	doc_1	doc_2	decision	id
cases	<table><tr><td>authors</td><td>• Hitomi Yanaka • K. Mineshima</td></tr><tr><td>title</td><td>Compositional Evaluation on Japanese Textual Entailment and Similarity</td></tr><tr><td>publication_date</td><td>2022-08-09 00:00:00</td></tr><tr><td>source</td><td>SupportedSources.SEMANTIC_SCHOLAR</td></tr><tr><td>journal</td><td>Transactions of the Association for Computational Linguistics</td></tr><tr><td>volume</td><td>10</td></tr><tr><td>doi</td><td>10.1162/tacl_a_00518</td></tr><tr><td>urls</td><td>• https://www.semanticscholar.org/paper/089fd688de463519a68bd25b11ae1c3eb57b207d</td></tr><tr><td>id</td><td>id83644125427564708 32</td></tr><tr><td>abstract</td><td>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.</td></tr><tr><td>versions</td><td></td></tr></table>	<table><tr><td>authors</td><td>• Hitomi Yanaka • Koji Mineshima</td></tr><tr><td>title</td><td>Compositional Evaluation on Japanese Textual Entailment and Similarity</td></tr><tr><td>publication_date</td><td>2022-08-09 15:10:56+00:00</td></tr><tr><td>source</td><td>SupportedSources.ARXIV</td></tr><tr><td>journal</td><td>None</td></tr><tr><td>volume</td><td></td></tr><tr><td>doi</td><td></td></tr><tr><td>urls</td><td>• http://arxiv.org/pdf/2208.04826v1 • http://arxiv.org/abs/2208.04826v1 • http://arxiv.org/pdf/2208.04826v1</td></tr><tr><td>id</td><td>id3687970207881057749</td></tr><tr><td>abstract</td><td>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.</td></tr><tr><td>versions</td><td></td></tr></table>	DUPLICATES	116