Syntactic Interchangeability in Word Embedding Models

Daniel Hershcovich & Assaf Toledo & Alon Halfon & Noam Slonim

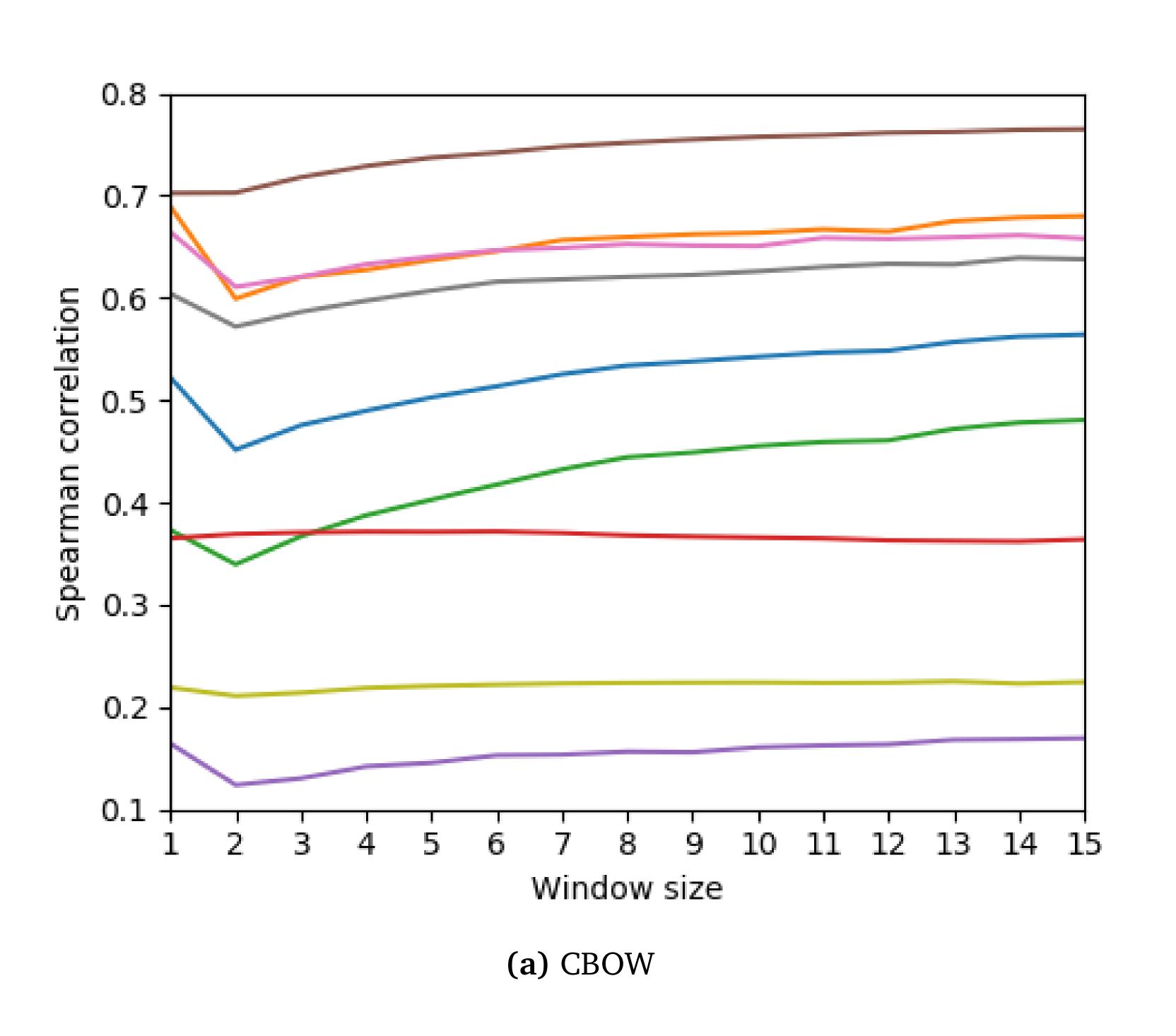
IBM Research

