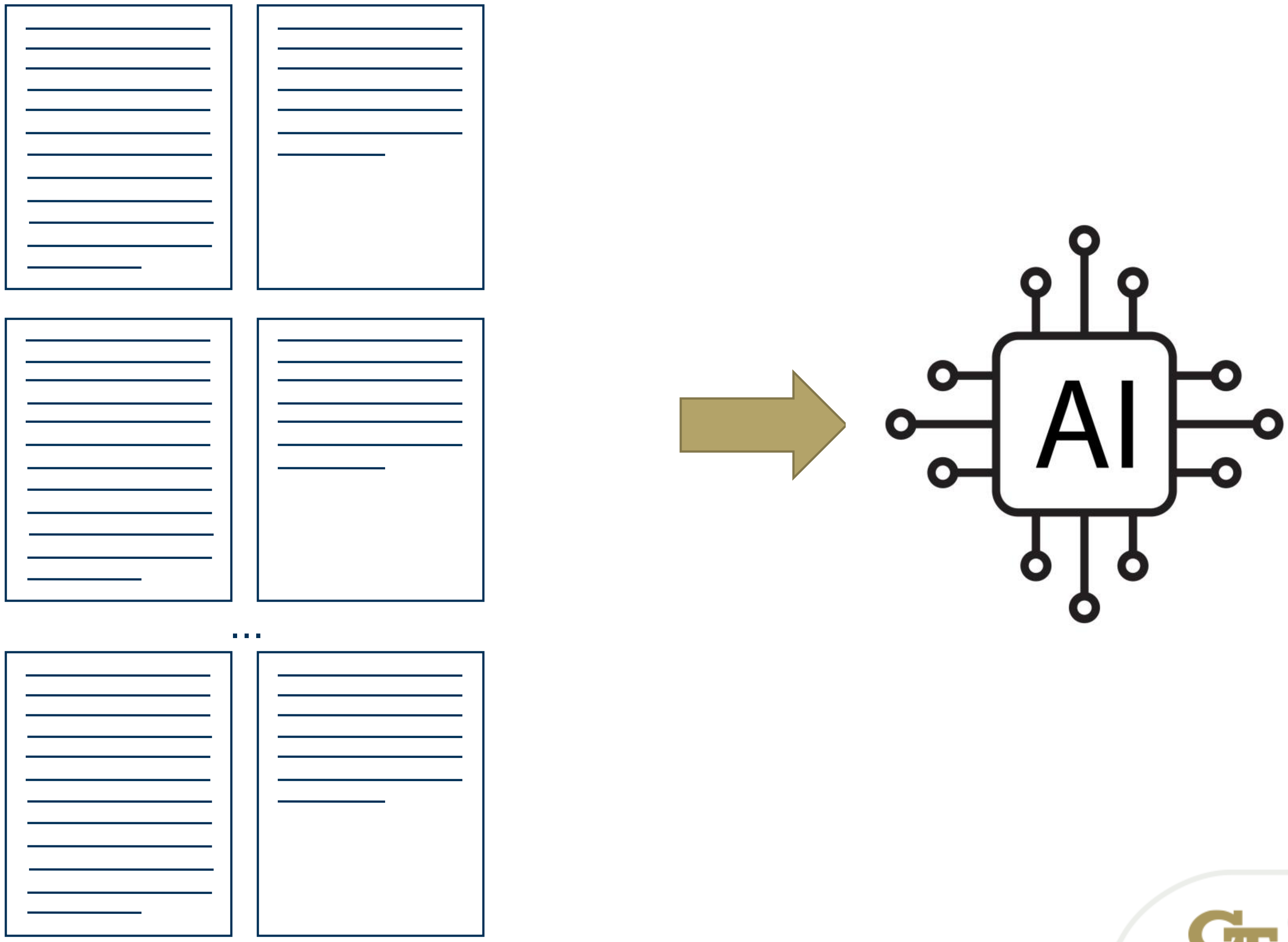


English
French
German
Italian
Spanish
Portuguese
Japanese
Russian
Arabic
Danish
Basque
Slovene
Urdu
Text Simplification Papers
0
100
200
300
400
500
2006
2008
2010
2012
2014
2016
2018
2020
Year
200+
Papers
Per Year

Automatic Text Simplification (How)



Automatic Text Simplification (How)



The Problem

This data is not readily available in languages besides English!

Our Contributions

(1) Comprehensive collection of all existing resources for text simplification.

- 1.7 million complex-simple sentence pairs in 12 different languages

(2) Experiments using multilingual language models to show the utility of this dataset especially in languages with limited data.

Part 1: The Dataset (MultiSim)

Collection Strategies

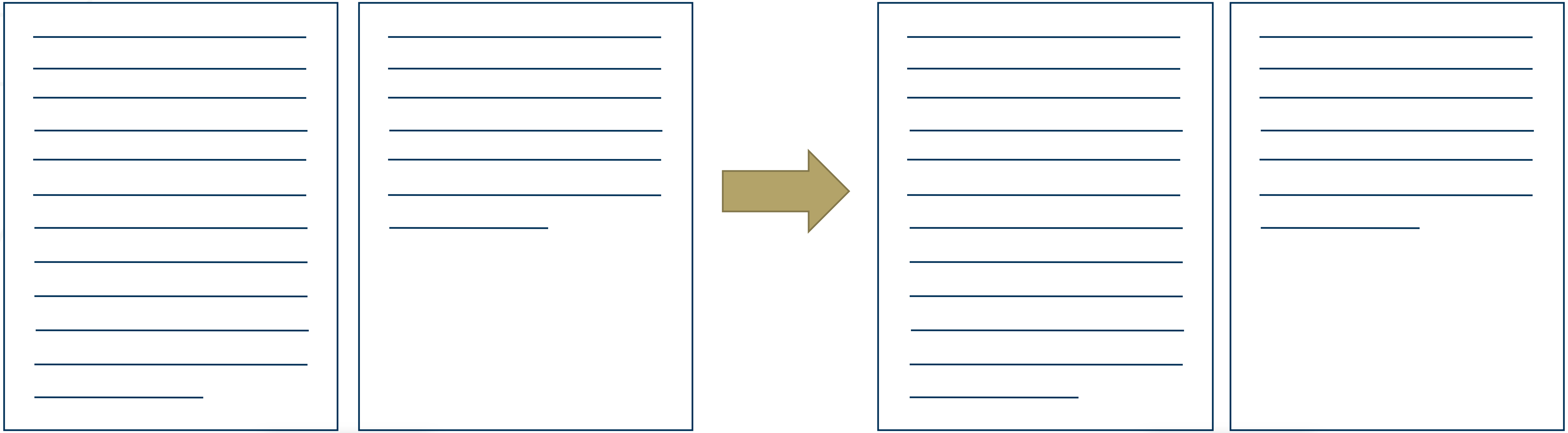
- Machine Translation
- Automatic Collection
- Manual Simplification
- Target Audience Resources

Lower Quality + Easy to Obtain

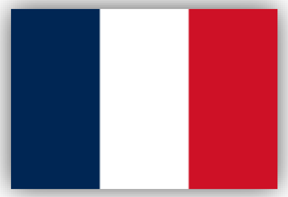
Higher Quality + Hard to Obtain

Machine Translation

- Low Quality, Easy to Obtain



2 Resources in MultiSim



Automatic Collection

- Low/Medium Quality, Easy/Moderate to Obtain

Simple English Wikipedia

Georgia Tech

39 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

The **Georgia Institute of Technology**, commonly referred to as **Georgia Tech** or, in the state of Georgia, as **Tech** or **The Institute**,[9] is a public research university and institute of technology in Atlanta, Georgia.[10] Established in 1885, it is part of the University System of Georgia and has satellite campuses in Savannah, Georgia; Metz, France; Shenzhen, China; and Singapore.

The school was founded as the Georgia School of Technology as part of Reconstruction plans to build an industrial economy in the post-Civil War Southern United States. Initially, it offered only a degree in mechanical engineering. By 1901, its curriculum had expanded to include electrical, civil, and chemical engineering. In 1948, the school changed its name to reflect its evolution from a trade school to a larger and more capable technical institute and research university. Today, Georgia Tech is organized into 6 colleges and contains about 31 departments and academic units, with emphasis on science and technology.

Student athletics, both organized and intramural, are a part of student and alumni life. The school's intercollegiate competitive sports teams, the four-time football national champion Yellow Jackets, and the nationally recognized fight song "Ramblin' Wreck from Georgia Tech", have helped keep Georgia Tech in the national spotlight. Georgia Tech fields eight men's and seven women's teams that compete in the NCAA Division I athletics and the Football Bowl Subdivision. Georgia Tech is a member of the Coastal Division in the Atlantic Coast Conference.

History [edit]

Main article: History of Georgia Tech

Wikipedia

Georgia Institute of Technology

39 languages

Page Talk Read Edit Change source View history Tools

From Simple English Wikipedia, the free encyclopedia

(Redirected from Georgia Tech)

The **Georgia Institute of Technology** is a public university (school for after high school) in Atlanta, Georgia. It is also known as **Georgia Tech**.

It focuses primarily on engineering, science and computer science,[1] but also has schools of management, architecture, and liberal arts.

When it opened in 1885, the school's main focus was on practical experience (learning things by doing them); students spent half of their time in class, and half in the shop.

It first admitted women in 1952.

In 1961, it was the first university in the South to integrate (let people of different races go to school together) without a court order.

The school's colors are white and old gold (mustard yellow). Their mascot is Buzz, a Yellow Jacket (a kind of bee-like wasp).

The Tech Tower, a big building at the **Georgia Institute of Technology**

Wikimedia Commons has media related to ***Georgia Institute of Technology***.

References [change | change source]

1. ↑ "High Tech: Georgia Tech, in Higher Education at the Crossroads of Disruption: the University of the 21st century, Andreas Kaplan, 2021, London: Emerald Publishers". *emerald.com*. Retrieved 2021-04-22.

- "Georgia Institute of Technology". *gatech.edu*. Retrieved 2008-06-28.

Automatic Collection

- Low/Medium Quality, Easy/Moderate to Obtain

The **Georgia Institute of Technology**, commonly referred to as **Georgia Tech** or, in the state of Georgia, as **Tech** or **The Institute**, is a public research university and institute of technology in Atlanta, Georgia.

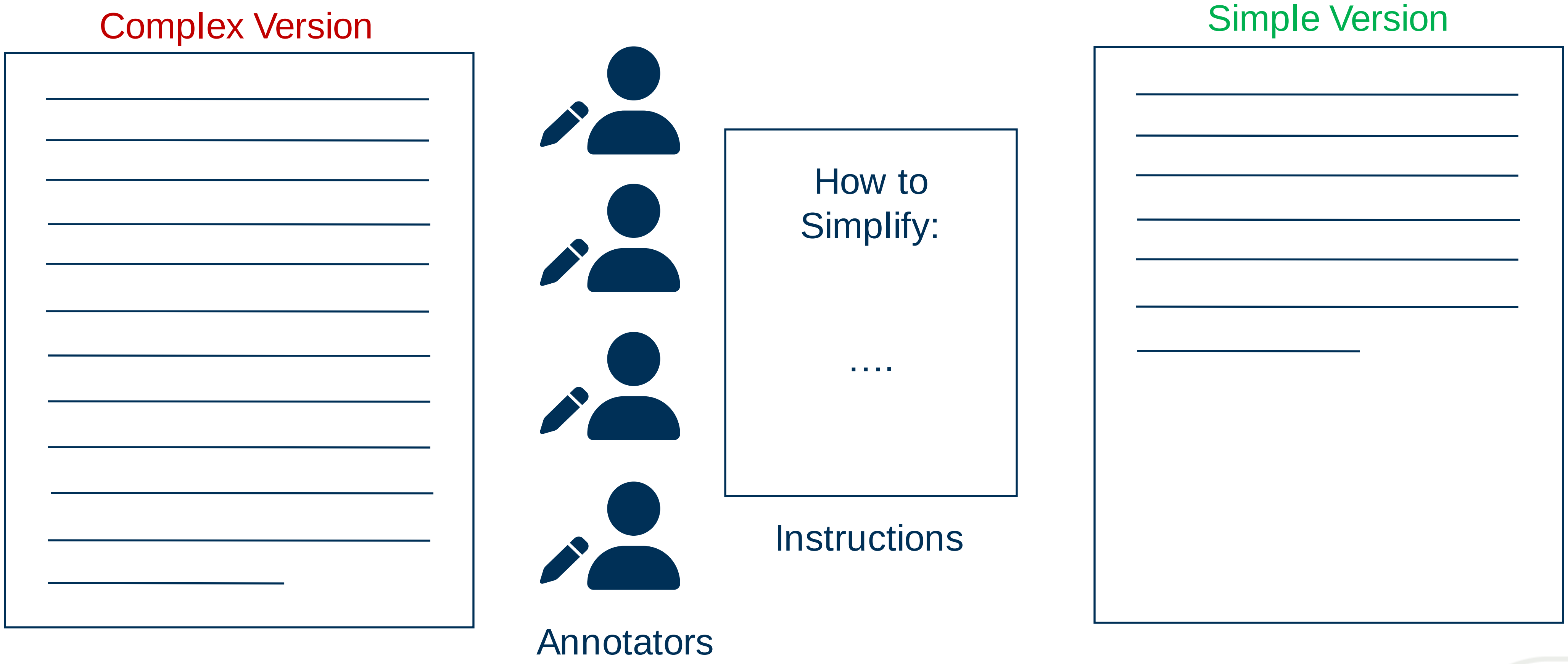
The **Georgia Institute of Technology** is a public university (school for after high school) in Atlanta, Georgia. It is also known as **Georgia Tech**.

