

[ ![Logo](../../\_static/logo.png) ](../../index.html)

## Getting Started

- \* [Installation](../../docs/installation.html)

- \* [Install with pip](../../docs/installation.html#install-with-pip)

- \* [Install with Conda](../../docs/installation.html#install-with-conda)

- \* [Install from Source](../../docs/installation.html#install-from-source)

- \* [Editable Install](../../docs/installation.html#editable-install)

- \* [Install PyTorch with CUDA support](../../docs/installation.html#install-pytorch-with-cuda-support)

- \* [Quickstart](../../docs/quickstart.html)

- \* [Sentence Transformer](../../docs/quickstart.html#sentence-transformer)

- \* [Cross Encoder](../../docs/quickstart.html#cross-encoder)

- \* [Next Steps](../../docs/quickstart.html#next-steps)

## Sentence Transformer

- \* [Usage](../../docs/sentence\_transformer/usage/usage.html)

- \* [Computing Embeddings](../../applications/computing-embeddings/README.html)

- \* [Initializing a Sentence Transformer Model](../../applications/computing-embeddings/README.html#initializing-a-sentence-transformer-model)

- \* [Calculating Embeddings](../../applications/computing-embeddings/README.html#calculating-embeddings)

- \* [Prompt Templates](../../applications/computing-embeddings/README.html#prompt-templates)

- \* [Input Sequence Length](../../applications/computing-embeddings/README.html#id1)

\* [Multi-Process / Multi-GPU

Encoding](../../applications/computing-embeddings/README.html#multi-process-multi-gpu-encoding)

\* [Semantic Textual

Similarity](../../docs/sentence\_transformer/usage/semantic\_textual\_similarity.html)

\* [Similarity

Calculation](../../docs/sentence\_transformer/usage/semantic\_textual\_similarity.html#similarity-calculation)

\* [Semantic Search](../../applications/semantic-search/README.html)

\* [Background](../../applications/semantic-search/README.html#background)

\* [Symmetric vs. Asymmetric Semantic

Search](../../applications/semantic-search/README.html#symmetric-vs-asymmetric-semantic-search)

\* [Manual

Implementation](../../applications/semantic-search/README.html#manual-implementation)

\* [Optimized

Implementation](../../applications/semantic-search/README.html#optimized-implementation)

\* [Speed Optimization](../../applications/semantic-search/README.html#speed-optimization)

\* [Elasticsearch](../../applications/semantic-search/README.html#elasticsearch)

\* [Approximate Nearest

Neighbor](../../applications/semantic-search/README.html#approximate-nearest-neighbor)

\* [Retrieve & Re-Rank](../../applications/semantic-search/README.html#retrieve-re-rank)

\* [Examples](../../applications/semantic-search/README.html#examples)

\* [Retrieve & Re-Rank](../../applications/retrieve\_rerank/README.html)

\* [Retrieve & Re-Rank

Pipeline](../../applications/retrieve\_rerank/README.html#retrieve-re-rank-pipeline)

\* [Retrieval: Bi-Encoder](../../applications/retrieve\_rerank/README.html#retrieval-bi-encoder)

\* [Re-Ranker:

Cross-Encoder](../../applications/retrieve\_rerank/README.html#re-ranker-cross-encoder)

\* [Example Scripts](../../applications/retrieve\_rerank/README.html#example-scripts)

\* [Pre-trained Bi-Encoders

(Retrieval)](../../applications/retrieve\_rerank/README.html#pre-trained-bi-encoders-retrieval)

\* [Pre-trained Cross-Encoders

(Re-Ranker)](../../applications/retrieve\_rerank/README.html#pre-trained-cross-encoders-re-ranker)

\* [Clustering](../../applications/clustering/README.html)

\* [k-Means](../../applications/clustering/README.html#k-means)

\* [Agglomerative Clustering](../../applications/clustering/README.html#agglomerative-clustering)

\* [Fast Clustering](../../applications/clustering/README.html#fast-clustering)

\* [Topic Modeling](../../applications/clustering/README.html#topic-modeling)

\* [Paraphrase Mining](../../applications/paraphrase-mining/README.html)

\*

[`paraphrase\_mining()'](../../applications/paraphrase-mining/README.html#sentence\_transformers.  
util.paraphrase\_mining)

\* [Translated Sentence Mining](../../applications/parallel-sentence-mining/README.html)

\* [Margin Based

Mining](../../applications/parallel-sentence-mining/README.html#margin-based-mining)

\* [Examples](../../applications/parallel-sentence-mining/README.html#examples)

\* [Image Search](../../applications/image-search/README.html)

\* [Installation](../../applications/image-search/README.html#installation)

\* [Usage](../../applications/image-search/README.html#usage)

\* [Examples](../../applications/image-search/README.html#examples)

\* [Embedding Quantization](../../applications/embedding-quantization/README.html)

\* [Binary

Quantization](../../applications/embedding-quantization/README.html#binary-quantization)

Quantization](../applications/embedding-quantization/README.html#scalar-int8-quantization)