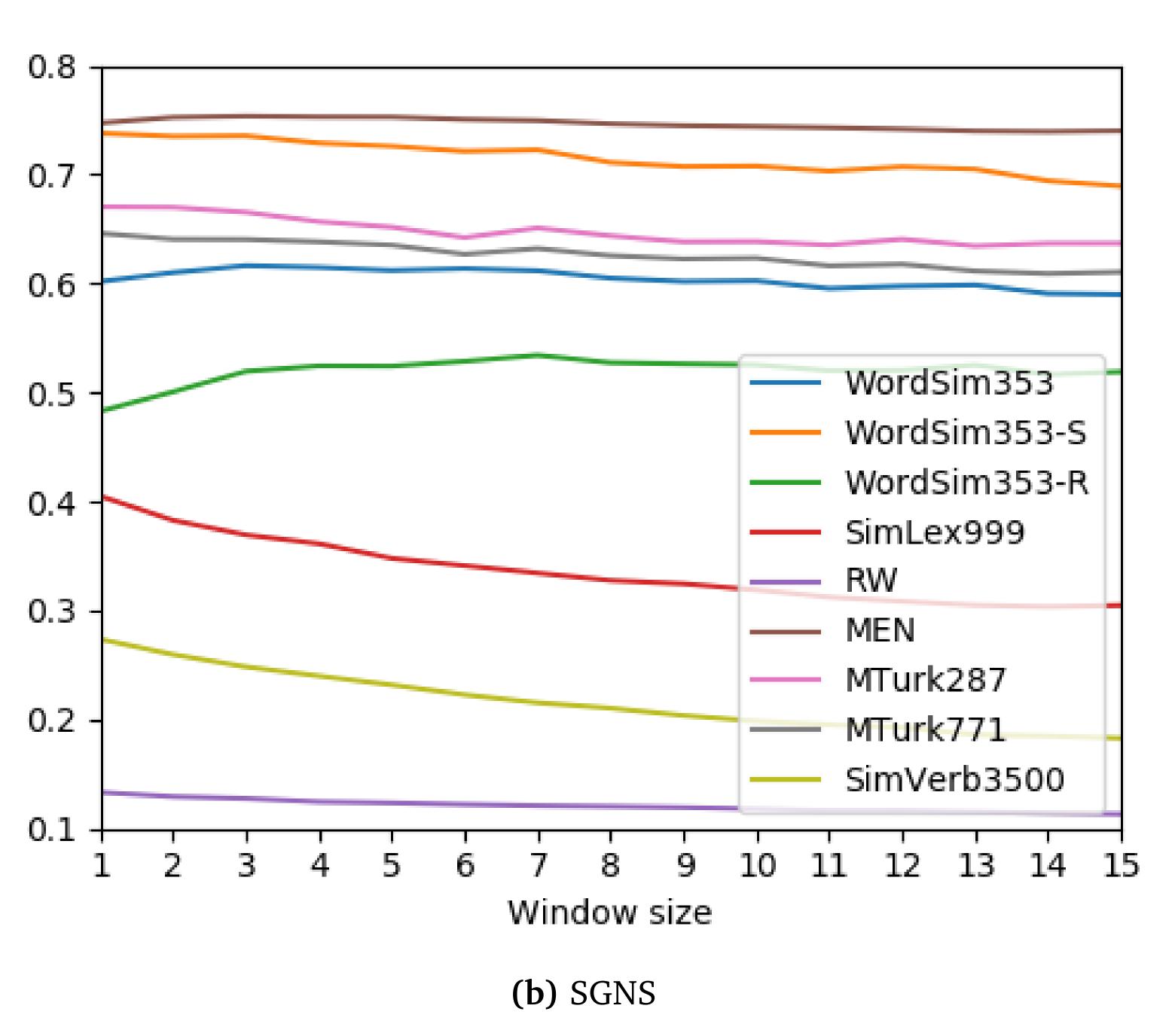
Do word embeddings capture part-of-speech? Explaining the effect of context window size on word similarity evaluation.