4 Resources in MultiSim

Georgia Tech

Manual Simplification

- Medium/High Quality, Difficult to Obtain



15 Resources in MultiSim

Target Audience Resources

- High Quality, Very Difficult to Collect (Company Driven)

newsela

DR

NHK

Japan Broadcasting Corporation

6 Resources in MultiSim

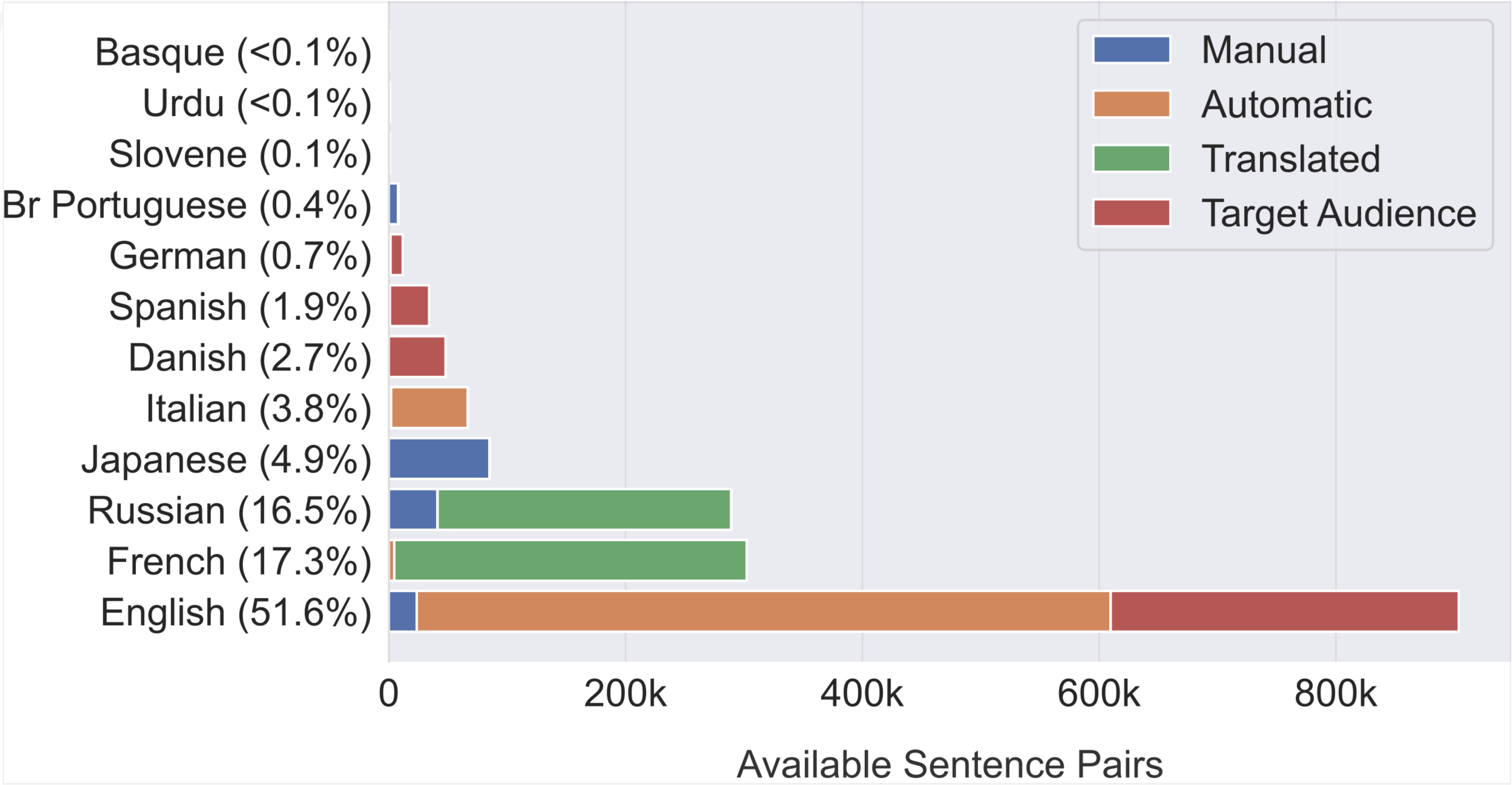
Georgia Tech

MultiSim Datasets

Corpus	Source(s)	Simplification Author	Collection Strategy	Alignment Level	Sentence Aligned	Complex Sentences	Simple Sentences	Access
Arabic Corpora								
Saaq al-Bambuu (Khallaf and Sharoff, 2022)	📕	writer	★	sentence	auto	2,980	2,980	private
Basque Corpora								
CBST (Gonzalez-Dios et al., 2018)	⚗	translator, teacher	✎	document	manual	458	591	on request
Brazilian Portuguese Corpora								
PorSimples (Aluísio and Gasperin, 2010)	📰 ⚗	linguist	✎	document	manual	7,902	10,174	on request
Danish Corpora								
DSim (Klerke and Søgaard, 2012)	📰	journalists	★	sentence	auto	47,887	60,528	on request
English Corpora†								
ASSET (Alva-Manchego et al., 2020)	W	crowdsource	✎	sentence	manual	2,359	23,590	open source
Newsela EN (Xu et al., 2015)	📰	experts	★	document	auto	393,798	402,222	on request
Wiki-Auto (Jiang et al., 2020)	W	crowdsource	⚙	document	auto	10,144,476	1,241,671	open source
French Corpora								
Alector (Gala et al., 2020)	🌐 ⚗	experts	✎	document	NA	1,230	1,192	open source
CLEAR (Grabar and Cardon, 2018)	W ✚	crowdsource, experts	⚙	sentence	auto	4,596	4,596	open source
WikiLarge FR (Cardon and Grabar, 2020)	W	crowdsource	🈂	sentence	auto	307,067	308,409	open source
German Corpora								
GEOLinoTest (Mallinson et al., 2020)	📰	linguist	✎	sentence	manual	1,198	1,198	open source
German News (Säuberli et al., 2020)	📰	news agency	★	document	auto	15,239	14,344	on request
Klexikon (Aumiller and Gertz, 2022)	W	crowdsource	⚙	document	NA	771,059	96,870	open source
Simple Patho (Trienes et al.)	✚	medical students	✎	paragraph	manual	22,191	26,551	private
Simple German (Battisti et al., 2020)	🌐	government	★	document	auto	12,806	8,400	on request*
TextComplexityDE (Naderi et al., 2019)	W	native speaker	✎	document	manual	250	250	open source
Italian Corpora								
AdminIT (Miliani et al., 2022)	🔨	researchers	✎	sentence	manual	777	763	open source
SIMPITIKI Wiki (Tonelli et al., 2016)	W	crowdsource	⚙	sentence	manual	575	575	open source
PaCCSS-IT (Brunato et al., 2016)	🌐	crowdsource	⚙	sentence	auto	63,006	63,006	open source
Teacher (Brunato et al., 2015)	🌐	teachers	✎	document	manual	204	195	open source
Terence (Brunato et al., 2015)	📕	experts	✎	document	manual	1,035	1,060	open source
Japanese Corpora								
EasyJapanese (Maruyama and Yamamoto, 2018)	📰 🌐	students	✎	sentence	manual	50,000	50,000	open source
EasyJapaneseExtended (Katsuta and Yamamoto, 2018)	📰 🌐	crowdsource	✎	sentence	manual	34,400	35,000	open source
Japanese News (Goto et al., 2015)	📰	journalists, teachers	★	document	auto	13,356	13,356	private
Russian Corpora								
RuAdapt Encyclopedia (Dmitrieva et al., 2021)	ℹ	researchers	✎	document	auto	9,729	10,230	open source
RuAdapt Fairytale (Dmitrieva et al., 2021)	📕	researchers	✎	document	auto	310	404	open source
RuAdapt Lit (Dmitrieva and Tiedemann, 2021)	📕	writers	✎	document	auto	24,152	28,259	on request
RSSE (Sakhovskiy et al., 2021)	W	crowdsource	✎	sentence	manual	2,000	6,804	open source
RuWikiLarge (Sakhovskiy et al., 2021)	W	crowdsource	🈂	sentence	auto	278,499	289,788	on request
Slovene Corpora								
SloTS (Gorenc and Robnik-Šikonja, 2022)	📕	experts	★	sentence	manual	1,181	1,287	open source
Spanish Corpora								
FIRST (Orasan et al., 2013)	📕 📰 ✚	experts	✎	document	manual	320	332	private
Newsela ES (Xu et al., 2015)	📰	experts	★	document	auto	46,256	45,519	on request
Simplext (Saggion et al., 2015)	📰	researchers	✎	document	manual	1,108	1,742	on request
Urdu Corpora								
SimplifyUREval (Qasmi et al., 2020)	📕 📰	expert	✎	sentence	manual	500	736	open source

Table 1: Important properties of text simplification parallel corpora. †Common English corpora included for comparison. Many other English corpora omitted. *Only scripts to replicate the corpus are available upon request. Simple German results differ from original paper because of changes to availability of online articles. *Sources*: 📕 Literature, ⚗ Science Communications, 📰 News, W Wikipedia, 🌐 Websites, ✚ Medical Documents, 🔨 Government, ℹ Encyclopedic. *Collection Strategies*: ⚙ Automatic, 🈂 Translation, ✎ Annotator, ★ Target Audience Resource.

MultiSim Data



MultiSim Splits

Language	Dataset	#train	#test	#dev
Open Source				
English	WikiAuto	576,126	5,012	5,012
	ASSET*	20,000	3,590	0
French	WikiLargeFR*	296,402	359	992
	CLEAR*	4,196	100	300
German	GEOLino	958	122	118
	TextCompDE	200	25	25
Italian	PaCCSS-IT	60,485	1,267	1,254
	Terence	809	101	102
	AdminIT	588	73	75
	Simpitiki	460	56	59
	Teacher	136	17	17
Japanese	Easy JA	48,000	1,000	1,000
	Easy JA Ext*	34,269	731	0
Russian	RuAdapt Ency	7,782	982	965
	RSSE Corpus*	3,406	3,398	0
	RuAdapt Fairy	248	31	31
Slovene	SloTS*	749	96	94
Urdu	SimplifyUR	594	68	74
	Total	1,055,408	17,028	10,118
Dataloaders Available (Data on Request)				
Basque	CBST	361	46	46
Br Portuguese	PorSimples	6,290	790	784
Danish	DSim Corpus	45,885	997	1,005
English	Newsela EN	291,969	991	1,008
German	German News	8,186	1,024	1,023
Russian	RuWikiLarge*	246,978	365	768
	RuAdapt Lit	22,152	1,000	1,000
Spanish	Newsela ES	30,910	1,001	1,001
	Simplext	737	92	93
	Total	653,468	6,306	6,728

Table 2: MULTISIM splits. *Original splits preserved

Part 2: The Experiments

Automatic Metrics

BLEU Scores

- Used in Machine Translation
- Decent measure of Adequacy/Fluency of model output
- Penalizes model from deviating from expected output

SARI Scores

- Used in Text Simplification
- Decent measure of output Simplicity
- Primary focus of our experiments
- Rewards model for removing unnecessary parts of the input while remaining close to the expected output.