- \* [Additional extensions](../applications/embedding-quantization/README.html#additional-extensions)
- \* [Demo](../applications/embedding-quantization/README.html#demo)
- \* [Try it yourself](../applications/embedding-quantization/README.html#try-it-yourself)
- \* [Speeding up Inference](../docs/sentence\_transformer/usage/efficiency.html)
- \* [PyTorch](../docs/sentence\_transformer/usage/efficiency.html#pytorch)
- \* [ONNX](../docs/sentence\_transformer/usage/efficiency.html#onnx)
- \* [OpenVINO](../docs/sentence\_transformer/usage/efficiency.html#openvino)
- \* [Benchmarks](../docs/sentence\_transformer/usage/efficiency.html#benchmarks)
- \* [Creating Custom Models](../docs/sentence\_transformer/usage/custom\_models.html)
- \* [Structure of Sentence Transformer Models](../docs/sentence\_transformer/usage/custom\_models.html#structure-of-sentence-transformer-models)
- \* [Sentence Transformer Model from a Transformers Model](../docs/sentence\_transformer/usage/custom\_models.html#sentence-transformer-model-from-a-transformers-model)
- \* [Pretrained Models](../docs/sentence\_transformer/pretrained\_models.html)
- \* [Original Models](../docs/sentence\_transformer/pretrained\_models.html#original-models)
- \* [Semantic Search Models](../docs/sentence\_transformer/pretrained\_models.html#semantic-search-models)
- \* [Multi-QA Models](../docs/sentence\_transformer/pretrained\_models.html#multi-qa-models)
- \* [MSMARCO Passage Models](../docs/sentence\_transformer/pretrained\_models.html#msmarco-passage-models)
- \* [Multilingual Models](../docs/sentence\_transformer/pretrained\_models.html#multilingual-models)

[\\* \[Semantic Similarity Models\]\(../../docs/sentence\\_transformer/pretrained\\_models.html#semantic-similarity-models\)](#)  
[\\* \[Bitext Mining\]\(../../docs/sentence\\_transformer/pretrained\\_models.html#bitext-mining\)](#)  
[\\* \[Image & Text-Models\]\(../../docs/sentence\\_transformer/pretrained\\_models.html#image-text-models\)](#)  
[\\* \[INSTRUCTOR models\]\(../../docs/sentence\\_transformer/pretrained\\_models.html#instructor-models\)](#)  
[\\* \[Scientific Similarity Models\]\(../../docs/sentence\\_transformer/pretrained\\_models.html#scientific-similarity-models\)](#)  
[\\* \[Training Overview\]\(../../docs/sentence\\_transformer/training\\_overview.html\)](#)  
[\\* \[Why Finetune?\]\(../../docs/sentence\\_transformer/training\\_overview.html#why-finetune\)](#)  
[\\* \[Training Components\]\(../../docs/sentence\\_transformer/training\\_overview.html#training-components\)](#)  
[\\* \[Dataset\]\(../../docs/sentence\\_transformer/training\\_overview.html#dataset\)](#)  
[\\* \[Dataset Format\]\(../../docs/sentence\\_transformer/training\\_overview.html#dataset-format\)](#)  
[\\* \[Loss Function\]\(../../docs/sentence\\_transformer/training\\_overview.html#loss-function\)](#)  
[\\* \[Training Arguments\]\(../../docs/sentence\\_transformer/training\\_overview.html#training-arguments\)](#)  
[\\* \[Evaluator\]\(../../docs/sentence\\_transformer/training\\_overview.html#evaluator\)](#)  
[\\* \[Trainer\]\(../../docs/sentence\\_transformer/training\\_overview.html#trainer\)](#)  
[\\* \[Callbacks\]\(../../docs/sentence\\_transformer/training\\_overview.html#callbacks\)](#)  
[\\* \[Multi-Dataset Training\]\(../../docs/sentence\\_transformer/training\\_overview.html#multi-dataset-training\)](#)  
[\\* \[Deprecated Training\]\(../../docs/sentence\\_transformer/training\\_overview.html#deprecated-training\)](#)  
[\\* \[Best Base Embedding Models\]\(../../docs/sentence\\_transformer/training\\_overview.html#best-base-embedding-models\)](#)

- \* [Dataset Overview](../../docs/sentence\_transformer/dataset\_overview.html)
- \* [Datasets on the Hugging Face Hub](../../docs/sentence\_transformer/dataset\_overview.html#datasets-on-the-hugging-face-hub)
- \* [Pre-existing Datasets](../../docs/sentence\_transformer/dataset\_overview.html#pre-existing-datasets)
- \* [Loss Overview](../../docs/sentence\_transformer/loss\_overview.html)
- \* [Loss modifiers](../../docs/sentence\_transformer/loss\_overview.html#loss-modifiers)
- \* [Distillation](../../docs/sentence\_transformer/loss\_overview.html#distillation)
- \* [Commonly used Loss Functions](../../docs/sentence\_transformer/loss\_overview.html#commonly-used-loss-functions)
- \* [Custom Loss Functions](../../docs/sentence\_transformer/loss\_overview.html#custom-loss-functions)
- \* [Training Examples](../../docs/sentence\_transformer/training/examples.html)
- \* [Semantic Textual Similarity](../sts/README.html)
- \* [Training data](../sts/README.html#training-data)
- \* [Loss Function](../sts/README.html#loss-function)
- \* [Natural Language Inference](../nli/README.html)
- \* [Data](../nli/README.html#data)
- \* [SoftmaxLoss](../nli/README.html#softmaxloss)
- \* [MultipleNegativesRankingLoss](../nli/README.html#multiplenegativesrankingloss)
- \* [Paraphrase Data](../paraphrases/README.html)
- \* [Pre-Trained Models](../paraphrases/README.html#pre-trained-models)
- \* [Quora Duplicate Questions](../quora\_duplicate\_questions/README.html)
- \* [Training](../quora\_duplicate\_questions/README.html#training)
- \* [MultipleNegativesRankingLoss](../quora\_duplicate\_questions/README.html#multiplenegativesrankingloss)

