

Domain		Source		
Sub-Domain	ar	en	hi	
WIKIPEDIA	wikipedia.com	wikipedia.com	wikipedia.com	
NEWS ARTICLES	(Alfonse and Gawich, 2022)	(Misra, 2022)	—	
RESEARCH				
Law	spu.sharjah.ac.ae	elgaronline.com	library.bjp.org	
Politics	jcopolity.uobaghdad.edu.iq	tandfonline.com	journal.ijarms.org	
Medical	—	onlinelibrary.wiley.com	—	
Literature	—	jstor.org/journal/jmodelite	hindijournal.com	
Economics	asjp.cerist.dz/index.php/en	aeaweb.org	journal.ijarms.org	
Science & Engineering	—	arxiv.org	—	
LITERATURE	hindawi.org/books/	gutenberg.org	Public Domain Books	
TEXTBOOKS	hindawi.org/books/	open.umn.edu	ncert.nic.in	
LEGAL				
Constitutions	presidency.gov.lb	constitutioncenter.org	legislative.gov.in	
Judicial Rulings	—	law.cornell.edu/supremecourt	HLDC (Kapoor et al., 2022)	
UN Parliament	United Nations Parallel Corpus (Ziems et al., 2016)			—
USER REVIEWS				
Products	(ElSahar and El-Beltagy, 2015)	MARC (Keung et al., 2020)	(Akhtar et al., 2016)	
Books	LABR (Aly and Atiya, 2013)	(Wan et al., 2019)	—	
Movies	—	JMURv1 (Chatterjee et al., 2021)	(HindiMovieReviews)	
Hotels	(ElSahar and El-Beltagy, 2015)	(Ray et al., 2021)	—	
Restaurants	(ElSahar and El-Beltagy, 2015)	(TripAdvisor)	—	
DIALOGUE				
Open-domain	ArabicED (Naous et al., 2020)	DailyDialog (Li et al., 2017)	MDIA (Zhang et al., 2022)	
Negotiation	—	CraigslistBargain (He et al., 2018)	—	
Task-oriented	xSID (van der Goot et al., 2021)	xSID (van der Goot et al., 2021)	HDRS (Malviya et al., 2021)	
FORUMS				
Reddit	Reddit Dump			
QA Websites	CQA-MD (Nakov et al., 2016)	quora.com (Quora.com, 2017)	(Howard et al., 2021)	
StackOverflow	—	(Tabassum et al., 2020)	—	
SOCIAL MEDIA				
Twitter	Stanceosaurus (Zheng et al., 2022)			
POLICIES				
Contracts	ejar.sa	honeybook.com	—	
Olympic Rules	—	resources.specialolympics.org/translated-resources	—	
Code of Conduct	—	fatimafellowship.com	lonza.com	
GUIDES				
User Manuals	—	samsung.com/us/support/downloads	—	
Online Tutorials	ar.wikihow.com	wikihow.com	hi.wikihow.com	
Cooking Recipes	ar.wikibooks.org	en.wikibooks.org	—	
Code Documentation	—	mathworks.com	—	
CAPTIONS				
Images	(ElJundi et al., 2020)	Flickr30K (Plummer et al., 2015)	(Rathi, 2020)	
Videos	—	Vatex (Wang et al., 2019)	(Singh et al., 2022)	
Movies	—	OpenSubtitles2016 (Lison and Tiedemann, 2016)	—	
YouTube	—	youtube.com	—	
MEDICAL TEXT				
Clinical Reports	—	i2b2/VA (Uzuner et al., 2011)	—	
DICTIONARIES	almaany.com	dictionary.com	—	
ENTERTAINMENT				
Jokes	(Al-Khalifa et al., 2022)	(Weller and Seppi, 2019)	123hindijokes.com	
FINANCE	—	(Malo et al., 2014)	—	
SPEECH				
Ted Talks	ted.com/talks	ted.com/talks	ted.com/talks	
Public Speech	state.gov/translations/arabic	whitehouse.gov	—	
STATEMENTS				
Rumours	Stanceosaurus (Zheng et al., 2022)			
Quotes	arabic-quotes.com	goodreads.com/quotes	storyshala.in	
POETRY	aldiwan.net	poetryfoundation.org	hindionlinejankari.com	
LETTERS	—	oflosttime.com	—	

Table 16: Dataset Sources (1/2). (—) denotes that no resource was found in the particular language.