Models

mT5: Multilingual Text to Text Transformer (580M Parameters)

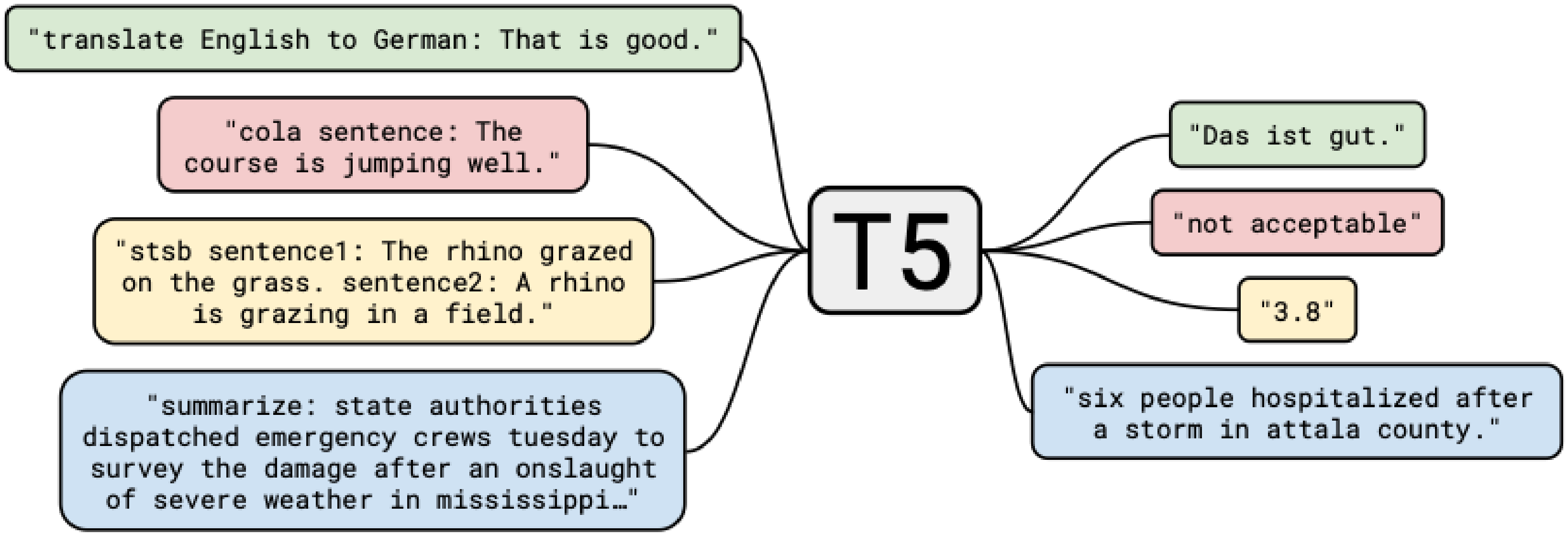


Image Credit: [Google Research]

Models

BLOOM (176B parameters)

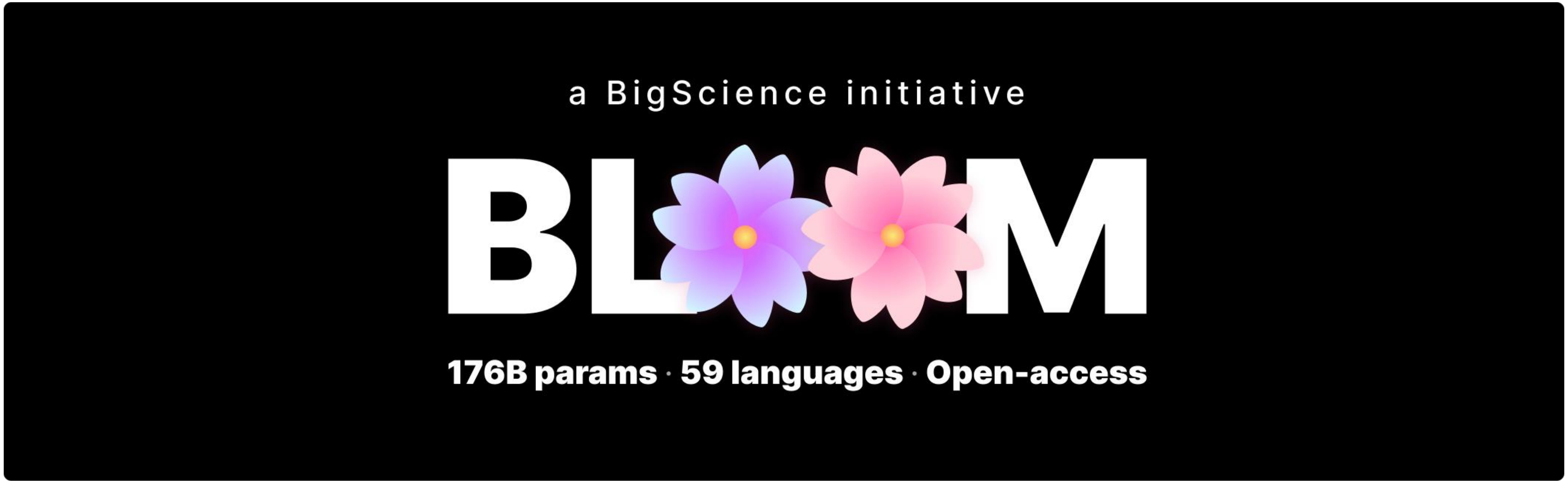


Image Credit: [BigScience]

Baseline: Identity

"This is the original sentence for the model to simplify"

↓

"This is the original sentence for the model to simplify"

Tends to score very highly on BLEU, but low in SARI

Baseline: Identity

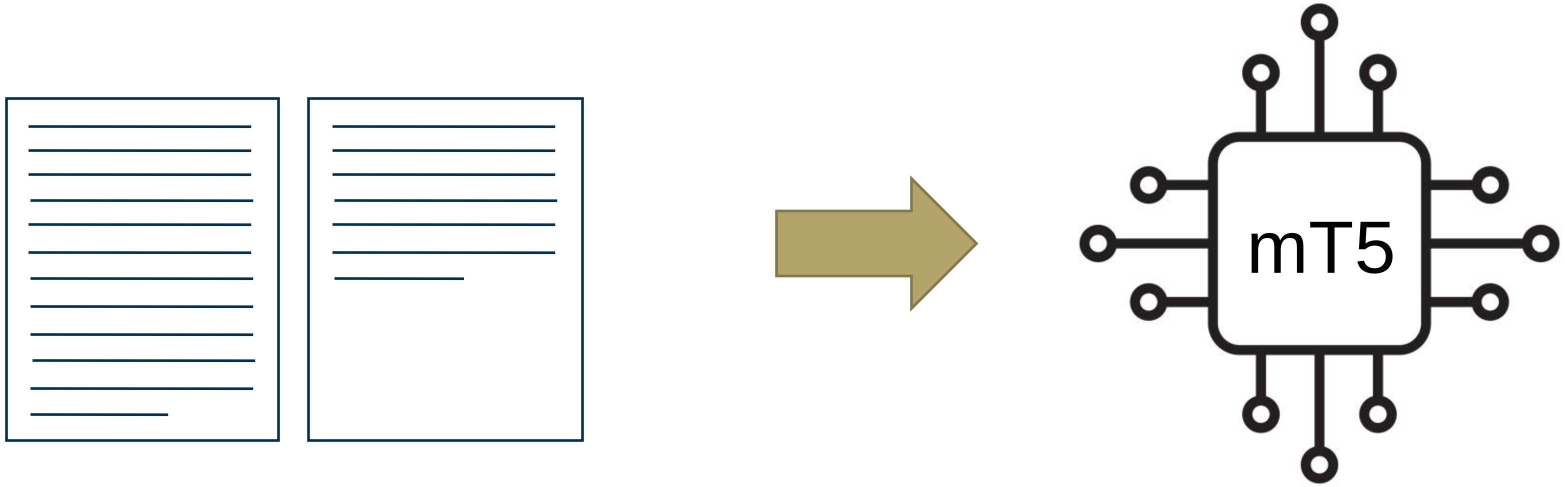
"This is the original sentence for the model to simplify"

↓

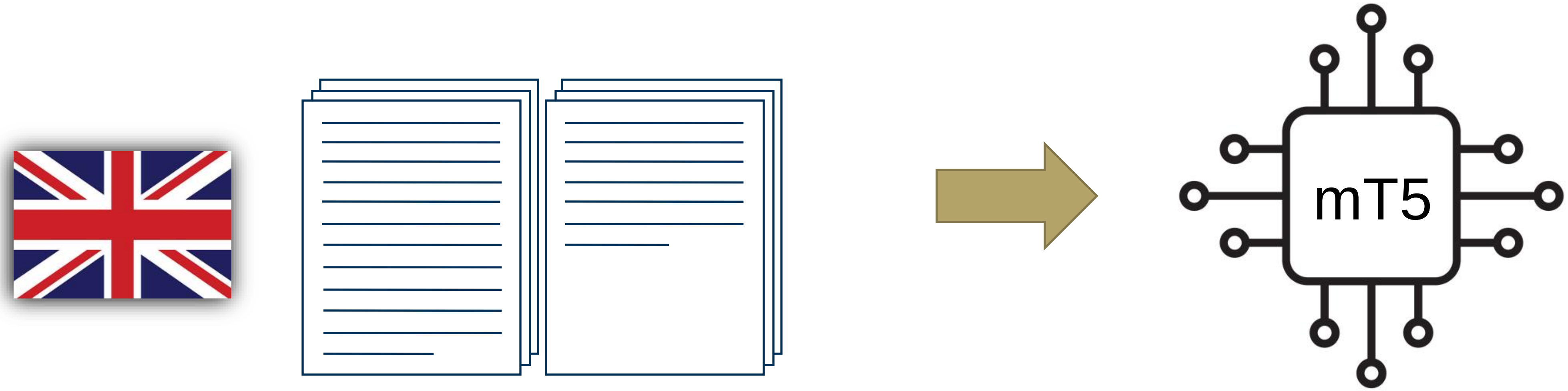
"This is the original sentence for the model"