Word Similarity and Relatedness Benchmarks

300-dimensional fastText embeddings trained on Wikipedia with the CBOW and SGNS algorithms.

Measuring cosine similarity between words in each benchmark pair.

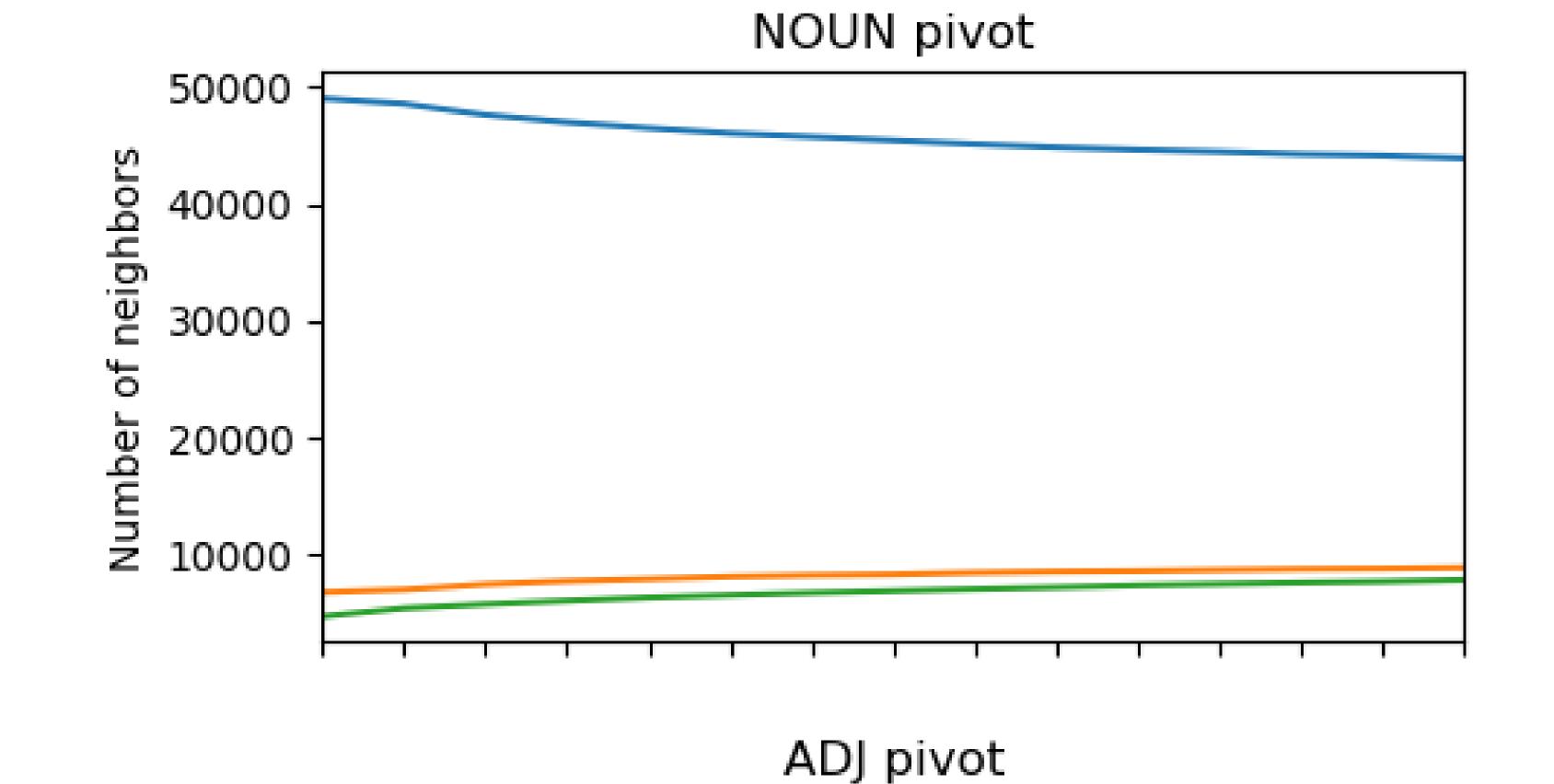