Domain	Source	
Sub-Domain	fr	ru
WIKIPEDIA	wikipedia.com	wikipedia.com
RESEARCH	hal.science	ruscorpora.ru
LITERATURE	gutenberg.org	gutenberg.org
LEGAL		
Constitutions	legifrance.gouv.fr	constitution.ru
Judicial Rulings	—	supcourt.ru
UN Parliament	United Nations Parallel Corpus (Ziemski et al., 2016)	
USER REVIEWS		
Products	—	RuReviews (Smetanin and Komarov, 2019)
DIALOGUE		
Open-domain	MDIA (Zhang et al., 2022)	MDIA (Zhang et al., 2022)
Task-oriented	M-CID (Arora et al., 2020)	—
FORUMS		
Reddit		Reddit Dump
QA Websites	(d’Hoffschmidt et al., 2020)	(Efimov et al., 2020)
SOCIAL MEDIA		
Twitter	(Kozlowski et al., 2020)	RuSentiTweet (Smetanin, 2022)
POLICIES		
Contracts	cesu.urssaf.fr	blanker.ru
Olympic Rules	resources.specialolympics.org/translated-resources	
GUIDES		
User Manuals	samsung.com/us/support/downloads	manuals.plus/ru
Online Tutorials		wikihow.com
Cooking Recipes		wikibooks.org
CAPTIONS		
Images	(Schamoni et al., 2018)	
Videos	citevideo-captions-fr	—
Movies	OpenSubtitles2016 (Lison and Tiedemann, 2016)	
ENTERTAINMENT		
Jokes	—	(Jokes)
FINANCE	(Daudert and Ahmadi, 2019)	ruscorpora.ru
SPEECH		
Ted Talks	ted.com/talks	ted.com/talks
Public Speech	—	ruscorpora.ru
STATEMENTS		
Quotes	evene.lefigaro.fr	infoselection.ru
POETRY	poesie-francaise.fr	ruscorpora.ru
LETTERS	gutenberg.org	runivers.ru

Table 17: Dataset Sources (1/2). (—) denotes that no resource was found in the particular language.

Rank and Rate Sentences on Readability

Signed in as Anonymous [Sign out](#)

[View Instructions](#) [View Examples](#)

Batch ID: 500

Sentences

There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.

[- Context](#)

The company also sponsors other postretirement benefit (OPEB) plans that provide medical and dental benefits, as well as life insurance for some active and qualifying retired employees.

[+ Context](#)

I also had to taste my Mom's multi-grain pumpkin pancakes with pecan butter and they were amazing, fluffy, and delicious!

[+ Context](#)

A certain type of generalization of the mean value theorem to vector-valued functions is obtained as follows: Let f be a continuously differentiable real-valued function defined on an open interval I , and let x as well as $x + h$ be points of I .

[+ Context](#)

The Chevron Incentive Plan is an annual cash bonus plan for eligible employees that links awards to corporate, business unit and individual performance in the prior year.

[+ Context](#)

[Submit and Continue](#)

Figure 16: Screenshot of the developed annotation interface for rating English readability sentences. Annotators first rank sentences according to their readability level by simply dragging the box as shown in the figure. An optional Context button is available to show the context of a sentence if available.

Rank and Rate Sentences on Readability

Signed in as Anonymous [Sign out](#)

[View Instructions](#) [View Examples](#)

Batch ID: 500

Sentences

I also had to taste my Mom's multi-grain pumpkin pancakes with pecan butter and they were amazing, fluffy, and delicious!

[- Context](#)

There are only two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle.

[- Context](#)

The Chevron Incentive Plan is an annual cash bonus plan for eligible employees that links awards to corporate, business unit and individual performance in the prior year.

[+ Context](#)

The company also sponsors other postretirement benefit (OPEB) plans that provide medical and dental benefits, as well as life insurance for some active and qualifying retired employees.

[+ Context](#)

A certain type of generalization of the mean value theorem to vector-valued functions is obtained as follows: Let f be a continuously differentiable real-valued function defined on an open interval I , and let x as well as $x + h$ be points of I .

[+ Context](#)

[Submit and Continue](#)

Figure 17: After ranking, annotators then assign a score for each sentence on a scale of 1 to 6 that corresponds to the CEFR levels. When done, annotators submit their scores and proceed to another batch of 5 sentences.