- \* [Pretrained Models](../quora\_duplicate\_questions/README.html#pretrained-models)
- \* [MS MARCO](../ms\_marco/README.html)
- \* [Bi-Encoder](../ms\_marco/README.html#bi-encoder)
- \* [Matryoshka Embeddings](../matryoshka/README.html)
- \* [Use Cases](../matryoshka/README.html#use-cases)
- \* [Results](../matryoshka/README.html#results)
- \* [Training](../matryoshka/README.html#training)
- \* [Inference](../matryoshka/README.html#inference)
- \* [Code Examples](../matryoshka/README.html#code-examples)
- \* [Adaptive Layers](../adaptive\_layer/README.html)
- \* [Use Cases](../adaptive\_layer/README.html#use-cases)
- \* [Results](../adaptive\_layer/README.html#results)
- \* [Training](../adaptive\_layer/README.html#training)
- \* [Inference](../adaptive\_layer/README.html#inference)
- \* [Code Examples](../adaptive\_layer/README.html#code-examples)
- \* Multilingual Models
  - \* Extend your own models
  - \* Training
  - \* Datasets
  - \* Sources for Training Data
  - \* Evaluation
  - \* Available Pre-trained Models
  - \* Usage
  - \* Performance
  - \* Citation
- \* [Model Distillation](../distillation/README.html)
- \* [Knowledge Distillation](../distillation/README.html#knowledge-distillation)

- \* [\[Speed - Performance Trade-Off\]\(../distillation/README.html#speed-performance-trade-off\)](#)
- \* [\[Dimensionality Reduction\]\(../distillation/README.html#dimensionality-reduction\)](#)
- \* [\[Quantization\]\(../distillation/README.html#quantization\)](#)
- \* [\[Augmented SBERT\]\(../data\\_augmentation/README.html\)](#)
- \* [\[Motivation\]\(../data\\_augmentation/README.html#motivation\)](#)
  - \* [\[Extend to your own datasets\]\(../data\\_augmentation/README.html#extend-to-your-own-datasets\)](#)
  - \* [\[Methodology\]\(../data\\_augmentation/README.html#methodology\)](#)
    - \* [\[Scenario 1: Limited or small annotated datasets \(few labeled sentence-pairs\)\]\(../data\\_augmentation/README.html#scenario-1-limited-or-small-annotated-dataset-s-few-labeled-sentence-pairs\)](#)
    - \* [\[Scenario 2: No annotated datasets \(Only unlabeled sentence-pairs\)\]\(../data\\_augmentation/README.html#scenario-2-no-annotated-datasets-only-unlabeled-sentence-pairs\)](#)
  - \* [\[Training\]\(../data\\_augmentation/README.html#training\)](#)
  - \* [\[Citation\]\(../data\\_augmentation/README.html#citation\)](#)
- \* [\[Training with Prompts\]\(../prompts/README.html\)](#)
  - \* [\[What are Prompts?\]\(../prompts/README.html#what-are-prompts\)](#)
    - \* [\[Why would we train with Prompts?\]\(../prompts/README.html#why-would-we-train-with-prompts\)](#)
    - \* [\[How do we train with Prompts?\]\(../prompts/README.html#how-do-we-train-with-prompts\)](#)
  - \* [\[Training with PEFT Adapters\]\(../peft/README.html\)](#)
    - \* [\[Compatibility Methods\]\(../peft/README.html#compatibility-methods\)](#)
    - \* [\[Adding a New Adapter\]\(../peft/README.html#adding-a-new-adapter\)](#)
    - \* [\[Loading a Pretrained Adapter\]\(../peft/README.html#loading-a-pretrained-adapter\)](#)
    - \* [\[Training Script\]\(../peft/README.html#training-script\)](#)
  - \* [\[Unsupervised Learning\]\(../unsupervised\\_learning/README.html\)](#)



\* [TSDAE](../unsupervised\_learning/README.html#tsdae)

\* [SimCSE](../unsupervised\_learning/README.html#simcse)

\* [CT](../unsupervised\_learning/README.html#ct)

\* [CT (In-Batch Negative Sampling)](../unsupervised\_learning/README.html#ct-in-batch-negative-sampling)

\* [Masked Language Model (MLM)](../unsupervised\_learning/README.html#masked-language-model-mlm)

\* [GenQ](../unsupervised\_learning/README.html#genq)

\* [GPL](../unsupervised\_learning/README.html#gpl)

