cases	doc_1		doc_2		decision	id
	authors	• Yiren Jian and Chongyang Gao and Soroush Vosoughi	**authors**	• Yiren Jian • Chongyang Gao • Soroush Vosoughi	DUPLICATES	148
	title	Non-Linguistic Supervision for Contrastive Learning of Sentence Embeddings	**title**	Non-Linguistic Supervision for Contrastive Learning of Sentence Embeddings		
	publication_date	2022-09-20 00:00:00	**publication_date**	2022-09-20 00:00:00		
	source	SupportedSources.INTERNET_ARCHIVE	**source**	SupportedSources.SEMANTIC_SCHOLAR		
	journal		**journal**	ArXiv		
	volume		**volume**	abs/2209.09433		
	doi		**doi**	10.48550/arXiv.2209.09433		
	urls	• https://web.archive.org/web/20220926133149/https://arxiv.org/pdf/2209.09433v1.pdf	**urls**	• https://www.semanticscholar.org/paper/898072d734344c1c2ea5583d070158dbea733c67		
	id	id-3003001293162505140	**id**	id7904089235936949420		
	abstract	Semantic representation learning for sentences is an important and well-studied problem in NLP. The current trend for this task involves training a Transformer-based sentence encoder through a contrastive objective with text, i.e., clustering sentences with semantically similar meanings and scattering others. In this work, we find the performance of Transformer models as sentence encoders can be improved by training with multi-modal multi-task losses, using unpaired examples from another modality (e.g., sentences and unrelated image/audio data). In particular, besides learning by the contrastive loss on text, our model clusters examples from a non-linguistic domain (e.g., visual/audio) with a similar contrastive loss at the same time. The reliance of our framework on unpaired non-linguistic data makes it language-agnostic, enabling it to be widely applicable beyond English NLP. Experiments on 7 semantic textual similarity benchmarks reveal that models trained with the additional non-linguistic (images/audio) contrastive objective lead to higher quality sentence embeddings. This indicates that Transformer models are able to generalize better by doing a similar task (i.e., clustering) with unpaired examples from different modalities in a multi-task fashion.	**abstract**	Semantic representation learning for sentences is an important and well-studied problem in NLP. The current trend for this task involves training a Transformer-based sentence encoder through a contrastive objective with text, i.e., clustering sentences with semantically similar meanings and scattering others. In this work, we ï¬nd the performance of Transformer models as sentence encoders can be improved by training with multi-modal multi-task losses, using unpaired examples from another modality (e.g., sentences and unrelated image/audio data). In particular, besides learning by the contrastive loss on text, our model clusters examples from a non-linguistic domain (e.g., visual/audio) with a similar contrastive loss at the same time. The reliance of our framework on unpaired non-linguistic data makes it language-agnostic, enabling it to be widely applicable beyond English NLP. Experiments on 7 semantic textual similarity benchmarks reveal that models trained with the additional non-linguistic (images/audio) contrastive objective lead to higher quality sentence embeddings. This indicates that Transformer models are able to generalize better by doing a similar task (i.e., clustering) with unpaired examples from different modalities in a multi-task fashion. The code is available at https://github.com/yiren-jian/NonLing-CSE. outperfom SimCSE in the transfer benchmarks, though some improvements are marginal. These ï¬ndings show that the representations learned by our framework can be successfully applied to downstream tasks.		
	versions		**versions**			