Tends to score highly on SARI, but lower in BLEU

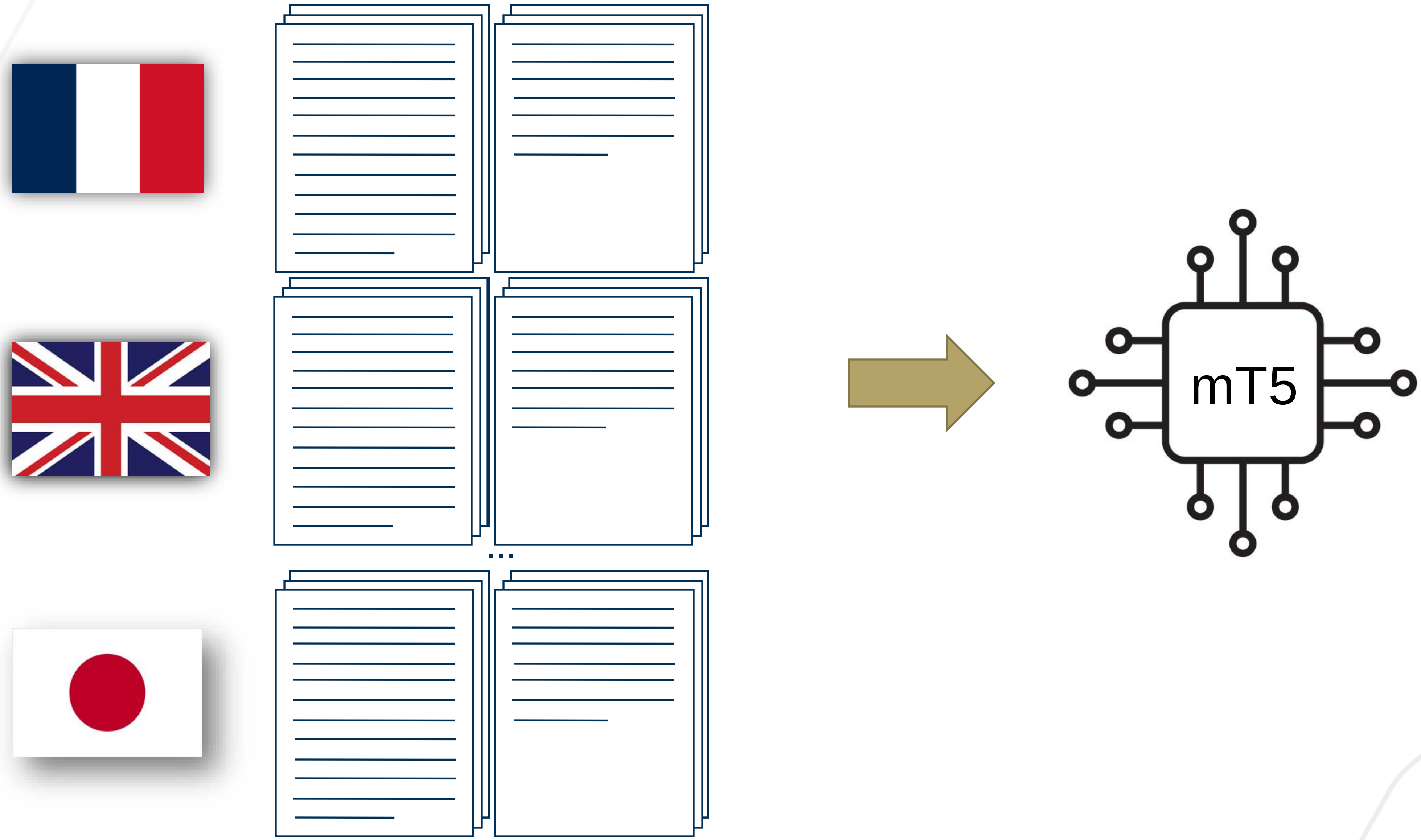
Finetuning Experiments: Single



Finetuning Experiments: Language



Finetuning Experiments: All



Finetuning Experiments

Training Setting

Evaluation Setting

			BLEU					SARI				
			Baseline		Finetune			Baseline		Finetune		
Lang	Dataset	Size	Identity	Trunc	Single	Lang	All	Identity	Trunc	Single	Lang	All
eu	CBST	218	72.02	57.87	—	—	66.75	23.46	32.58	—	—	**32.83**
ur	SimplifyUR	470	58.85	41.11	—	—	56.23	24.84	31.30	—	—	**51.74**
sl	SloTS	188	7.76	6.09	—	—	7.63	5.93	19.03	—	—	**30.52**
pt-br	PorSimples	1,949	73.67	51.93	—	—	63.85	28.21	31.25	—	—	**44.27**
de	TextCompDE	144	26.77	19.98	—	—	24.53	15.42	26.81	—	—	**41.15**
	GEOLino	437	69.86	50.03	—	—	71.90	27.45	30.70	—	—	**50.75**
	GermanNews	1,748	7.29	7.13	—	—	6.57	5.61	17.69	—	—	**31.58**
es	Simplext	157	13.91	13.15	—	14.42	12.25	7.94	20.27	—	19.91	**32.68**
	NewselaES	17,022	58.18	43.06	51.78	53.12	48.94	24.21	31.64	29.89	28.56	**35.36**
da	DSim	25,524	31.39	28.85	33.66	33.66	27.25	16.25	26.10	31.40	31.40	**38.44**
it	Teacher	83	34.49	29.05	—	32.21	29.76	17.41	27.75	—	29.98	**30.97**
	AdminIT	114	52.50	45.63	—	40.09	43.80	20.89	28.22	—	34.72	**36.21**
	Terence	394	67.24	49.72	—	59.33	50.65	26.83	32.82	—	**37.77**	36.92
	Simpitiki	24	95.23	74.48	—	24.40	36.28	32.45	32.00	—	20.10	**24.27**
	PaCCSS-IT	55,274	36.76	28.77	49.57	48.31	42.87	18.14	28.26	**57.30**	55.98	54.43
ja	EasyJA	27,600	58.09	8.43	65.83	68.12	66.04	24.64	24.28	67.36	**70.95**	70.11
	EasyJAExt	32,248	20.23	0.00	33.07	35.67	31.50	9.00	35.32	43.15	50.26	**53.49**
ru	RuAdaptFairy	97	12.56	8.03	—	13.11	11.01	10.63	24.84	—	23.77	**26.55**
	RSSE	1,477	38.23	34.69	—	36.94	31.78	10.91	22.72	—	29.49	**35.08**
	RuAdapt Ency	1,450	84.15	59.66	—	76.06	61.83	29.90	31.09	—	**34.73**	34.40
	RuAdapt Lit	10,515	51.22	41.64	49.94	53.74	48.54	22.66	31.94	41.75	**42.03**	42.01
	RuWikiLarge	135,191	57.82	44.38	55.03	51.97	40.82	24.24	31.87	32.01	34.95	**37.59**
fr	CLEAR	3,179	55.00	45.10	25.45	53.72	48.57	23.73	32.17	34.86	30.85	**35.37**
	WikiLargeFR	148,276	58.51	46.67	52.43	51.16	43.57	24.44	32.23	35.20	38.22	**39.23**
en	ASSET	14,814	92.81	88.11	88.26	81.20	85.90	20.73	29.66	35.98	**42.77**	41.56
	NewselaEN	129,387	68.71	52.30	62.78	51.51	55.68	26.17	32.90	38.60	**40.18**	38.80
	WikiAuto	315,018	45.40	41.31	37.95	35.30	36.91	20.93	31.45	42.46	**42.48**	42.00

Finetuning Experiments

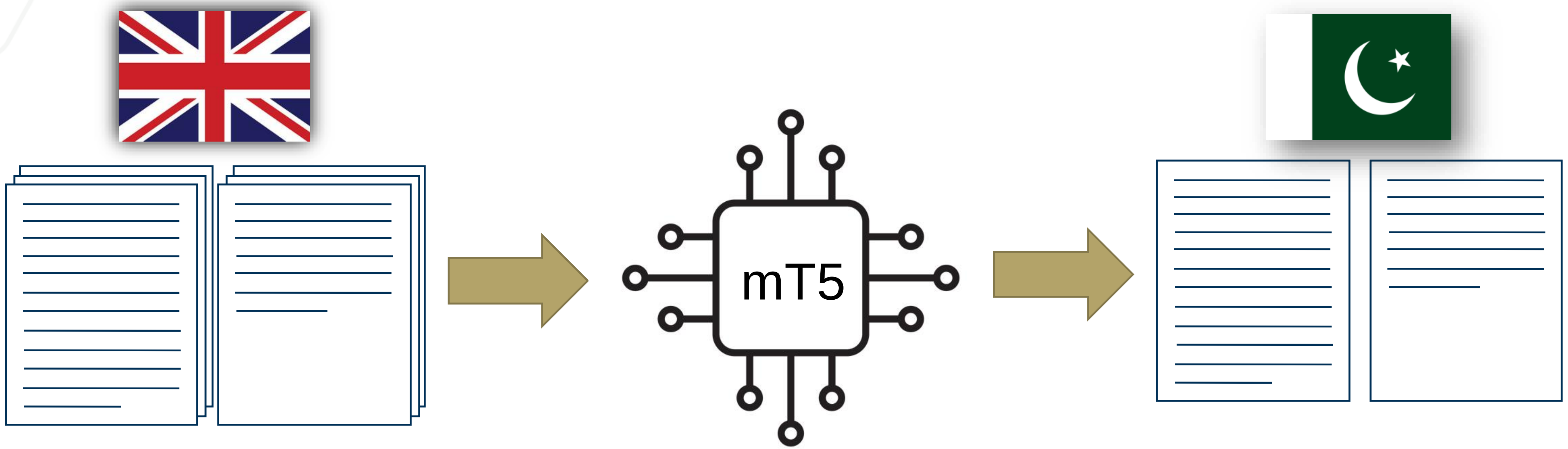
Less Data

More Data

			BLEU					SARI				
			Baseline		Finetune			Baseline		Finetune		
Lang	Dataset	Size	Identity	Trunc	Single	Lang	All	Identity	Trunc	Single	Lang	All
eu	CBST	218	72.02	57.87	—	—	66.75	23.46	32.58	—	—	**32.83**
ur	SimplifyUR	470	58.85	41.11	—	—	56.23	24.84	31.30	—	—	**51.74**
sl	SloTS	188	7.76	6.09	—	—	7.63	5.93	19.03	—	—	**30.52**
pt-br	PorSimples	1,949	73.67	51.93	—	—	63.85	28.21	31.25	—	—	**44.27**
de	TextCompDE	144	26.77	19.98	—	—	24.53	15.42	26.81	—	—	**41.15**
	GEOLino	437	69.86	50.03	—	—	71.90	27.45	30.70	—	—	**50.75**
	GermanNews	1,748	7.29	7.13	—	—	6.57	5.61	17.69	—	—	**31.58**
es	Simplext	157	13.91	13.15	—	14.42	12.25	7.94	20.27	—	19.91	**32.68**
	NewselaES	17,022	58.18	43.06	51.78	53.12	48.94	24.21	31.64	29.89	28.56	**35.36**
da	DSim	25,524	31.39	28.85	33.66	33.66	27.25	16.25	26.10	31.40	31.40	**38.44**
it	Teacher	83	34.49	29.05	—	32.21	29.76	17.41	27.75	—	29.98	**30.97**
	AdminIT	114	52.50	45.63	—	40.09	43.80	20.89	28.22	—	34.72	**36.21**
	Terence	394	67.24	49.72	—	59.33	50.65	26.83	32.82	—	**37.77**	36.92
	Simpitiki	24	95.23	74.48	—	24.40	36.28	32.45	32.00	—	20.10	**24.27**
	PaCCSS-IT	55,274	36.76	28.77	49.57	48.31	42.87	18.14	28.26	**57.30**	55.98	54.43
ja	EasyJA	27,600	58.09	8.43	65.83	68.12	66.04	24.64	24.28	67.36	**70.95**	70.11
	EasyJAExt	32,248	20.23	0.00	33.07	35.67	31.50	9.00	35.32	43.15	50.26	**53.49**
ru	RuAdaptFairy	97	12.56	8.03	—	13.11	11.01	10.63	24.84	—	23.77	**26.55**
	RSSE	1,477	38.23	34.69	—	36.94	31.78	10.91	22.72	—	29.49	**35.08**
	RuAdapt Ency	1,450	84.15	59.66	—	76.06	61.83	29.90	31.09	—	**34.73**	34.40
	RuAdapt Lit	10,515	51.22	41.64	49.94	53.74	48.54	22.66	31.94	41.75	**42.03**	42.01
	RuWikiLarge	135,191	57.82	44.38	55.03	51.97	40.82	24.24	31.87	32.01	34.95	**37.59**
fr	CLEAR	3,179	55.00	45.10	25.45	53.72	48.57	23.73	32.17	34.86	30.85	**35.37**
	WikiLargeFR	148,276	58.51	46.67	52.43	51.16	43.57	24.44	32.23	35.20	38.22	**39.23**
en	ASSET	14,814	92.81	88.11	88.26	81.20	85.90	20.73	29.66	35.98	**42.77**	41.56
	NewselaEN	129,387	68.71	52.30	62.78	51.51	55.68	26.17	32.90	38.60	**40.18**	38.80
	WikiAuto	315,018	45.40	41.31	37.95	35.30	36.91	20.93	31.45	42.46	**42.48**	42.00

Languages with less data tended to benefit from multilingual training.

Transfer Experiments



Transfer Experiments

Small Datasets

Transfer to Italian: Terence						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	7.96	23.92
				RuAdaptLit (ru)	57.07	35.19
			W	RuWikiLarge (ru)	10.76	25.85
✓			W	WikiAuto (en)	21.90	30.89
✓	✓		W	WikiLargeFR (fr)	8.66	25.20
✓	✓	✓	W	PaCCSS-IT (it)	40.59	36.99

Transfer to Basque: CBST						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	2.31	24.26
				EasyJA (ja)	1.87	26.67
				RuAdaptLit (ru)	47.31	37.89
✓				NewselaEN (en)	24.37	31.09
✓				PorSimples (pt-br)	7.97	29.44

Transfer to Urdu: SimplifyUR						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	5.20	33.18
			W	WikiAuto (en)	8.34	26.09
				EasyJA (ja)	0.00	17.87
				NewselaEN (en)	0.25	18.03
				RuAdaptLit (ru)	34.78	32.61
				PorSimples (pt-br)	5.24	32.87

Transfer Experiments

Small Datasets

Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
Transfer to Italian: Terence						
				Finetuned	7.96	23.92
				RuAdaptLit (ru)	57.07	35.19
			W	RuWikiLarge (ru)	10.76	25.85
✓			W	WikiAuto (en)	21.90	30.89
✓	✓		W	WikiLargeFR (fr)	8.66	25.20
✓	✓	✓	W	PaCCSS-IT (it)	40.59	36.99
Transfer to Basque: CBST						
				Finetuned	2.31	24.26
				EasyJA (ja)	1.87	26.67
				RuAdaptLit (ru)	47.31	37.89
✓				NewselaEN (en)	24.37	31.09
✓				PorSimples (pt-br)	7.97	29.44
Transfer to Urdu: SimplifyUR						
				Finetuned	5.20	33.18
			W	WikiAuto (en)	8.34	26.09
				EasyJA (ja)	0.00	17.87
				NewselaEN (en)	0.25	18.03
				RuAdaptLit (ru)	34.78	32.61
				PorSimples (pt-br)	5.24	32.87