\* [Performance Comparison](../unsupervised\_learning/README.html#performance-comparison)

\* [Domain Adaptation](../domain\_adaptation/README.html)

\* [Domain Adaptation vs. Unsupervised Learning](../domain\_adaptation/README.html#domain-adaptation-vs-unsupervised-learning)

\* [Adaptive Pre-Training](../domain\_adaptation/README.html#adaptive-pre-training)

\* [GPL: Generative Pseudo-Labeling](../domain\_adaptation/README.html#gpl-generative-pseudo-labeling)

\* [Hyperparameter Optimization](../hpo/README.html)

\* [HPO Components](../hpo/README.html#hpo-components)

\* [Putting It All Together](../hpo/README.html#putting-it-all-together)

\* [Example Scripts](../hpo/README.html#example-scripts)

\* [Distributed Training](../docs/sentence\_transformer/training/distributed.html)

\* [Comparison](../docs/sentence\_transformer/training/distributed.html#comparison)

\* [FSDP](../docs/sentence\_transformer/training/distributed.html#fsdp)

Cross Encoder

\* [Usage](../../docs/cross\_encoder/usage/usage.html)

\* [Retrieve & Re-Rank](../../applications/retrieve\_rerank/README.html)

\* [Retrieve & Re-Rank Pipeline](../../applications/retrieve\_rerank/README.html#retrieve-re-rank-pipeline)

\* [Retrieval: Bi-Encoder](../../applications/retrieve\_rerank/README.html#retrieval-bi-encoder)

\* [Re-Ranker: Cross-Encoder](../../applications/retrieve\_rerank/README.html#re-ranker-cross-encoder)

\* [Example Scripts](../../applications/retrieve\_rerank/README.html#example-scripts)

\* [Pre-trained Bi-Encoders (Retrieval)](../../applications/retrieve\_rerank/README.html#pre-trained-bi-encoders-retrieval)

\* [Pre-trained Cross-Encoders (Re-Ranker)](../../applications/retrieve\_rerank/README.html#pre-trained-cross-encoders-re-ranker)

\* [Pretrained Models](../../docs/cross\_encoder/pretrained\_models.html)

\* [MS MARCO](../../docs/cross\_encoder/pretrained\_models.html#ms-marco)

\* [SQuAD (QNLI)](../../docs/cross\_encoder/pretrained\_models.html#squad-qnli)

\* [STSbenchmark](../../docs/cross\_encoder/pretrained\_models.html#stsbenchmark)

\* [Quora Duplicate Questions](../../docs/cross\_encoder/pretrained\_models.html#quora-duplicate-questions)

\* [NLI](../../docs/cross\_encoder/pretrained\_models.html#nli)

\* [Community Models](../../docs/cross\_encoder/pretrained\_models.html#community-models)

\* [Training Overview](../../docs/cross\_encoder/training\_overview.html)

\* [Training Examples](../../docs/cross\_encoder/training/examples.html)

\* [MS MARCO](../ms\_marco/cross\_encoder\_README.html)

\* [Cross-Encoder](../ms\_marco/cross\_encoder\_README.html#cross-encoder)

\* [Cross-Encoder Knowledge Distillation](../ms\_marco/cross\_encoder\_README.html#cross-encoder-knowledge-distillation)

## Package Reference

\* [Sentence Transformer](../../docs/package\_reference/sentence\_transformer/index.html)

\*

[SentenceTransformer](../../docs/package\_reference/sentence\_transformer/SentenceTransformer.html)

\*

[SentenceTransformer](../../docs/package\_reference/sentence\_transformer/SentenceTransformer.html#id1)

\*

[SentenceTransformerModelCardData](../../docs/package\_reference/sentence\_transformer/SentenceTransformer.html#sentencetransformermodelcarddata)

\*

[SimilarityFunction](../../docs/package\_reference/sentence\_transformer/SentenceTransformer.html#similarityfunction)

\* [Trainer](../../docs/package\_reference/sentence\_transformer/trainer.html)

\*

[SentenceTransformerTrainer](../../docs/package\_reference/sentence\_transformer/trainer.html#sentencetransformertrainer)

\* [Training Arguments](../../docs/package\_reference/sentence\_transformer/training\_args.html)

\*

[SentenceTransformerTrainingArguments](../../docs/package\_reference/sentence\_transformer/training\_args.html#sentencetransformertrainingarguments)

\* [Losses](../../docs/package\_reference/sentence\_transformer/losses.html)

\*

[BatchAllTripletLoss](../../docs/package\_reference/sentence\_transformer/losses.html#batchalltripletloss)

\*

[BatchHardSoftMarginTripletLoss](../../docs/package\_reference/sentence\_transformer/losses.html#batchhardsoftmargintripletloss)

\*

[BatchHardTripletLoss](../../docs/package\_reference/sentence\_transformer/losses.html#batchhardtripletloss)

\*

[BatchSemiHardTripletLoss](../../docs/package\_reference/sentence\_transformer/losses.html#batchsemi-hardtripletloss)

\*

[ContrastiveLoss](../../docs/package\_reference/sentence\_transformer/losses.html#contrastiveloss)

