

Paraphrase detection

Paraphrase detection is the task of analyzing two segments of text and determining if they have the same meaning despite the differences in structure and wording.

It is hard to apply in real case scenarios, since each lexical form has a definite connotation.

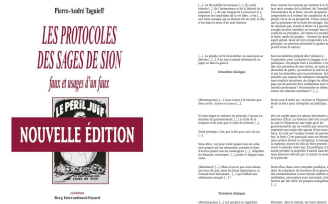
(quasi-) paraphrases, i.e., sentences or phrases that convey approximately the same meaning using different words (Bhagat, and Hovy, 2013).

Drawbacks of current benchmarks

- ⊖ are mainly in English
- ⊖ Area mainly constructed automatically
- ⊖ comprise short sentences
- ⊖ consist of paraphrase pairs with high word overlap
- ⊖ have simple syntactic structures
- ⊖ are mostly formulated as binary classification
- ⊖ do not provide retrieval tasks

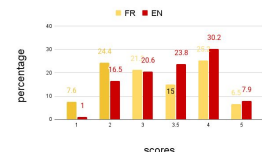
Dataset	Sent. length # ↑	Word overlap % ↓
SemEval2017T1	8.7 ± 3.34	23.8 ± 11.8
SICK	9.6 ± 3.69	29.2 ± 12.1
SemEval2015T2	11.5 ± 6.38	15.0 ± 11.8
SemEval2016T1	14.3 ± 19.45	26.2 ± 11.9
MRPC	19.7 ± 16.03	28.0 ± 8.1
PAWS	21.5 ± 5.42	40.4 ± 4.7
Our work	23.5 ± 13.64	10.3 ± 6.4

Datasets



Score	Label	Description
5	Very high	The two sentences are identical, or almost identical.
4	High	The key words are kept untouched (or using very close synonyms) in the paraphrase.
3.5	Medium-high	A few words are kept untouched, but the key words may have been replaced with synonyms or periphrastics.
3	Medium	The idea is similar but the terms and formulation are not obviously reused.
2	Low	It can be understood that the essence of both sentences is related upon a somewhat detailed reading
1	Very low	Based on the context or a subjective understanding, the two sentences seem to be somewhat related.

S	L	Protocoles	Dialogues
5	FR	Le résultat justifie les moyens.	La loi justifie les moyens.
4	EN	The ends justify the means.	The ends justify the means.
3.5	FR	La liberté politique est une idée et non un fait.	La liberté politique n'est qu'une idée relative.
3	EN	Political liberty is only a secondary idea.	Political freedom is an idea but not a fact.
3.5	FR	Il n'est pas que le monde puisse à sacrifier les biens de tous pour atteindre son propre bien.	Tout ce que nous ne pouvons pas sacrifier les biens de tous pour atteindre son propre bien.
3	EN	It is not that the world can sacrifice the goods of all for its own benefit.	All, or nearly all, are ready to sacrifice another's rights to their own interests.
3	FR	Il n'y a pas de loi qui ne soit basée sur la justice, de faire valoir les droits qui vaudraient sans faire de l'oppression et qui soient reconnus.	Dans un système politique, celui qui peut frapper immédiatement, et à tout bout, ne peut être élu, mais qu'il est élu.
3	EN	There have will evidence one by one all the intelligence and theories which have been permitted by the system, and our kingdom will be strengthened by a doctrine of such magnificent proportion as to be at any moment and in every place in a position to wipe out any system who appears to be dead or weak.	It calls for a vast system of legislation that takes back to its full all the theories that had been ignorantly founded - it calls a gigantic doctrine that could strike immediately and at any time all who resist or oppose.
2	FR	Pourquoi aurions-nous inventé et répété aux châtiments sans politique, sans loi, sans droit, sans justice, pourquoi, si ce n'est pour atteindre quelque chose que nous ne pouvons pas atteindre autrement ?	À quel service la politique, si ce n'est pour protéger, par des voies obligées, la loi qui est le point d'arrivée par la loi divine ?
2	EN	For what purpose then have we invented this whole policy and institution if it is the result of the gods without giving them any chance to examine its underlying meaning?	What's the use of political maneuvering if it can't attain the desired goal by devious ways, when straight ones are inadequate?
1	FR	L'idée de la liberté est indéfinissable, parce que personne ne sait en user dans son intérêt.	Une certaine latitude de l'homme, il y a des peuples incapables de modération dans l'exercice de la liberté.
1	EN	The idea of freedom is impossible of realization because no one knows how to use it with moderation.	In certain regions of Europe, there are people incapable of moderation in the exercise of liberty.