1. Matching script and language improve transfer performance

Transfer Experiments

Small Datasets

Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
Transfer to Italian: Terence						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	7.96	23.92
				RuAdaptLit (ru)	57.07	35.19
			W	RuWikiLarge (ru)	10.76	25.85
✓			W	WikiAuto (en)	21.90	30.89
✓	✓		W	WikiLargeFR (fr)	8.66	25.20
✓	✓	✓	W	PaCCSS-IT (it)	40.59	36.99
Transfer to Basque: CBST						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	2.31	24.26
				EasyJA (ja)	1.87	26.67
				RuAdaptLit (ru)	47.31	37.89
✓				NewselaEN (en)	24.37	31.09
✓				PorSimples (pt-br)	7.97	29.44
Transfer to Urdu: SimplifyUR						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	5.20	33.18
			W	WikiAuto (en)	8.34	26.09
				EasyJA (ja)	0.00	17.87
				NewselaEN (en)	0.25	18.03
				RuAdaptLit (ru)	34.78	32.61
				PorSimples (pt-br)	5.24	32.87

1. Matching script and language improve transfer performance

2. Domain match can help regardless of script

Transfer Experiments

Small Datasets

Transfer to Italian: Terence						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	7.96	23.92
				RuAdaptLit (ru)	57.07	35.19
			W	RuWikiLarge (ru)	10.76	25.85
✓			W	WikiAuto (en)	21.90	30.89
✓	✓		W	WikiLargeFR (fr)	8.66	25.20
✓	✓	✓	W	PaCCSS-IT (it)	40.59	36.99

Transfer to Basque: CBST						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	2.31	24.26
				EasyJA (ja)	1.87	26.67
				RuAdaptLit (ru)	47.31	37.89
✓				NewselaEN (en)	24.37	31.09
✓				PorSimples (pt-br)	7.97	29.44

Transfer to Urdu: SimplifyUR						
Scr	Fam	Lang	Dom	Dataset	BLEU	SARI
				Finetuned	5.20	33.18
			W	WikiAuto (en)	8.34	26.09
				EasyJA (ja)	0.00	17.87
				NewselaEN (en)	0.25	18.03
				RuAdaptLit (ru)	34.78	32.61
				PorSimples (pt-br)	5.24	32.87

1. Matching script and language improve transfer performance
2. Domain match can help regardless of script
3. Russian is a good candidate language for crosslingual transfer

Fewshot Experiments

N example sentence pairs from training set

Original: "[EXAMPLE 1 ORIGINAL]"

Simple: "[EXAMPLE 1 SIMPLIFICATION]"

...

Original: "[EXAMPLE N ORIGINAL]"

Simple: "[EXAMPLE N ORIGINAL]"

Test sentence from test set

Original: "[TEST ORIGINAL]"

Simple: "

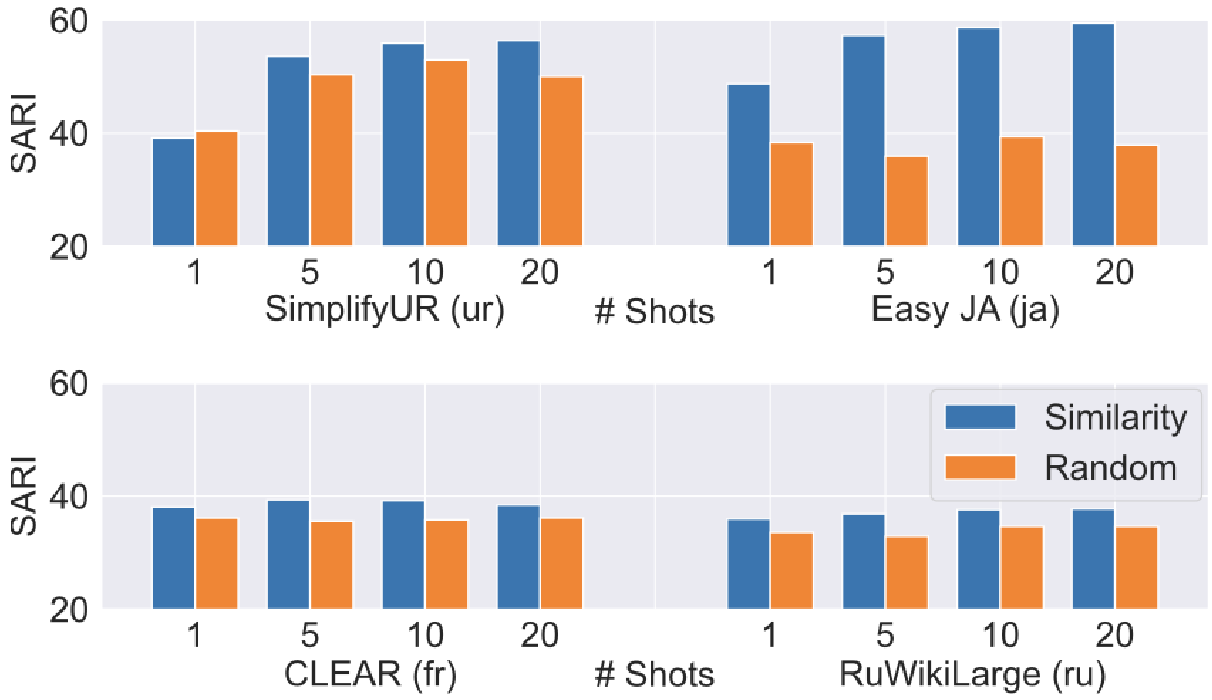
But how do we pick these examples?

2 Approaches:

- Random sample from training set ("Random")
- K-nearest neighbors search on LASER embedding of original sentence ("Similarity")
  - Ideally LASER embeddings will measure something about the semantics and syntax so we can find similar sentences to act as examples in the prompt

Georgia Tech

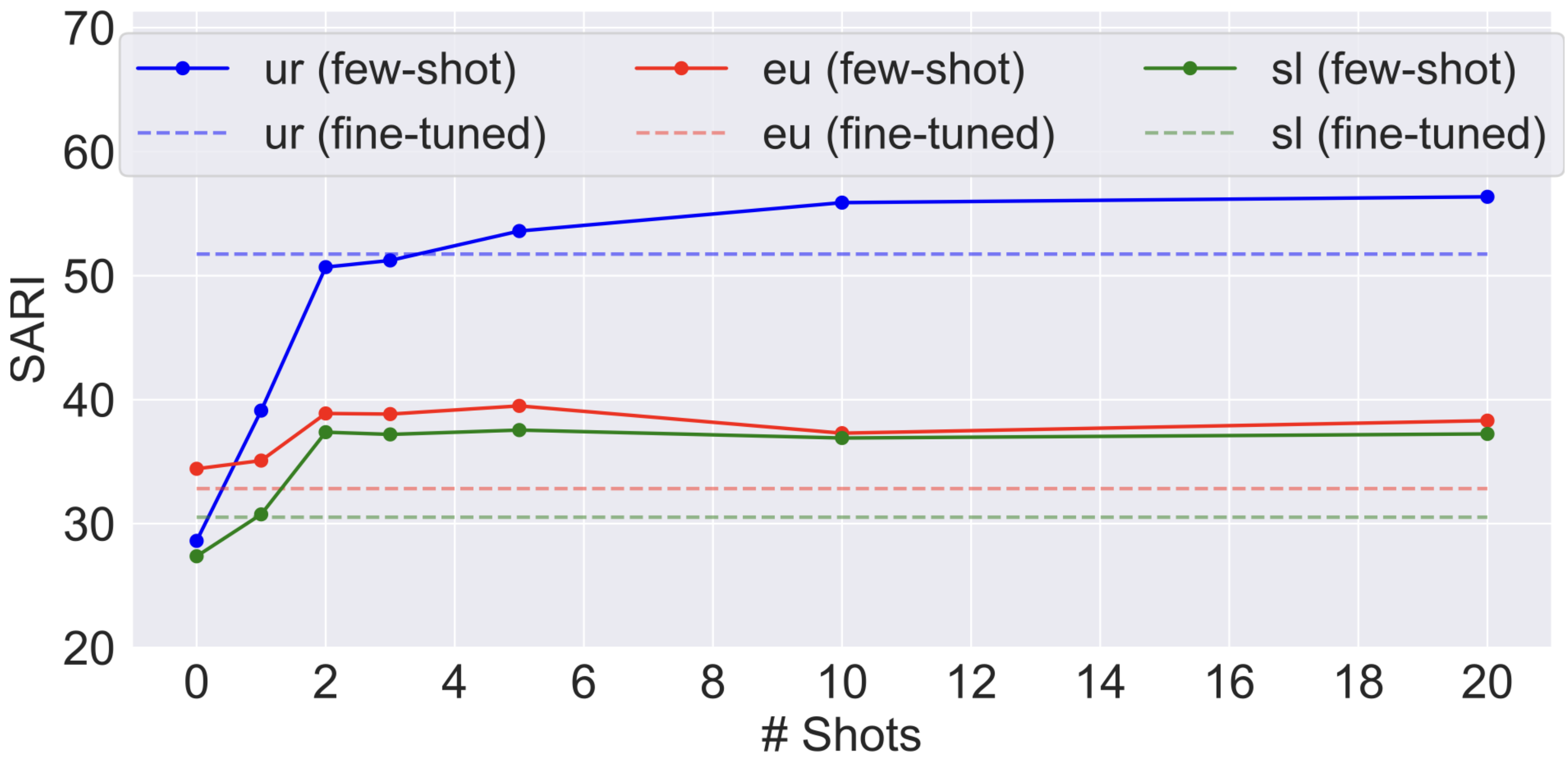
Semantic Similarity vs Random Sample Prompting



