

# **Evaluating Multilingual Sentence Representation Models on a Real Case** Scenario







# Rocco Tripodi, Rexhina Blloshmi and Simon Levis Sullam

## Paraphrase detection

## Definition 1.

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.

Definition 2.

(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 $\# \uparrow$	Word overlap %↓
SemEval2017T1	$8.7 \pm 3.34$	$23.8 \pm 11.8$
SICK	$9.6 \pm 3.69$	$29.2 \pm 12.1$
SemEval2015T2	$11.5 \pm 6.38$	$15.0 \pm 11.8$
SemEval2016T1	$14.3 \pm 19.45$	$26.2 \pm 11.9$
MRPC	$19.7 \pm 16.03$	$28.0 \pm 8.1$
PAWS	$21.5 \pm 5.42$	$40.4 \pm 4.7$
Our work	$\textbf{23.5} \pm \textbf{13.64}$	$\textbf{10.3} \pm \textbf{6.4}$

## Datasets



engages for analytics of the little school [] and for analytic good determined the school of the little school of the school	compressions i anciene in consistente de sa polições est este prinçontes filosomorphismis que la professione de habitale seu completa framente, aporta professione de habitale seu completa promotiva en emplemente de monitor seu resulta entre la condicione se proceptos a destinata ha fonctione de condicione se proceptos a destinata ha fonctione de professione de processo primates ha positiva de la policita com esta de monitor de la policita de policita com esta destina.
(-) Le pople limit à lui mêtre, se aureupe se devaire. (-) ¿ le si sencionale administra, et pape, et la limit paren. Considere d'adapue	Next on childrich project des de missen à l' percenting per le manifer le françage et le stabil platiques als que qui en et la materia c'entan. de se appresson de ma sollan, es voje que be stancage et la perin, qu'encenten a sui die jonnest, a que le bibilité qu'en primeration a sui die jonnest, a que le bibilité qu'en primeration, si de de perindica l'amortine, de dépays les allatent plus prepurante passen les sanisheises autries anché promotes à l'aprendiction de départ de propurante passen les sanisheises autries anché promotes à l'aprendiction de des les constitutes de l'aprendiction de des les constitutes de l'aprendiction de des les constitutes de l'aprendiction de de l'aprendiction de l'aprendiction de l'aprendiction de de l'aprendiction de l'aprendiction de l'aprendiction de l'aprendiction de de l'aprendiction de
Observation [] was store it is bracke per described in both of large [ . ].	Name and a wise and is discover Department. Search Serviced strengther or publique, nations.
We were highly involved as an appealant. J assess on a matter follower belong E. [1] is don't the symmetric size plan participate, the follower []. Your principe, it've que has pass and size of []. Your first principe, it've participate design and size of the first section of the passed belongs as the same quanticipate advanced in passed by land for the first section of the passed of the land first section of the first section of the land first section of the first section of the section.	Given a soluble dustion in bittoms information com- bination of first of processing and processing as a source of expressions are change to be a formation of the complete are change as for a common and a source of a source of a first of the complete and a source of a common and a source of a first of the complete and a first of the complete and a first of the complete and a a mallown common as to one principal and a mallown common as a principal and a mallown common as a principal and a procession.
(Martiani) [] Main rive or parms native, springs they done to their depotypes. In the set of the depotypes. In the set of the depotypes. In the set of	Name than Amounts recognite participa, a lo- dress de exceptor de l'emerce de l'apparen par de contrabatement a como mobile confront plans publishes, escenaires pour reconnels units arrives qui bié abile se peupon emeglément. 11
Treatme damps	
Опторона 1.1 за рафи и пребит	Portion, then is more, to more phane, figure,