\*

[OnlineContrastiveLoss](../../docs/package\_reference/sentence\_transformer/losses.html#onlinecontrastiveloss)

\*

[ContrastiveTensionLoss](../../docs/package\_reference/sentence\_transformer/losses.html#contrastivetensionloss)

\*

[ContrastiveTensionLossInBatchNegatives](../../docs/package\_reference/sentence\_transformer/losses.html#contrastivetensionlossinbatchnegatives)

\* [CoSENTLoss](../../docs/package\_reference/sentence\_transformer/losses.html#cosentloss)

\* [AngleLoss](../../docs/package\_reference/sentence\_transformer/losses.html#angleloss)

\*

[CosineSimilarityLoss](../../docs/package\_reference/sentence\_transformer/losses.html#cosinesimilarityloss)

\*

[DenoisingAutoEncoderLoss](../../docs/package\_reference/sentence\_transformer/losses.html#denoisingautoencoderloss)

osingautoencoderloss)

\*

[GISTEmbedLoss](../../docs/package\_reference/sentence\_transformer/losses.html#gistembedloss  
)

\*

[CachedGISTEmbedLoss](../../docs/package\_reference/sentence\_transformer/losses.html#cachedgistembedloss)

\* [MSELoss](../../docs/package\_reference/sentence\_transformer/losses.html#mseloss)

\*

[MarginMSELoss](../../docs/package\_reference/sentence\_transformer/losses.html#marginmseloss  
)

\*

[MatryoshkaLoss](../../docs/package\_reference/sentence\_transformer/losses.html#matryoshkaloss  
)

\*

[Matryoshka2dLoss](../../docs/package\_reference/sentence\_transformer/losses.html#matryoshka2dloss)

\*

[AdaptiveLayerLoss](../../docs/package\_reference/sentence\_transformer/losses.html#adaptivelayerloss)

\*

[MegaBatchMarginLoss](../../docs/package\_reference/sentence\_transformer/losses.html#megabatchmarginloss)

\*

[MultipleNegativesRankingLoss](../../docs/package\_reference/sentence\_transformer/losses.html#multiplenegativesrankingloss)

\*

[CachedMultipleNegativesRankingLoss](../../docs/package\_reference/sentence\_transformer/losses.html#cachedmultiplenegativesrankingloss)

\*

[MultipleNegativesSymmetricRankingLoss](../../docs/package\_reference/sentence\_transformer/losses.html#multiplenegativessymmetricrankingloss)

\*

[CachedMultipleNegativesSymmetricRankingLoss](../../docs/package\_reference/sentence\_transformer/losses.html#cachedmultiplenegativessymmetricrankingloss)

\* [SoftmaxLoss](../../docs/package\_reference/sentence\_transformer/losses.html#softmaxloss)

\* [TripletLoss](../../docs/package\_reference/sentence\_transformer/losses.html#tripletloss)

\* [Samplers](../../docs/package\_reference/sentence\_transformer/sampler.html)

\*

[BatchSamplers](../../docs/package\_reference/sentence\_transformer/sampler.html#batchsamplers)

\*

[MultiDatasetBatchSamplers](../../docs/package\_reference/sentence\_transformer/sampler.html#multidatasetbatchsamplers)

\* [Evaluation](../../docs/package\_reference/sentence\_transformer/evaluation.html)

\*

[BinaryClassificationEvaluator](../../docs/package\_reference/sentence\_transformer/evaluation.html#binaryclassificationevaluator)

\*

[EmbeddingSimilarityEvaluator](../../docs/package\_reference/sentence\_transformer/evaluation.html#embeddingsimilarityevaluator)

\*

[InformationRetrievalEvaluator](../../docs/package\_reference/sentence\_transformer/evaluation.html#informationretrievalevaluator)

\*

[NanoBEIREvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#nanobe  
irevaluator)

\*

[MSEEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#mseevaluator  
)

\*

[ParaphraseMiningEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#  
paraphraseminingevaluator)

\*

[RerankingEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#rerankin  
gevaluator)

\*

[SentenceEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#sentenc  
eevaluator)

\*

[SequentialEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#sequen  
tiaevaluator)

\*

[TranslationEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#translat  
ionevaluator)

\*

[TripletEvaluator](../../../../docs/package\_reference/sentence\_transformer/evaluation.html#tripletevalua  
tor)

\* [Datasets](../../../../docs/package\_reference/sentence\_transformer/datasets.html)

\*

[ParallelSentencesDataset](../../../../docs/package\_reference/sentence\_transformer/datasets.html#par

allelsentencesdataset)

\*

[SentenceLabelDataset](../../docs/package\_reference/sentence\_transformer/datasets.html#sentence-label-dataset)

\*

[DenoisingAutoEncoderDataset](../../docs/package\_reference/sentence\_transformer/datasets.html#denoising-auto-encoder-dataset)

\*

[NoDuplicatesDataLoader](../../docs/package\_reference/sentence\_transformer/datasets.html#no-duplicates-data-loader)

\* [Models](../../docs/package\_reference/sentence\_transformer/models.html)

\*

[Main

Classes](../../docs/package\_reference/sentence\_transformer/models.html#main-classes)