1. Similarity based example selection is consistently better than Random sampling

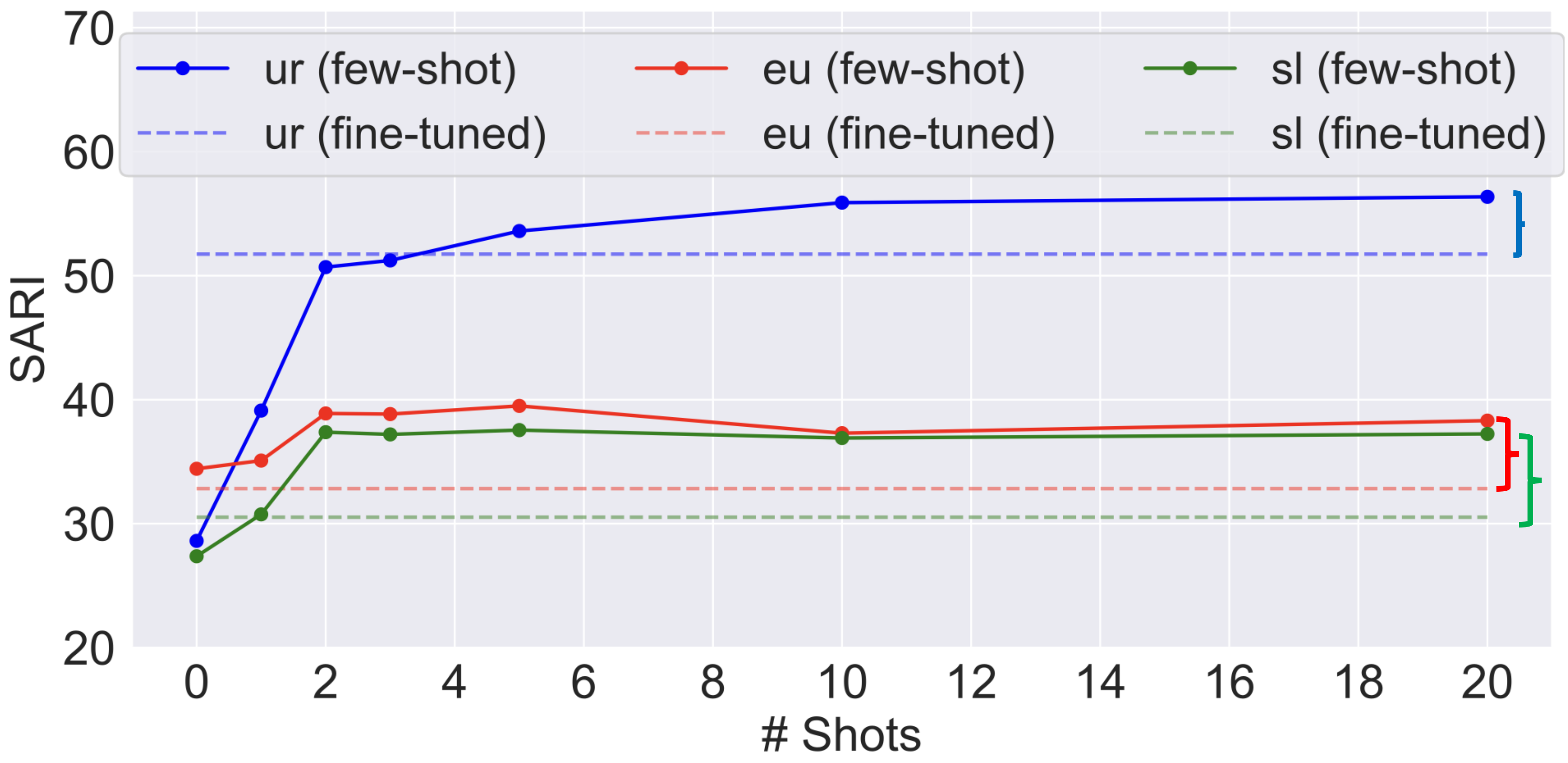
Georgia Tech

Fewshot Experiments



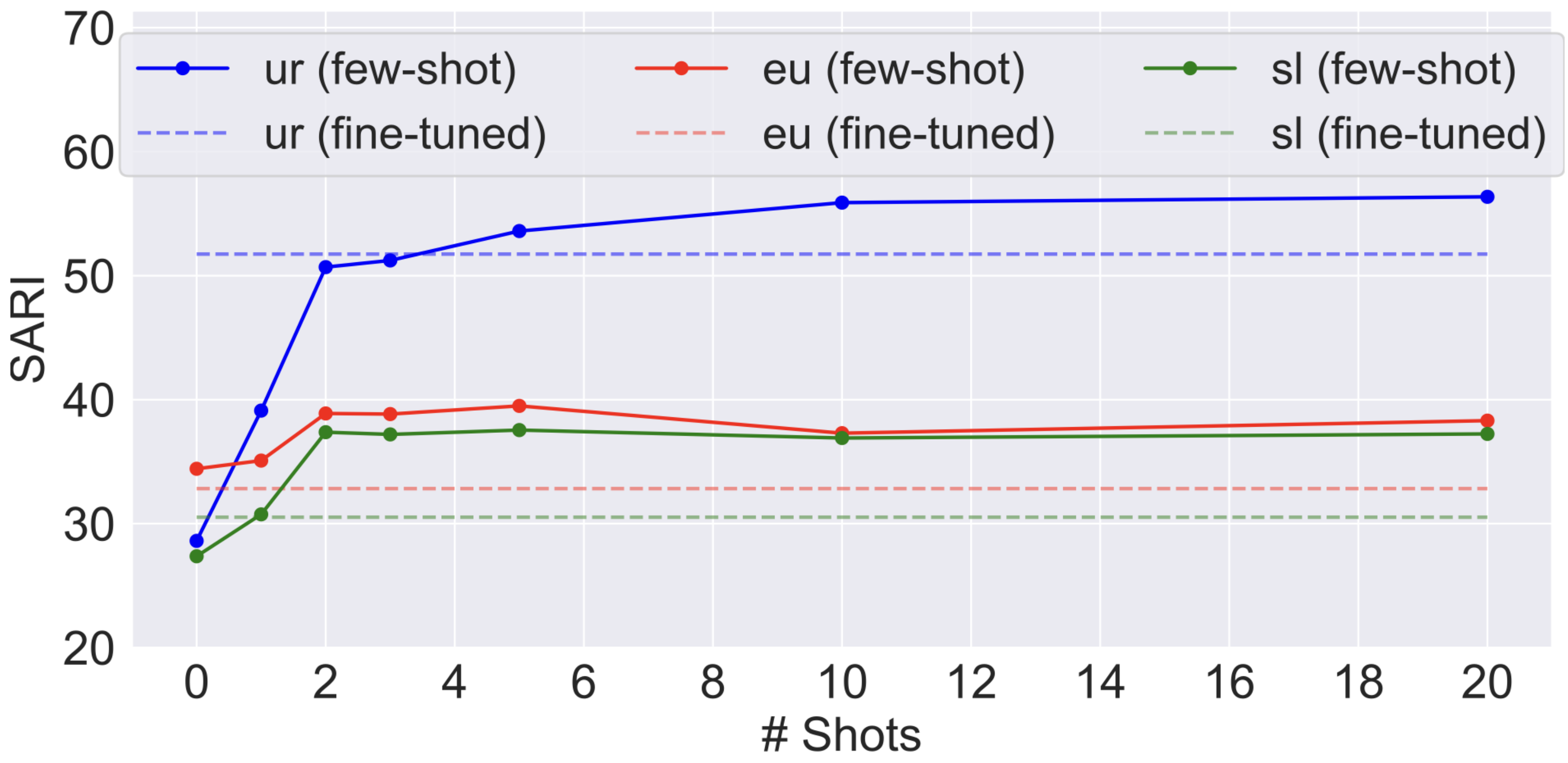
Georgia Tech

Fewshot Experiments



1. Fewshot prompting works better than finetuning for low resource languages

Fewshot Experiments



1. Fewshot prompting works better than finetuning for low resource languages

2. More examples improve performance, but past 5 there is marginal benefit.

Manual Evaluation

Adequacy (is the meaning preserved?)	
1:	The subject of the sentence has changed entirely and is entirely unrelated
2:	The meaning has been seriously altered (negated or changed)
3:	**Two or more important pieces** of information have been added or removed
4:	Meaning is similar but **one** piece of information has been added or removed
5:	Meaning is preserved aside from minor unimportant information
Fluency (is the simplification eloquent/grammatical?)	
1:	The simplification is completely unreadable
2:	The simplification suffers from many serious grammar issues (nearly unreadable)
3:	The simplification has **two or more** grammatical mistakes
4:	The simplification has a minor grammatical issue or is written strangely in one place
5:	The simplification is perfectly eloquent as if written by a human
Simplicity (is the simplification actually simpler?)	
1:	The simplification is actually harder to understand (ex. more complex terms used)
2:	The simplification is about the same difficulty as the original
3:	The simplification is mildly simpler, but this simplification does not help readability
4:	The simplification is actually simpler
5:	The simplification is vastly simpler and could help someone better understand

Table 9: Manual evaluation key provided to annotators

Manual Evaluation

2 English, 2 Russian, 2 Italian, 2 Urdu (20 sentences per model)

	English (ASSET)			Russian (RuAdaptLit)			Italian (Terence)			Urdu (SimplifyUR)		
	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity
Reference	$4.60_{\pm 0.22}$	$4.85_{\pm 0.11}$	$4.13_{\pm 0.26}$	$3.70_{\pm 0.47}$	$4.45_{\pm 0.32}$	$2.50_{\pm 0.24}$	$4.73_{\pm 0.55}$	$4.88_{\pm 0.33}$	$2.98_{\pm 1.37}$	$4.83_{\pm 0.38}$	$5.00_{\pm 0.00}$	$4.25_{\pm 1.17}$
mT5 Single	$4.45_{\pm 0.25}$	$4.95_{\pm 0.07}$	$3.00_{\pm 0.34}$	$4.50_{\pm 0.31}$	$4.78_{\pm 0.19}$	$2.25_{\pm 0.15}$	$1.23_{\pm 0.73}$	$1.15_{\pm 0.66}$	$1.15_{\pm 0.66}$	$2.38_{\pm 1.37}$	$2.10_{\pm 1.32}$	$1.10_{\pm 0.30}$
mT5 Joint Language	$4.65_{\pm 0.19}$	$4.98_{\pm 0.05}$	$3.38_{\pm 0.25}$	$4.78_{\pm 0.26}$	$5.00_{\pm 0.00}$	$2.48_{\pm 0.24}$	$4.78_{\pm 0.48}$	$4.83_{\pm 0.55}$	$2.55_{\pm 1.01}$	—	—	—
mT5 Joint All	$4.64_{\pm 0.18}$	$4.93_{\pm 0.08}$	$2.94_{\pm 0.21}$	$4.23_{\pm 0.38}$	$4.85_{\pm 0.21}$	$2.75_{\pm 0.32}$	$4.18_{\pm 1.20}$	$4.65_{\pm 0.70}$	$2.50_{\pm 1.09}$	$4.25_{\pm 1.13}$	$4.88_{\pm 0.65}$	$2.95_{\pm 1.30}$
mT5 English Transfer	—	—	—	$2.18_{\pm 0.49}$	$1.70_{\pm 0.40}$	$1.18_{\pm 0.12}$	$1.73_{\pm 1.45}$	$1.88_{\pm 1.54}$	$1.25_{\pm 0.59}$	$1.83_{\pm 1.08}$	$1.40_{\pm 0.63}$	$1.00_{\pm 0.00}$
mT5 Russian Transfer	$4.25_{\pm 0.33}$	$3.93_{\pm 0.16}$	$2.63_{\pm 0.30}$	—	—	—	$4.53_{\pm 1.01}$	$4.70_{\pm 0.72}$	$2.35_{\pm 0.92}$	$3.70_{\pm 1.73}$	$3.60_{\pm 1.85}$	$1.60_{\pm 0.50}$
BLOOM 5 Shot (Rand)	$4.63_{\pm 0.25}$	$4.75_{\pm 0.18}$	$3.10_{\pm 0.40}$	$4.65_{\pm 0.27}$	$4.78_{\pm 0.22}$	$2.33_{\pm 0.20}$	$4.33_{\pm 1.02}$	$4.53_{\pm 0.85}$	$2.45_{\pm 1.11}$	$4.95_{\pm 0.22}$	$4.93_{\pm 0.35}$	$3.28_{\pm 1.69}$
BLOOM 5 Shot (Sim)	$4.63_{\pm 0.22}$	$4.80_{\pm 0.13}$	$2.88_{\pm 0.32}$	$4.63_{\pm 0.28}$	$4.95_{\pm 0.07}$	$2.43_{\pm 0.24}$	$4.00_{\pm 1.47}$	$4.58_{\pm 1.01}$	$2.38_{\pm 1.19}$	$4.83_{\pm 0.68}$	$4.85_{\pm 0.70}$	$3.28_{\pm 1.71}$

Manual Evaluation

2 English, 2 Russian, 2 Italian, 2 Urdu (20 sentences per model)

	English (ASSET)			Russian (RuAdaptLit)			Italian (Terence)			Urdu (SimplifyUR)		
	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity
Reference	4.60±0.22	4.85±0.11	4.13±0.26	3.70±0.47	4.45±0.32	2.50±0.24	4.73±0.55	4.88±0.33	2.98±1.37	4.83±0.38	5.00±0.00	4.25±1.17
mT5 Single	4.45±0.25	4.95±0.07	3.00±0.34	4.50±0.31	4.78±0.19	2.25±0.15	1.23±0.73	1.15±0.66	1.15±0.66	2.38±1.37	2.10±1.32	1.10±0.30
mT5 Joint Language	4.65±0.19	4.98±0.05	3.38±0.25	4.78±0.26	5.00±0.00	2.48±0.24	4.78±0.48	4.83±0.55	2.55±1.01	—	—	—
mT5 Joint All	4.64±0.18	4.93±0.08	2.94±0.21	4.23±0.38	4.85±0.21	2.75±0.32	4.18±1.20	4.65±0.70	2.50±1.09	4.25±1.13	4.88±0.65	2.95±1.30
mT5 English Transfer	—	—	—	2.18±0.49	1.70±0.40	1.18±0.12	1.73±1.45	1.88±1.54	1.25±0.59	1.83±1.08	1.40±0.63	1.00±0.00
mT5 Russian Transfer	4.25±0.33	3.93±0.16	2.63±0.30	—	—	—	4.53±1.01	4.70±0.72	2.35±0.92	3.70±1.73	3.60±1.85	1.60±0.50
BLOOM 5 Shot (Rand)	4.63±0.25	4.75±0.18	3.10±0.40	4.65±0.27	4.78±0.22	2.33±0.20	4.33±1.02	4.53±0.85	2.45±1.11	4.95±0.22	4.93±0.35	3.28±1.69
BLOOM 5 Shot (Sim)	4.63±0.22	4.80±0.13	2.88±0.32	4.63±0.28	4.95±0.07	2.43±0.24	4.00±1.47	4.58±1.01	2.38±1.19	4.83±0.68	4.85±0.70	3.28±1.71

1. Fewshot outperforms Finetuned

Manual Evaluation

2 English, 2 Russian, 2 Italian, 2 Urdu (20 sentences per model)

	English (ASSET)			Russian (RuAdaptLit)			Italian (Terence)			Urdu (SimplifyUR)		
	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity	Adequacy	Fluency	Simplicity
Reference	4.60±0.22	4.85±0.11	4.13±0.26	3.70±0.47	4.45±0.32	2.50±0.24	4.73±0.55	4.88±0.33	2.98±1.37	4.83±0.38	5.00±0.00	4.25±1.17
mT5 Single	4.45±0.25	4.95±0.07	3.00±0.34	4.50±0.31	4.78±0.19	2.25±0.15	1.23±0.73	1.15±0.66	1.15±0.66	2.38±1.37	2.10±1.32	1.10±0.30
mT5 Joint Language	4.65±0.19	4.98±0.05	3.38±0.25	4.78±0.26	5.00±0.00	2.48±0.24	4.78±0.48	4.83±0.55	2.55±1.01	—	—	—
mT5 Joint All	4.64±0.18	4.93±0.08	2.94±0.21	4.23±0.38	4.85±0.21	2.75±0.32	4.18±1.20	4.65±0.70	2.50±1.09	4.25±1.13	4.88±0.65	2.95±1.30
mT5 English Transfer	—	—	—	2.18±0.49	1.70±0.40	1.18±0.12	1.73±1.45	1.88±1.54	1.25±0.59	1.83±1.08	1.40±0.63	1.00±0.00
mT5 Russian Transfer	4.25±0.33	3.93±0.16	2.63±0.30	—	—	—	4.53±1.01	4.70±0.72	2.35±0.92	3.70±1.73	3.60±1.85	1.60±0.50
BLOOM 5 Shot (Rand)	4.63±0.25	4.75±0.18	3.10±0.40	4.65±0.27	4.78±0.22	2.33±0.20	4.33±1.02	4.53±0.85	2.45±1.11	4.95±0.22	4.93±0.35	3.28±1.69
BLOOM 5 Shot (Sim)	4.63±0.22	4.80±0.13	2.88±0.32	4.63±0.28	4.95±0.07	2.43±0.24	4.00±1.47	4.58±1.01	2.38±1.19	4.83±0.68	4.85±0.70	3.28±1.71

1. Fewshot outperforms Finetuned
2. Russian outperforms English as a transfer language

Conclusion

Conclusion

- We release the MultiSim Benchmark with **1.7 million sentence pairs** spanning **12 diverse languages**.