Experimental Setup

Tasks

Correlation: evaluate the correlation between the paraphrase similarity scores calculated using different representation models, and that assigned by human annotators.

Classification: given an input sentence from the *Protocols* marked as a paraphrase by the annotators, find the corresponding sentence in the full text of the *Dialogue* book.

Retrieval: using the full text of both books to find how many paraphrases a model is able to detect

correlation

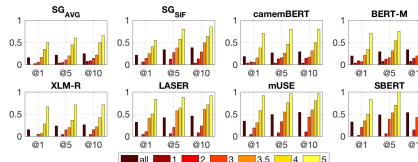
	SG _{AVG}	SG _{SIF}	camemBERT	BERT-T _{FF-DF}	BERT-M _{T_{FF}-DF}	XLM-R _{T_{FF}-DF}	LASER	mUSE	SBERT
ρ	0.30	0.55	0.44	0.38	0.38	0.19	0.58	0.66	0.59
r_s	0.33	0.54	0.45	0.40	0.40	0.19	0.58	0.66	0.60

TABLE 2. Pearson (ρ) and Spearman (r_s) correlation for each model on the French dataset.

	ConceptNet _{SIF}	SG _{SIF}	BERT-T _{FF-DF}	XLM-R _{T_{FF}-DF}	LASER	USE	mUSE	SBERT	AMR
ρ	0.54	0.51	0.35	0.26	0.41	0.64	0.56	0.62	0.50
r_s	0.52	0.48	0.38	0.27	0.49	0.61	0.54	0.59	0.44

TABLE 3. Pearson (ρ) and Spearman (r_s) correlation for each model on English texts.

Classification FR



Conclusion

- Dataset for qualitative evaluation
- Small but challenging
- We conducted the evaluation on two languages, and we plan to extend it to German, Italian, and Spanish
- It would be difficult to use the existing models in real-world scenarios, since they can mainly detect highly similar paraphrase sentences
- As much as our presented evaluation represents a real task, it is rather simplified. This is because we already presented to the systems the two books in which the similarities actually exist. In a real case scenario instead, the search space could be broader including an extensive collection of heterogeneous texts to search from
- Apart from the contribution of providing an evaluation of sentence representations in a real-world scenario, the more noticeable impact of this paper is to encourage the use of language technology in different fields with social impact. We aim to highlight the need to develop more efficient technologies to solve pressing issues, such as plagiarism detection or the spread of misinformation.

Models

Static embeddings: mean pooling of word2vec and ConceptNet Numberbat embeddings with different weighting schemas.

Contextualized embeddings: mean pooling of BERT, Camem-BERT and XLM-R embeddings (cosine similarity)

Sentence embeddings: LASER, USE, mUSE and SBERT (cosine similarity)

Explicit representation: Abstract Meaning Representation graphs (SMATCH score)

retrieval

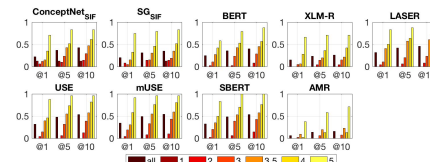
SG _{AVG}	SG _{SIF}	camemBERT	BERT-M	XLM-R _{T_{FF}-DF}	LASER	mUSE	SBERT
0.55	0.54	0.55	0.59	0.55	0.65	0.63	0.61

TABLE 0. AUC measure for each model on the French dataset.

ConceptNet _{SIF}	SG _{SIF}	BERT-T _{FF-DF}	XLM-R _{T_{FF}-DF}	LASER	USE	mUSE	SBERT	AMR
0.62	0.59	0.61	0.55	0.61	0.63	0.65	0.63	0.52

TABLE 1. AUC measure for each model on the English dataset.

Classification EN



Resources