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

S	L	Protocols	Dialogue
5	FR	Le résultat justifie les moyens.	La fin justifie les moyens.
	EN	The ends justify the means.	The result justifies the means.
4	FR	La liberté politique est une idée et non un fait.	La liberté politique n'est qu'une idée relative.
	EN	Political liberty is only a secondary idea.	Political freedom is an idea but not a fact.
3.5	FR	Il en est peu qui ne soient prêts à sacrifier les biens de tous pour atteindre leur propre bien.	Tous ou presque tous sont prêts à sacrifier les droits d'autrui à leurs innérêts.
	EN	and rare indeed are the men who would not be willing to sacrifice the welfare of all for the sake of securing their own welfare.	All, or nearly all, are ready to sacrifice another's rights to their own interests.
3	FR	Notre règne se signalera par un despotisme si majestueux qu'il sera en état, en tout temps et en tout lieu, de faire tuire les chrétiens qui voudront nous faire de l'opposition et qui seront mécontents.	Dans un despotisme gigantesque, enfin, qui puisse frapper immédiatement, et à toute heure, tout ce qui résiste, tout ce qui se plaint.
	EN	These loves will withdraw one by one all the indulgences and liberties which have been permitted by the poytin, and our intigation will be distinguished by a despotism of such magnificent proportions as to be at any moment and in every place in a position to wipe out any goytim who oppone us by deed or wool.	It calls for a wast system of legislation that takes back bit by bit all the liberties that had been improdently bestowed – in sum, a gigantic despoisan that could strike immediately and at any time all who resist and complain.
2	FR	Pourquoi aurions-nous inventé et inspiré aux chrétiens toute cette politique, suns leur donner les moyens de la pénétrer, pourquoi, si ce n'est pour atteindre secrétement ce que notre race dispersée ne pouvait atteindre directement ?	À quoi servirait la politique, si l'on ne pouvait gagner, par des voies obliques, le but qui ne peut s'atteindre par la ligne droite ?
	EN	For what purpose then have we invented this whole policy and insinuated it into the minds of the goys 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 irréalisable, parce que personne ne sait en user dans une juste mesure.	Sous certaines latitudes de l'Europe, il y a des peuples incupables de modération dans l'exercice de la liberté.
		The idea of freedom is impossible of realization because no one knows how	In certain regions of Europe, there are regule incanable of moderation



### **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

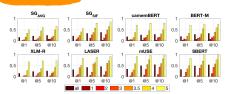
	SGAVG	$SG_{SIF}$	camen	BERT <sub>TF-IDF</sub>	BERT-MTF-IDF	XLM-R <sub>T</sub>	F-IDF	LASER	mUSE	SBERT
ρ	0.30	0.55		0.44	0.38	0.28		0.57	0.66	0.59
$r_s$	0.33	0.54		0.45	0.40	0.19		0.58	0.66	0.60
	Inoic 4	. Pearso	on (p) a	nd Spearman	$(r_s)$ correlation	for each n	odel o	on the Fre	nch datas	et.
_				•						
0	Concept 0.54		on (ρ) as SG <sub>SIF</sub> 0.51	BERT <sub>TF-IDF</sub>	(r <sub>s</sub> ) correlation XLM-R <sub>TF-IDF</sub> 0.26	LASER 0.41	USE 0.64	mUSE	SBERT 0.62	AMI 0.50

ranse 3: Pearson ( $\rho$ ) and Spearman ( $r_s$ ) correlation for each model on English texts

SGSIF BERTTF-IDF XLM-RTF-IDF LASER USE MUSE SBERT AMR 0.55 0.61

TADJE /. AUC measure for each model on the English dataset.

#### Classification FR



#### Conclusion

- we already presented to the systems the two books in which the similarities actually exist. In a real
- 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

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

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

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

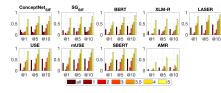
Explicit representation: Abstract Meaning Representation graphs

### retrieval

0.55	0.54	0.55	0.59	0.55	0.65	0.63	0.61		
Table 6: AUC measure for each model on the French dataset.									

ATTC	and the second state of the second	. Particle dates	

#### Classification EN



#### Resources



https://github.com/roccotrip/protocols