- We show that the large resource can help improve performance in text simplification for low resource languages.

- We make recommendations for best practices to use this data based on empirical evidence.
  - Match Script and Domain for transfer
  - Russian is a good pivot language

Georgia Tech

Acknowledgements

- Thank you to Tarek Naous for his excellent guidance in designing and executing the experiments involving multilingual language models.

- Thank you to Dr. Wei Xu for her continuous support and advising on this project.

- This project would not have been possible without their support!

Georgia Tech

Thank you!

Questions?

Appendix

MultiSim Splits

Language	Dataset	#train	#test	#dev
Open Source				
English	WikiAuto	576,126	5,012	5,012
	ASSET*	20,000	3,590	0
French	WikiLargeFR*	296,402	359	992
	CLEAR*	4,196	100	300
German	GEOLino	958	122	118
	TextCompDE	200	25	25
Italian	PaCCSS-IT	60,485	1,267	1,254
	Terence	809	101	102
	AdminIT	588	73	75
	Simpitiki	460	56	59
	Teacher	136	17	17
Japanese	Easy JA	48,000	1,000	1,000
	Easy JA Ext*	34,269	731	0
Russian	RuAdapt Ency	7,782	982	965
	RSSE Corpus*	3,406	3,398	0
	RuAdapt Fairy	248	31	31
Slovene	SloTS*	749	96	94
Urdu	SimplifyUR	594	68	74
	Total	1,055,408	17,028	10,118
Dataloaders Available (Data on Request)				
Basque	CBST	361	46	46
Br Portuguese	PorSimples	6,290	790	784
Danish	DSim Corpus	45,885	997	1,005
English	Newsela EN	291,969	991	1,008
German	German News	8,186	1,024	1,023
Russian	RuWikiLarge*	246,978	365	768
	RuAdapt Lit	22,152	1,000	1,000
Spanish	Newsela ES	30,910	1,001	1,001
	Simplext	737	92	93
	Total	653,468	6,306	6,728

Table 2: MULTISIM splits. *Original splits preserved

Georgia Tech

Basic Stats

Corpus	Lang	Vocab Size		Token Count		Avg Tok/Sent		Avg Char/Tok		Avg Sent/Doc	
		orig ↑	simp ↓	orig ↑	simp ↓	orig ↑	simp ↓	orig ↑	simp ↓	orig	simp
DSim	da	57,308	40,220	953,201	796,201	19.91	13.15	5.57	5.36	—	—
GEOLino	de	4,467	4,266	19,185	17,889	16.01	14.93	**5.68**	**5.74**	—	—
German News A2	de	23,542	7,764	147,905	78,946	20.30	11.23	6.40	5.79	4.03	3.89
German News B2	de	25,039	10,473	160,188	93,283	20.14	12.75	6.43	5.99	4.96	4.56
Klexikon	de	706,243	55,868	15,240,505	1,239,694	19.77	12.80	6.42	5.53	266.53	33.48
Simple German	de	35,763	18,753	313,622	199,861	24.50	23.79	6.60	6.38	57.16	37.50
TextComplexityDE	de	3,068	2,760	7,485	7,092	29.94	28.37	6.78	6.62	10.87	10.87
ASSET	en	**11,998**	**19,320**	521,940	448,376	22.13	19.01	5.28	5.18	—	—
Newsela EN 0-1	en	68,972	61,115	2,187,046	1,881,631	23.95	19.83	5.08	5.06	48.52	50.41
Newsela EN 1-2	en	61,115	53,673	1,881,631	1,733,011	19.83	16.84	5.06	4.98	50.41	54.67
Newsela EN 2-3	en	53,673	42,879	1,733,011	1,458,744	16.84	13.93	4.98	4.86	54.67	55.65
Newsela EN 3-4	en	42,879	34,104	1,458,744	1,144,534	13.93	11.48	4.86	4.75	55.65	53.00
WikiAuto	en	2,009,681	419,496	265,352,569	22,170,411	26.16	17.86	5.19	4.96	—	—
Newsela ES 0-1	es	27,950	23,452	323,034	257,905	28.28	23.31	5.37	5.36	47.01	45.52
Newsela ES 1-2	es	23,452	20,582	257,905	225,659	23.31	19.35	5.36	5.31	45.52	48.00
Newsela ES 2-3	es	20,582	16,148	225,659	178,117	19.35	14.72	5.31	5.21	48.00	49.79
Newsela ES 3-4	es	16,148	11,695	178,117	122,064	14.72	11.42	5.21	5.13	49.79	43.98
Simplext	es	8,071	3,191	38,731	25,409	34.96	14.59	5.47	5.34	5.74	9.03
CBST Intuitive	eu	1,697	1,586	4,575	4,447	19.98	14.53	**6.34**	**6.43**	76.33	102.00
CBST Structural	eu	1,697	1,654	**4,575**	**4,793**	19.98	16.82	6.34	6.29	76.33	95.00
Alector	fr	5,728	5,024	28,283	26,179	22.99	21.96	4.78	4.73	15.57	15.09
CLEAR	fr	11,743	11,205	119,465	118,212	25.99	25.72	**5.72**	**5.73**	—	—
WikiLargeFR	fr	205,933	173,827	8,763,745	6,384,020	28.54	20.70	5.03	4.94	—	—
AdminIT	it	3,420	3,394	29,581	28,784	38.07	37.72	6.08	5.90	—	—
PaCCSS-IT	it	10,478	9,853	580,389	519,211	9.21	8.24	**4.75**	**4.79**	—	—
SimpitikiWiki	it	9,188	9,175	41,899	41,375	72.87	71.96	5.60	5.60	—	—
Teacher	it	1,485	1,061	4,225	3,367	20.71	17.27	4.89	4.76	11.33	10.83
Terence	it	3,681	3,219	19,455	18,881	18.80	17.81	5.13	5.04	32.34	33.12
Easy Japanese	ja	10,331	3,401	**489,302**	**517,651**	**9.79**	**10.35**	1.51	1.49	—	—
Easy Japanese Ext	ja	18,888	5,305	**433,341**	**503,035**	**12.38**	**14.37**	1.55	1.49	—	—
PorSimples Natural	pt-br	9,983	9,527	**64,610**	**65,174**	20.97	13.52	**5.39**	**5.53**	20.01	31.31
PorSimples Strong	pt-br	**9,527**	**9,601**	**65,174**	**65,552**	13.52	12.25	**5.53**	**5.57**	31.31	34.76
RSSE Corpus	ru	**16,467**	**24,307**	138,319	95,067	20.33	13.97	6.73	6.54	—	—
RuAdapt Ency A-B	ru	4,609	3,842	11,085	9,804	12.58	9.96	6.08	5.84	14.44	16.13
RuAdapt Ency A-C	ru	4,927	3,844	11,931	9,809	13.59	9.96	6.05	5.84	14.16	15.89
RuAdapt Ency B-C	ru	27,268	26,200	113,817	110,463	14.28	13.37	6.10	6.08	29.96	31.06
RuAdapt Fairytales	ru	1,688	1,512	4,391	4,289	14.16	10.62	**5.08**	**5.32**	34.44	44.89
RuAdapt Literature	ru	55,321	42,655	368,499	327,228	15.26	11.58	5.14	5.08	168.90	197.62
RuWikiLarge	ru	331,063	275,644	5,760,207	4,540,009	20.70	15.68	5.95	5.79	—	—
SloTS	sl	5,871	2,723	21,804	10,646	18.46	8.27	4.72	4.44	—	—
SimplifyUR	ur	**1,469**	**1,475**	6,580	6,561	8.94	8.91	4.25	4.22	—	—

Table 6: Basic statistics about all of the non-English corpora we analyzed. Typically the vocab size, token count, average tokens per sentence, and average characters per token all decrease from original to simplified texts. Outliers of this trend are highlighted in bold.

Edit Operations

Corpus	Deleted 1:0 (%)	Split 1:n (%)	Same 1:1 (%)	Changed 1:1 (%)	Merged n:1 (%)	Inserted 0:1 (%)
English						
Newsela EN 0-1	22.7	12.6	39.5	25.2	0.0	12.2
Newsela EN 1-2	17.4	13.4	36.4	32.8	0.0	10.7
Newsela EN 2-3	27.2	11.8	23.5	37.5	0.0	15.9
Newsela EN 3-4	33.0	10.4	21.2	35.4	0.0	17.8
WikiAuto	94.2	0.9	1.2	3.7	0.0	44.8
Russian						
RuAdapt Fairytales	0.0	22.6	4.2	73.2	0.0	0.0
RuAdapt Ency B-C	0.0	3.5	79.3	17.2	0.0	0.0
RuAdapt Ency A-C	0.0	10.8	31.8	57.4	0.0	0.0
RuAdapt Ency A-B	0.0	10.7	35.4	53.9	0.0	0.0
RuAdapt Literature	0.0	11.8	36.6	51.6	0.0	0.0
Italian						
Terence	0.7	4.1	35.4	57.0	2.9	0.4
Teacher	6.9	9.3	8.3	59.3	16.2	1.5
Spanish						
Newsela ES 0-1	29.1	20.0	19.1	31.6	0.2	0.7
Newsela ES 1-2	18.3	19.8	24.2	37.7	0.1	0.5
Newsela ES 2-3	27.5	22.9	13.3	36.0	0.3	0.4
Newsela ES 3-4	38.5	19.1	11.2	31.1	0.2	0.3
Simplext	16.2	32.2	3.5	47.4	0.7	19.3
German						
TextComplexityDE	0.0	0.0	0.4	99.6	0.0	0.0
German News A2	0.0	22.5	0.8	37.6	39.1	0.0
German News B2	0.0	23.0	1.4	33.2	42.4	0.0
Brazilian Portuguese						
PorSimples Natural	0.6	38.7	22.1	38.3	0.3	0.8
PorSimples Strong	0.2	9.9	73.7	16.0	0.1	0.1
Basque						
Structural CBST	0.0	22.3	24.0	51.5	2.2	0.0
Intuitive CBST	0.0	25.8	27.1	45.9	1.3	0.0

Table 7: Edit operations for all document-level corpora with sentence alignment. All operations are reported as a percentage of original sentences besides "inserted" which is reported as a percentage of simplified sentences.

Edit Distances

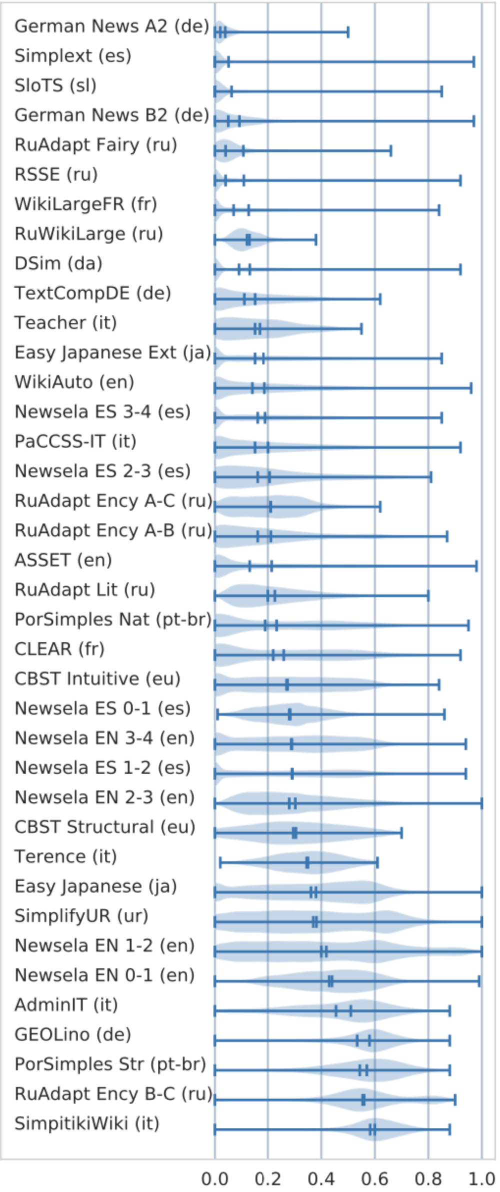


Figure 5: Violin Plots showing the minimum, maximum, mean, and median values for edit distances for all of the sentences in each corpus. Distributions estimated using Gaussian kernel density estimation.