\*

[Further

Classes](../../docs/package\_reference/sentence\_transformer/models.html#further-classes)

\* [quantization](../../docs/package\_reference/sentence\_transformer/quantization.html)

\*

[`quantize\_embeddings()`](../../docs/package\_reference/sentence\_transformer/quantization.html#sentence-transformers.quantization.quantize\_embeddings)

\*

[`semantic\_search\_faiss()`](../../docs/package\_reference/sentence\_transformer/quantization.html#sentence-transformers.quantization.semantic\_search\_faiss)

\*

[`semantic\_search\_usearch()`](../../docs/package\_reference/sentence\_transformer/quantization.html#sentence-transformers.quantization.semantic\_search\_usearch)

\* [Cross Encoder](../../docs/package\_reference/cross\_encoder/index.html)

\* [CrossEncoder](../../docs/package\_reference/cross\_encoder/cross\_encoder.html)



\* [CrossEncoder](../../docs/package\_reference/cross\_encoder/cross\_encoder.html#id1)

\*

[Training

Inputs](../../docs/package\_reference/cross\_encoder/cross\_encoder.html#training-inputs)

\* [Evaluation](../../docs/package\_reference/cross\_encoder/evaluation.html)

\*

[CEBinaryAccuracyEvaluator](../../docs/package\_reference/cross\_encoder/evaluation.html#cebinaryaccuracyevaluator)

\*

[CEBinaryClassificationEvaluator](../../docs/package\_reference/cross\_encoder/evaluation.html#cebinaryclassificationevaluator)

\*

[CECorrelationEvaluator](../../docs/package\_reference/cross\_encoder/evaluation.html#cecorrelationevaluator)

\* [CEF1Evaluator](../../docs/package\_reference/cross\_encoder/evaluation.html#cef1evaluator)

\*

[CESoftmaxAccuracyEvaluator](../../docs/package\_reference/cross\_encoder/evaluation.html#cesoftmaxaccuracyevaluator)

\*

[CERerankingEvaluator](../../docs/package\_reference/cross\_encoder/evaluation.html#cererankingevaluator)

\* [util](../../docs/package\_reference/util.html)

\* [Helper Functions](../../docs/package\_reference/util.html#module-sentence\_transformers.util)

\*

[`community\_detection()`](../../docs/package\_reference/util.html#sentence\_transformers.util.community\_detection)

\* [`http\_get()`](../../docs/package\_reference/util.html#sentence\_transformers.util.http\_get)

\*

[`is\_training\_available()`](../../docs/package\_reference/util.html#sentence\_transformers.util.is\_training\_available)

\*

[`mine\_hard\_negatives()`](../../docs/package\_reference/util.html#sentence\_transformers.util.mine\_hard\_negatives)

\*

[`normalize\_embeddings()`](../../docs/package\_reference/util.html#sentence\_transformers.util.normalize\_embeddings)

\*

[`paraphrase\_mining()`](../../docs/package\_reference/util.html#sentence\_transformers.util.paraphrase\_mining)

\*

[`semantic\_search()`](../../docs/package\_reference/util.html#sentence\_transformers.util.semantic\_search)

\*

[`truncate\_embeddings()`](../../docs/package\_reference/util.html#sentence\_transformers.util.truncate\_embeddings)

\*

[Model

Optimization](../../docs/package\_reference/util.html#module-sentence\_transformers.backend)

\*

[`export\_dynamic\_quantized\_onnx\_model()`](../../docs/package\_reference/util.html#sentence\_transformers.backend.export\_dynamic\_quantized\_onnx\_model)

\*

[`export\_optimized\_onnx\_model()`](../../docs/package\_reference/util.html#sentence\_transformers.backend.export\_optimized\_onnx\_model)

\*

[`export\_static\_quantized\_openvino\_model()`](../../docs/package\_reference/util.html#sentence\_tra

nsformers.backend.export\_static\_quantized\_openvino\_model)

\* [Similarity Metrics](../../docs/package\_reference/util.html#module-sentence\_transformers.util)

\* [`cos_sim()`](../../docs/package\_reference/util.html#sentence\_transformers.util.cos\_sim)

\* [`dot_score()`](../../docs/package\_reference/util.html#sentence\_transformers.util.dot\_score)

\*

[`euclidean_sim()`](../../docs/package\_reference/util.html#sentence\_transformers.util.euclidean\_sim)

\*

[`manhattan_sim()`](../../docs/package\_reference/util.html#sentence\_transformers.util.manhattan\_sim)

\*

[`pairwise_cos_sim()`](../../docs/package\_reference/util.html#sentence\_transformers.util.pairwise\_cos\_sim)

\*

[`pairwise_dot_score()`](../../docs/package\_reference/util.html#sentence\_transformers.util.pairwise\_dot\_score)

\*

[`pairwise_euclidean_sim()`](../../docs/package\_reference/util.html#sentence\_transformers.util.pairwise\_euclidean\_sim)

\*

[`pairwise_manhattan_sim()`](../../docs/package\_reference/util.html#sentence\_transformers.util.pairwise\_manhattan\_sim)

\_\_[Sentence Transformers](../../index.html)

\* [(../../index.html)](../../index.html)

\* [Training Examples](../../docs/sentence\_transformer/training/examples.html)

\* Multilingual Models

\*

[

Edit

on

GitHub](<https://github.com/UKPLab/sentence-transformers/blob/master/examples/training/multilingual/README.md>)

\* \* \*

# Multilingual Models

The issue with multilingual BERT (mBERT) as well as with XLM-RoBERTa is that those produce rather bad sentence representation out-of-the-box. Further, the vectors spaces between languages are not aligned, i.e., the sentences with the same content in different languages would be mapped to different locations in the vector space.

