

TFDS now supports the [Croissant](https://mlcommons.org/croissant)  format (<https://mlcommons.org/croissant>)! Read the [documentation](https://www.tensorflow.org/datasets/format_specific_dataset_builders#croissantbuilder) (https://www.tensorflow.org/datasets/format_specific_dataset_builders#croissantbuilder) to know more.

ted_hrlr_translate

- **Description:**

Data sets derived from TED talk transcripts for comparing similar language pairs where one is high resource and the other is low resource.

- **Homepage:** <https://github.com/neulab/word-embeddings-for-nmt>
(<https://github.com/neulab/word-embeddings-for-nmt>)
- **Source code:** [tfds.datasets.ted_hrlr_translate.Builder](https://github.com/tensorflow/datasets/tree/master/tensorflow_datasets/datasets/ted_hrlr_translate/ted_hrlr_translate_dataset_builder.py)
(https://github.com/tensorflow/datasets/tree/master/tensorflow_datasets/datasets/ted_hrlr_translate/ted_hrlr_translate_dataset_builder.py)
- **Versions:**
 - **1.0.0** (default): New split API (<https://tensorflow.org/datasets/splits>)
(<https://tensorflow.org/datasets/splits>)
- **Download size:** 124.94 MiB
- **Auto-cached** ([documentation](https://www.tensorflow.org/datasets/performances#auto-caching) (<https://www.tensorflow.org/datasets/performances#auto-caching>)): Yes
- **Figure** ([tfds.show_examples](https://www.tensorflow.org/datasets/api_docs/python/tfds/visualization/show_examples) (https://www.tensorflow.org/datasets/api_docs/python/tfds/visualization/show_examples)): Not supported.
- **Citation:**

```
@inproceedings{Ye2018WordEmbeddings,  
  author = {Ye, Qi and Devendra, Sachan and Matthieu, Felix and Sarguna, Padmanabhar  
  title = {When and Why are pre-trained word embeddings useful for Neural Machine T  
  booktitle = {HLT-NAACL},
```

```
year    = {2018},  
}
```

ted_hrlr_translate/az_to_en (default config)

- **Config description:** Translation dataset from az to en in plain text.
- **Dataset size:** 1.61 MiB
- **Splits:**

Split	Examples
'test'	903
'train'	5,946
'validation'	671

- **Feature structure:**

```
Translation({  
    'az': Text(shape=(), dtype=string),  
    'en': Text(shape=(), dtype=string),  
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
az	Text		string	
en	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('az' , 'en')

- **Examples** ([tfds.as_dataframe](https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/aztr_to_en

- **Config description:** Translation dataset from az_tr to en in plain text.
- **Dataset size:** 42.54 MiB
- **Splits:**

Split	Examples
'test'	903
'train'	188,396
'validation'	671

- **Feature structure:**

```
Translation({
  'az_tr': Text(shape=(), dtype=string),
  'en': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
az_tr	Text		string	
en	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('az_tr', 'en')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/be_to_en

- **Config description:** Translation dataset from be to en in plain text.
- **Dataset size:** 1.47 MiB
- **Splits:**

Split	Examples
'test'	664
'train'	4,509
'validation'	248

- **Feature structure:**

```
Translation({
  'be': Text(shape=(), dtype=string),
  'en': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			

Feature	Class	Shape	Dtype	Description
be	Text		string	
en	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('be' , 'en')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/beru_to_en

- **Config description:** Translation dataset from be_ru to en in plain text.
- **Dataset size:** 62.45 MiB
- **Splits:**

Split	Examples
'test'	664
'train'	212,614
'validation'	248

- **Feature structure:**

```
Translation({
  'be_ru': Text(shape=(), dtype=string),
  'en': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
be_ru	Text		string	
en	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('be_ru' , 'en')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/es_to_pt

- **Config description:** Translation dataset from es to pt in plain text.
- **Dataset size:** 9.62 MiB
- **Splits:**

Split	Examples
'test'	1,763
'train'	44,938
'validation'	1,016

- **Feature structure:**

```
Translation({
  'es': Text(shape=(), dtype=string),
```

```
'pt': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
Translation				
es	Text		string	
pt	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): (' es ' , ' pt ')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/fr_to_pt

- **Config description:** Translation dataset from fr to pt in plain text.
- **Dataset size:** 9.74 MiB
- **Splits:**

Split	Examples
'test'	1,494
'train'	43,873
'validation'	1,131

- **Feature structure:**