Georgia Tech

Document Level Compression

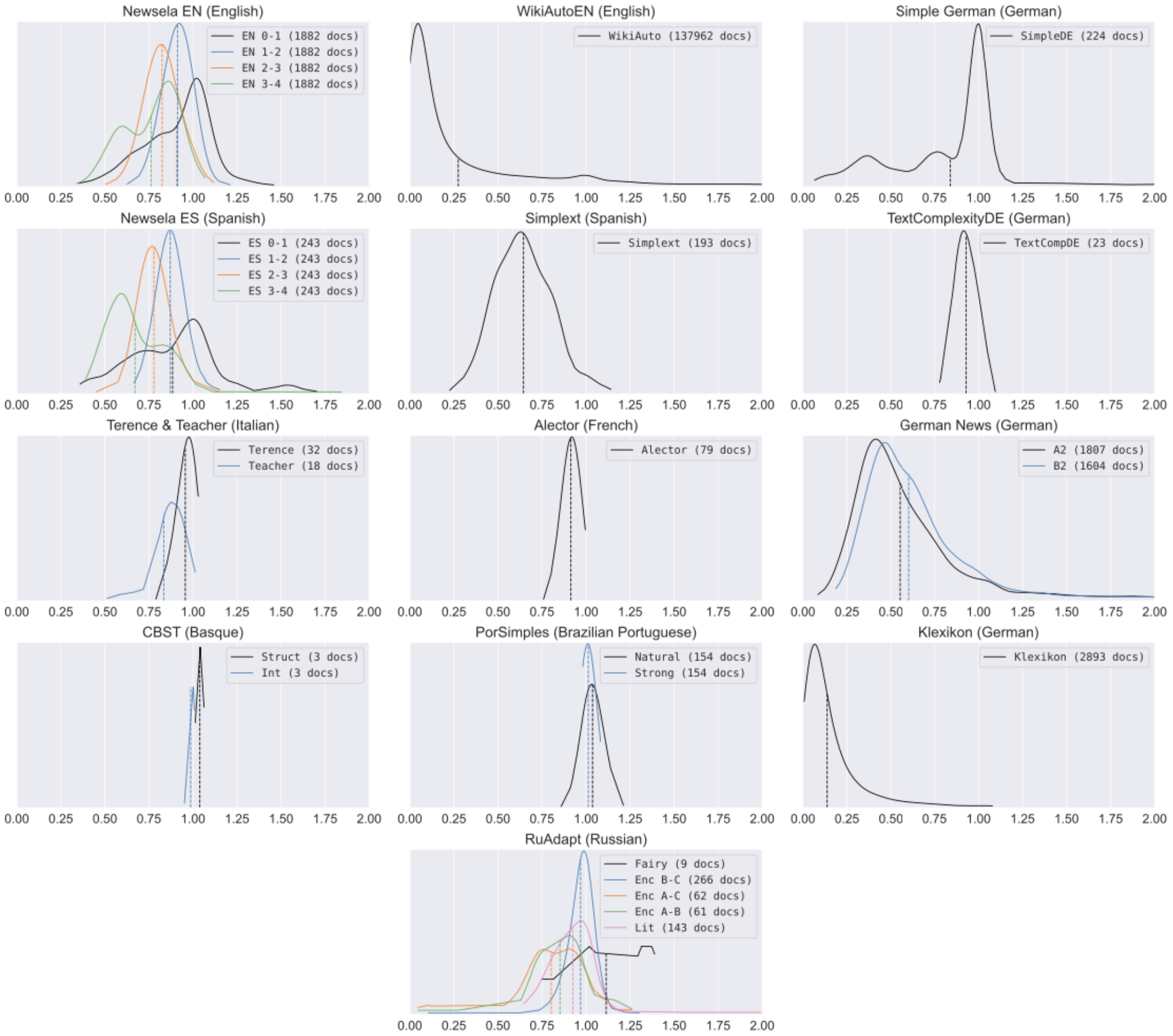


Figure 6: Distribution of document-level compression ratio for document-aligned corpora, smoothed by Gaussian kernel density estimation. Means are marked by dashed lines.

Fewshot Experiments

Lang →		en			ru					da
Approach	# shots	ASSET	Newsela EN	WikiAuto EN	RuWikiLarge	RSSE	RuAdapt Ency	RuAdapt Fairy	RuAdapt Lit	DSim
Fine-Tuned	NA	35.98	38.60	42.46	32.01	31.66	26.42	34.79	41.75	31.40
Zero Shot	0	35.52	33.22	34.73	31.47	20.09	33.20	12.74	30.95	35.84
Semantic Similarity	1	36.21	35.29	40.24	35.95	30.18	40.21	33.49	37.05	38.27
	2	36.37	37.70	40.97	36.93	29.75	42.22	35.51	37.61	38.84
	3	36.53	37.71	41.66	37.22	29.79	40.76	37.07	39.06	38.09
	5	36.79	38.82	42.25	36.84	31.08	41.01	38.43	39.89	37.57
	10	36.19	38.65	42.66	37.59	31.33	39.45	40.93	40.44	31.71
	20	36.80	39.26	42.83	37.71	31.22	39.54	38.95	40.32	29.88
Random Sampling	1	35.21	34.02	35.16	33.61	28.23	33.45	20.89	32.26	34.96
	2	35.77	34.87	36.40	34.20	28.44	34.07	24.24	33.14	35.29
	3	35.37	34.31	36.21	34.19	29.40	34.24	22.84	33.41	35.39
	5	35.89	34.39	36.00	32.89	29.30	33.17	25.91	33.45	34.81
	10	36.10	35.35	36.86	34.63	28.69	33.19	29.16	33.92	30.15
	20	36.21	34.73	37.14	34.63	29.60	33.71	31.83	33.89	27.98

Lang →		de			it					pt-br
Approach	# shots	German News	TextCompDE	GEOLino	PaCCSS-IT	Terence	AdminIT	Simpitiki	Teacher	PorSimples
Fine-Tuned	NA	36.04	30.26	26.44	57.30	23.92	23.42	4.11	29.84	31.54
Zero Shot	0	32.48	32.26	29.59	35.42	35.91	32.43	18.43	28.75	35.38
Semantic Similarity	1	36.19	37.63	38.16	51.42	34.95	37.06	27.73	33.97	36.72
	2	36.68	38.60	39.65	49.15	37.25	36.69	26.42	34.14	37.46
	3	36.78	41.03	39.44	48.00	34.95	35.67	27.06	29.41	38.85
	5	37.79	38.81	39.5	45.48	35.94	38.16	26.94	39.10	38.96
	10	37.69	38.93	39.7	37.31	35.39	35.21	27.20	32.62	41.34
	20	36.76	38.93	39.44	33.45	35.17	35.21	27.73	33.46	39.94
Random Sampling	1	32.58	34.94	35.89	38.84	33.60	33.31	21.96	33.94	37.68
	2	34.09	35.37	36.11	39.11	34.15	34.77	23.79	25.01	36.82
	3	34.92	34.13	35.22	38.51	33.96	35.98	25.22	31.41	36.57
	5	34.71	36.68	34.5	37.41	32.01	34.24	25.01	32.30	36.53
	10	35.58	38.07	35.42	35.01	31.60	35.67	25.04	30.82	35.93
	20	35.53	38.07	34.62	30.29	34.38	35.67	25.04	34.39	35.31

Lang →		fr		sl	ja		es		ur	eu
Approach	# shots	WikiLargeFR	CLEAR	SloTS	Easy JA	Easy JA Ext	NewselaES	Simplext	SimplifyUR	CBST
Fine-Tuned	NA	35.20	34.86	36.56	67.36	43.15	29.89	35.62	33.18	24.26
Zero Shot	0	35.71	35.75	27.37	41.71	30.53	34.15	25.36	28.61	34.43
Semantic Similarity	1	36.29	38.06	30.75	48.71	46.08	37.07	32.50	39.12	35.09
	2	35.22	39.03	37.38	54.89	49.39	36.90	38.05	50.69	38.89
	3	36.40	40.16	37.20	55.38	47.01	38.18	39.28	51.23	38.84
	5	36.75	39.34	37.55	57.29	49.30	38.12	40.26	53.60	39.50
	10	36.33	39.21	36.91	58.67	47.50	38.42	39.90	55.89	37.30
	20	37.72	38.45	37.24	59.42	46.55	38.42	39.75	56.36	38.31
Random Sampling	1	35.64	36.12	27.43	38.35	40.70	34.25	29.57	40.38	36.40
	2	36.67	34.64	31.83	37.85	41.38	34.55	33.51	49.00	36.00
	3	36.07	36.92	33.95	35.90	40.04	34.16	35.56	49.28	36.10
	5	36.17	35.54	33.60	35.92	42.55	33.64	36.81	50.34	37.39
	10	36.70	35.80	34.18	39.34	42.50	34.27	37.40	52.93	36.83
	20	35.80	36.13	35.30	37.89	42.11	33.91	39.27	50.02	38.86

Table 8: SARI Scores for BLOOM Fewshot Experiments

Georgia Tech

Prompting

We used a temperature of 1.0, and a repetition penalty of 0.0

Original: "[EXAMPLE 1 ORIGINAL]"
Simple: "[EXAMPLE 1 SIMPLIFICATION]"

Original: "[EXAMPLE N ORIGINAL]"
Simple: "[EXAMPLE N ORIGINAL]"

Original: "[TEST ORIGINAL]"
Simple: "

Preprocessing

- We compute sentence BLEU scores between the original and reference simplifications. For any sentence pairs with an S-BLEU score below 10 or above 70 we remove it from the training set.

- Following Martin et al., 2020 we also add control tokens to the input sentences.
  - We include a character length compression token <NC_[\#]>
  - a Levenshtein similarity token <LS_[\#]>
  - a dependency tree depth ratio token \texttt{<DR_[\#]>}
  - a word frequency rank token \texttt{<WR_[\#]>}
  - For each token we compute the respective measure on both the original and simple sentences and include the ratio between [0.05, 2] in increments of 0.05.

Training

- We used the sentence piece tokenizer
- Limited inputs/targets to a length of 128 tokens.
- Learning rate of 5e-5
- AdamW Optimizer
- Beam search with 4 beams
- Train batch size was set to 8.
- We train for 5 epochs.
- For any dataset without a development set we removed 10% of the training set up to 1000 sentences to create our own dev set.
- Training was performed on three NVIDIA A40 GPUs.

Examples

DSim (Danish)		
Original	Stigende vandstand i floderne i det østlige Tjekkiet forvandlede i aftes hundredvis af boliger i området til dødsfælder .	Rising water levels in the rivers in the eastern Czech Republic last night turned hundreds of homes in the area into death traps.
Simple	I det østlige Tjekkiet stiger vandstanden i floderne .	In the eastern Czech Republic, the water level in the rivers is rising.
EasyJA (Japanese)		
Original	君が言ったことで、僕はびっくりした。	What you said surprised me.
Simple	あなたが言ったことで、私は驚いた。	What you said surprised me.
EasyJAExt (Japanese)		
Original	彼の不注意にはあきれてしまった。	I was appalled at his carelessness.
Simple	彼の不注意には言葉を失う。	His carelessness leaves me speechless.
GEOLino (German)		
Original	Denn sie sind zwar mutig, aber durchaus nicht lebensmüde.	Because they are courageous, but by no means tired of life.
Simple	Denn sie sind zwar mutig, aber nicht lebensmüde.	Because they are courageous, but not tired of life.
GermanNews (German)		
Original	Jedes Kalb erhält spätestens sieben Tage nach der Geburt eine eindeutig identifizierbare Lebensnummer, die in Form von Ohrmarken beidseitig eingezogen wird.	Each calf receives a clearly identifiable life number no later than seven days after birth, which is recorded on both sides in the form of ear tags.
Simple	In Österreich bekommt jedes Kalb kurz nach der Geburt eine Nummer	In Austria, every calf is given a number shortly after birth.