In my publication [Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation](<https://arxiv.org/abs/2004.09813>) I describe an easy approach to extend sentence embeddings to further languages.

Chien Vu also wrote a nice blog article on this technique: [A complete guide to transfer learning from English to other Languages using Sentence Embeddings BERT Models](<https://towardsdatascience.com/a-complete-guide-to-transfer-learning-from-english-to-other-languages-using-sentence-embeddings-8c427f8804a9>)

## Extend your own models

![[Multilingual Knowledge

Distillation]](<https://raw.githubusercontent.com/UKPLab/sentence-transformers/master/docs/img/multilingual-distillation.png>)

The idea is based on a fixed (monolingual) **teacher model** that produces sentence embeddings with our desired properties in one language (e.g. English). The **student model** is supposed to mimic the teacher model, i.e., the same English sentence should be mapped to the same vector by the teacher and by the student model. Additionally, in order to make the student model work for other languages, we train the student model on parallel (translated) sentences. The translation of each sentence should also be mapped to the same vector as the original sentence.

In the above figure, the student model should map `_Hello World_` and the German translation `_Hallo Welt_` to the vector of ``teacher_model('Hello World')``. We achieve this by training the student model using mean squared error (MSE) loss.

In our experiments we initialized the student model with the multilingual [XLM-RoBERTa model](<https://huggingface.co/FacebookAI/xlm-roberta-base>).

## Training

For a **fully automatic code example**, see

[`make_multilingual.py`]([https://github.com/UKPLab/sentence-transformers/tree/master/examples/training/multilingual/make\\_multilingual.py](https://github.com/UKPLab/sentence-transformers/tree/master/examples/training/multilingual/make_multilingual.py)).

This script downloads the parallel sentences corpus, a corpus with transcripts and translations from talks. It then extends a monolingual model to several languages (en, de, es, it, fr, ar, tr). This corpus contains parallel data for more than 100 languages, hence, you can simply change the script and train a multilingual model in your favorite languages.

## ## Datasets

As training data we require parallel sentences, i.e., sentences translated in various languages. In particular, we will use

[Dataset](https://huggingface.co/docs/datasets/main/en/package\_reference/main\_classes#datasets.Dataset)

"(in datasets main)" instances with "english" and "non\_english" columns. We have prepared a large collection of such datasets in our [Parallel Sentences dataset collection](https://huggingface.co/collections/sentence-transformers/parallel-sentences-datasets-6644d644123d31ba5b1c8785).

The training script will take the "english" column and add a "label" column containing the embeddings of the english texts. Then, the student model "english" and "non\_english" will be trained to be similar to this "label". You can load such a training dataset like so:

```
from datasets import load_dataset
```

```
train_dataset = load_dataset("sentence-transformers/parallel-sentences-talks", "en-de",
```

```
split="train")
```

```
print(train_dataset[0])
```

```
# {"english": "So I think practicality is one case where it's worth teaching people by hand.",  
"non_english": "Ich denke, dass es sich aus diesem Grund lohnt, den Leuten das Rechnen von  
Hand beizubringen."}
```

## ## Sources for Training Data

A great website for a vast number of parallel (translated) datasets is [OPUS](<http://opus.nlpl.eu/>). There, you find parallel datasets for more than 400 languages. You can use these to create your own parallel sentence datasets, if you wish.

## ## Evaluation

Training can be evaluated in different ways. For an example how to use these evaluation methods, see

[make\_multilingual.py]([https://github.com/UKPLab/sentence-transformers/tree/master/examples/training/multilingual/make\\_multilingual.py](https://github.com/UKPLab/sentence-transformers/tree/master/examples/training/multilingual/make_multilingual.py)).

## ### MSE Evaluation

You can measure the mean squared error (MSE) between the student embeddings and teacher embeddings.

```
from datasets import load_dataset
```

```
eval_dataset = load_dataset("sentence-transformers/parallel-sentences-talks", "en-fr", split="dev")
```

```
dev_mse = MSEEvaluator(  
    source_sentences=eval_dataset["english"],  
    target_sentences=eval_dataset["non_english"],  
    name="en-fr-dev",  
    teacher_model=teacher_model,  
    batch_size=32,  
)
```

This evaluator computes the teacher embeddings for the `source_sentences`, for example, for English. During training, the student model is used to compute embeddings for the `target_sentences`, for example, for French. The distance between teacher and student embeddings is measured. Lower scores indicate a better performance.

### ### Translation Accuracy

You can also measure the translation accuracy. As inputs, this evaluator accepts a list of `source_sentences` (e.g. English), and a list of `target_sentences` (e.g. Spanish), such that `target_sentences[i]` is a translation of `source_sentences[i]`.