```
Translation({
  'fr': Text(shape=(), dtype=string),
  'pt': Text(shape=(), dtype=string),
})
```

• **Feature documentation:**

Feature	Class	Shape	Dtype	Description
Translation				
fr	Text		string	
pt	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('fr' , 'pt')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/gl_to_en

- **Config description:** Translation dataset from gl to en in plain text.
- **Dataset size:** 2.41 MiB
- **Splits:**

Split	Examples
'test'	1,007
'train'	10,017

Split	Examples
'validation'	682

- **Feature structure:**

```
Translation({
  'en': Text(shape=(), dtype=string),
  'gl': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
en	Text		string	
gl	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('gl', 'en')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/glpt_to_en

- **Config description:** Translation dataset from gl_pt to en in plain text.
- **Dataset size:** 12.90 MiB
- **Splits:**

Split	Examples
'test'	1,007
'train'	61,802
'validation'	682

- **Feature structure:**

```
Translation({
  'en': Text(shape=(), dtype=string),
  'gl_pt': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
en	Text		string	
gl_pt	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('gl_pt', 'en')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/he_to_pt

- **Config description:** Translation dataset from he to pt in plain text.

- **Dataset size:** 11.71 MiB
- **Splits:**

Split	Examples
'test'	1,623
'train'	48,511
'validation'	1,145

- **Feature structure:**

```
Translation({
  'he': Text(shape=(), dtype=string),
  'pt': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
he	Text		string	
pt	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('he' , 'pt')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/it_to_pt

- **Config description:** Translation dataset from it to pt in plain text.
- **Dataset size:** 9.94 MiB
- **Splits:**

Split	Examples
'test'	1,669
'train'	46,259
'validation'	1,162

- **Feature structure:**

```
Translation({
  'it': Text(shape=(), dtype=string),
  'pt': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
it	Text		string	
pt	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('it', 'pt')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/pt_to_en

- **Config description:** Translation dataset from pt to en in plain text.
- **Dataset size:** 10.89 MiB
- **Splits:**

Split	Examples
'test'	1,803
'train'	51,785
'validation'	1,193

- **Feature structure:**

```
Translation({
  'en': Text(shape=(), dtype=string),
  'pt': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
en	Text		string	
pt	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('pt', 'en')

- **Examples** ([tfds.as_dataframe](https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/ru_to_en

- **Config description:** Translation dataset from ru to en in plain text.
- **Dataset size:** 63.22 MiB
- **Splits:**

Split	Examples
'test'	5,476
'train'	208,106
'validation'	4,805

- **Feature structure:**

```
Translation({
  'en': Text(shape=(), dtype=string),
  'ru': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
en	Text		string	
ru	Text		string	

- **Supervised keys** (See `as_supervised` doc (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('ru' , 'en')
- **Examples** (`tfds.as_dataframe` (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/ru_to_pt

- **Config description:** Translation dataset from ru to pt in plain text.
- **Dataset size:** 13.00 MiB
- **Splits:**

Split	Examples
'test'	1,588
'train'	47,278
'validation'	1,184

- **Feature structure:**

```
Translation({
  'pt': Text(shape=(), dtype=string),
  'ru': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			

Feature	Class	Shape	Dtype	Description
pt	Text		string	
ru	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('ru' , 'pt')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

ted_hrlr_translate/tr_to_en

- **Config description:** Translation dataset from tr to en in plain text.
- **Dataset size:** 42.33 MiB
- **Splits:**

Split	Examples
'test'	5,029
'train'	182,450
'validation'	4,045

- **Feature structure:**

```
Translation({
  'en': Text(shape=(), dtype=string),
  'tr': Text(shape=(), dtype=string),
})
```

- **Feature documentation:**

Feature	Class	Shape	Dtype	Description
	Translation			
en	Text		string	
tr	Text		string	

- **Supervised keys** (See [as_supervised doc](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/load#args)): ('tr' , 'en')
- **Examples** ([tfds.as_dataframe](#) (https://www.tensorflow.org/datasets/api_docs/python/tfds/as_dataframe)):

[Display examples...](#)

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](https://developers.google.com/site-policies) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2023-02-11 UTC.