

cases	doc_1		doc_2		decision	id
					DUPLICATES	112
			authors	<ul style="list-style-type: none">Hitomi YanakaKoji Mineshima		
	authors	<ul style="list-style-type: none">Yanaka, H.Mineshima, K.	title	Compositional Evaluation on Japanese Textual Entailment and Similarity		
	title	Compositional Evaluation on Japanese Textual Entailment and Similarity	publication_date	2022-08-09 15:10:56+00:00		
	publication_date	2022-01-01 00:00:00	source	SupportedSources.ARXIV		
	source	SupportedSources.CROSSREF	journal	None		
	journal		volume			
	volume		doi			
	doi	10.1162/tacl_a_00518	urls	<ul style="list-style-type: none">http://arxiv.org/pdf/2208.04826v1http://arxiv.org/abs/2208.04826v1http://arxiv.org/pdf/2208.04826v1		
	urls	<ul style="list-style-type: none">https://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00518/2060724/tacl_a_00518.pdfhttps://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00518/2060724/tacl_a_00518.pdfhttp://dx.doi.org/10.1162/tacl_a_00518	id	id3687970207881057749		
	id	id-2736493335266370182	abstract	Natural Language Inference (NLI) and Semantic Textual Similarity (STS) are widely used benchmark tasks for compositional evaluation of pre-trained language models. Despite growing interest in linguistic universals, most NLI/STS studies have focused almost exclusively on English. In particular, there are no available multilingual NLI/STS datasets in Japanese, which is typologically different from English and can shed light on the currently controversial behavior of language models in matters such as sensitivity to word order and case particles. Against this background, we introduce JSICK, a Japanese NLI/STS dataset that was manually translated from the English dataset SICK. We also present a stress-test dataset for compositional inference, created by transforming syntactic structures of sentences in JSICK to investigate whether language models are sensitive to word order and case particles. We conduct baseline experiments on different pre-trained language models and compare the performance of multilingual models when applied to Japanese and other languages. The results of the stress-test experiments suggest that the current pre-trained language models are insensitive to word order and case marking.		
	abstract		versions			
	versions					