For each sentence pair, we check if `source\_sentences[i]` we check if  
`target\_sentences[i]` has the highest similarity out of all target sentences.  
If this is the case, we have a hit, otherwise an error. This evaluator reports  
accuracy (higher = better).

```
from datasets import load_dataset
```

```
eval_dataset = load_dataset("sentence-transformers/parallel-sentences-talks", "en-fr", split="dev")
```

```
dev_trans_acc = TranslationEvaluator(  
    source_sentences=eval_dataset["english"],  
    target_sentences=eval_dataset["non_english"],  
    name="en-fr-dev",  
    batch_size=32,  
)
```

### ### Multilingual Semantic Textual Similarity *if*•

You can also measure the semantic textual similarity (STS) between sentence  
pairs in different languages:

```
from datasets import load_dataset
```

```

test_dataset = load_dataset("mteb/sts17-crosslingual-sts", "nl-en", split="test")

test_emb_similarity = EmbeddingSimilarityEvaluator(
    sentences1=test_dataset["sentence1"],
    sentences2=test_dataset["sentence2"],
    scores=[score / 5.0 for score in test_dataset["score"]], # Convert 0-5 scores to 0-1 scores
    batch_size=32,
    name=f"sts17-nl-en-test",
    show_progress_bar=False,
)

```

Where `sentences1` and `sentences2` are lists of sentences and score is numeric value indicating the semantic similarity between `sentences1[i]` and `sentences2[i]`.

## ## Available Pre-trained Models¶

For a list of available models, see [Pretrained Models](../../docs/sentence\_transformer/pretrained\_models.html#multilingual-models).

## ## Usage¶

You can use the models in the following way:

```

from sentence_transformers import SentenceTransformer

model = SentenceTransformer("paraphrase-multilingual-MiniLM-L12-v2")

embeddings = model.encode(["Hello World", "Hallo Welt", "Hola mundo", "Bye, Moon!"])

similarities = model.similarity(embeddings, embeddings)

# tensor([[1.0000, 0.9429, 0.8880, 0.4558],
#
#         [0.9429, 1.0000, 0.9680, 0.5307],
#
#         [0.8880, 0.9680, 1.0000, 0.4933],
#
#         [0.4558, 0.5307, 0.4933, 1.0000]])

```

## ## Performance

The performance was evaluated on the [Semantic Textual Similarity (STS) 2017 dataset]([http://ixa2.si.ehu.es/stswiki/index.php/Main\\_Page](http://ixa2.si.ehu.es/stswiki/index.php/Main_Page)). The task is to predict the semantic similarity (on a scale 0-5) of two given sentences.

STS2017 has monolingual test data for English, Arabic, and Spanish, and cross-lingual test data for English-Arabic, -Spanish and -Turkish.

We extended the STS2017 and added cross-lingual test data for English-German, French-English, Italian-English, and Dutch-English ([STS2017-extended.zip](<https://public.ukp.informatik.tu-darmstadt.de/reimers/sentence-transformers/datasets/STS2017-extended.zip>)).

The performance is measured using Spearman correlation between the predicted similarity score and the gold score.

Model | AR-AR | AR-EN | ES-ES | ES-EN | EN-EN | TR-EN | EN-DE | FR-EN | IT-EN | NL-EN |

Average

---|---|---|---|---|---|---|---|---|---|---

XLM-RoBERTa mean pooling | 25.7 | 17.4 | 51.8 | 10.9 | 50.7 | 9.2 | 21.3 | 16.6 | 22.9 | 26.0 | 25.2

mBERT mean pooling | 50.9 | 16.7 | 56.7 | 21.5 | 54.4 | 16.0 | 33.9 | 33.0 | 34.0 | 35.6 | 35.3

LASER | 68.9 | 66.5 | 79.7 | 57.9 | 77.6 | 72.0 | 64.2 | 69.1 | 70.8 | 68.5 | 69.5

**\*\*Sentence Transformer Models\*\***

distiluse-base-multilingual-cased | 75.9 | 77.6 | 85.3 | 78.7 | 85.4 | 75.5 | 80.3 | 80.2 | 80.5 | 81.7 |

80.1

## Citation

If you use the code for multilingual models, feel free to cite our publication

[Making Monolingual Sentence Embeddings Multilingual using Knowledge

Distillation](<https://arxiv.org/abs/2004.09813>):

```
@article{reimers-2020-multilingual-sentence-bert,
```

```
  title = "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation",
```

```
  author = "Reimers, Nils and Gurevych, Iryna",
```

```
  journal= "arXiv preprint arXiv:2004.09813",
```

```
  month = "04",
```

```
  year = "2020",
```

```
  url = "http://arxiv.org/abs/2004.09813",
```

```
}
```

[ Previous](../adaptive\_layer/README.html "Adaptive Layers") [Next  
(../distillation/README.html "Model Distillation")

\* \* \*

(C) Copyright 2025.

Built with [Sphinx](https://www.sphinx-doc.org/) using a  
[theme](https://github.com/readthedocs/sphinx\_rtd\_theme) provided by [Read the  
Docs](https://readthedocs.org).