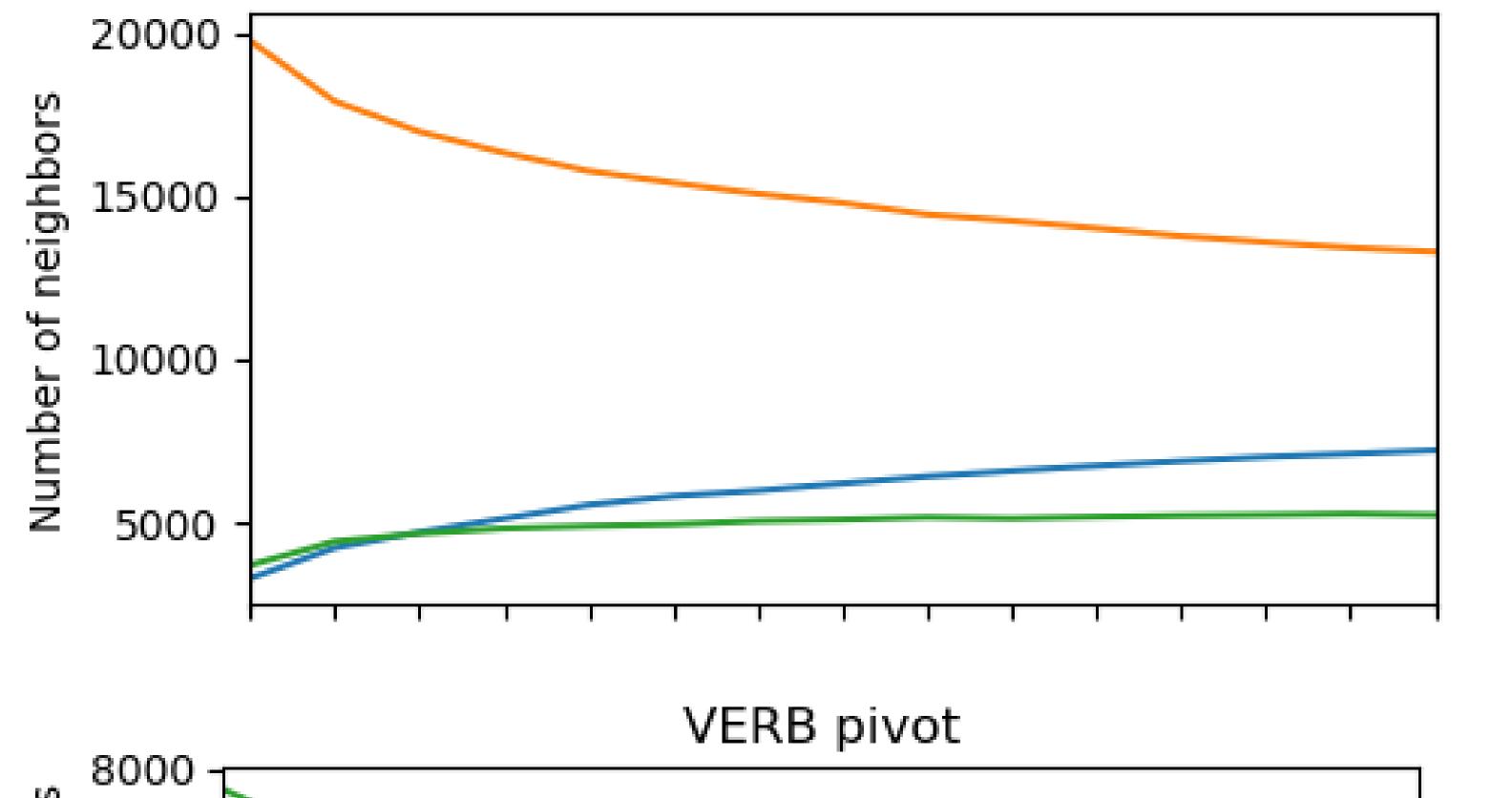


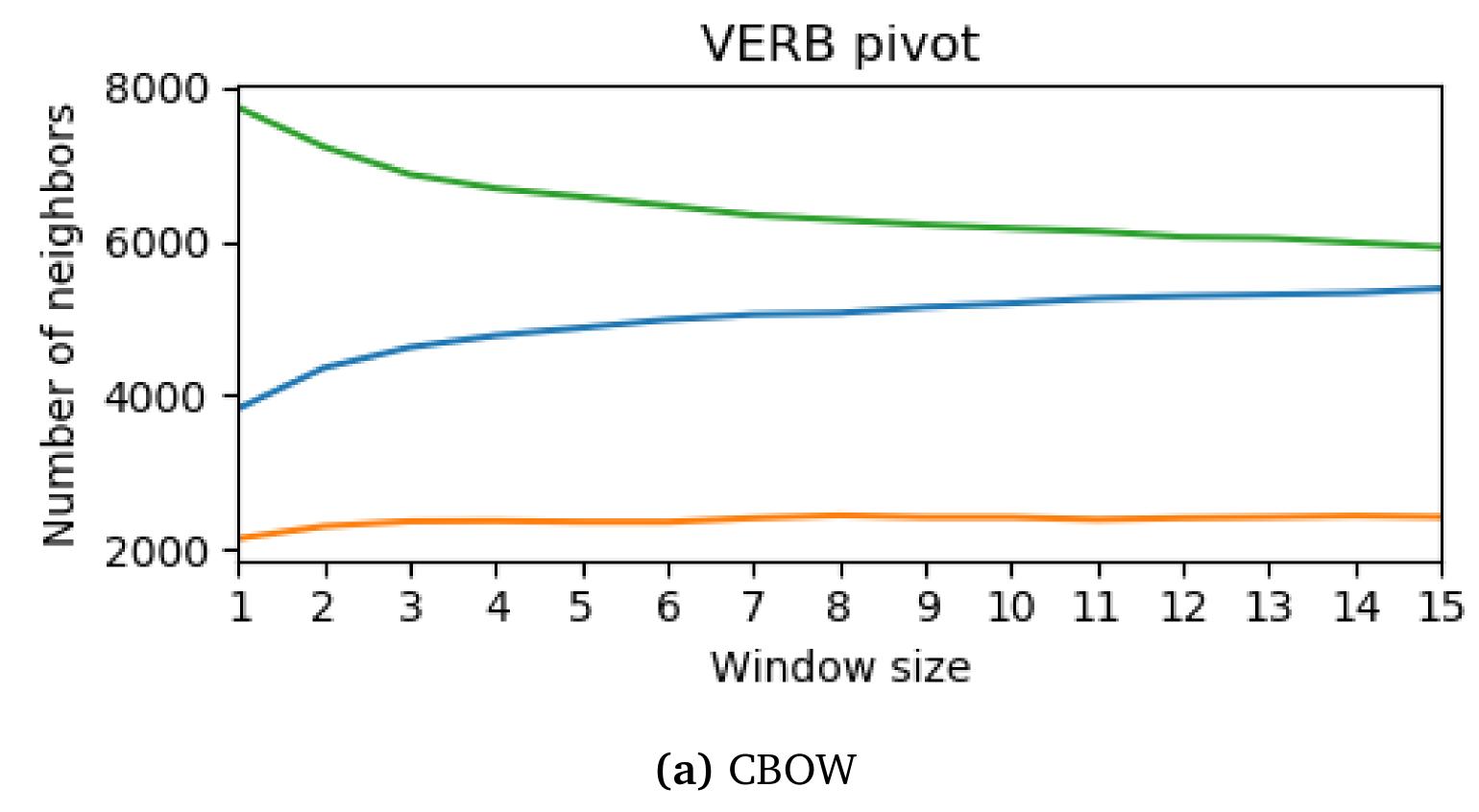
Number of Neighbors per POS

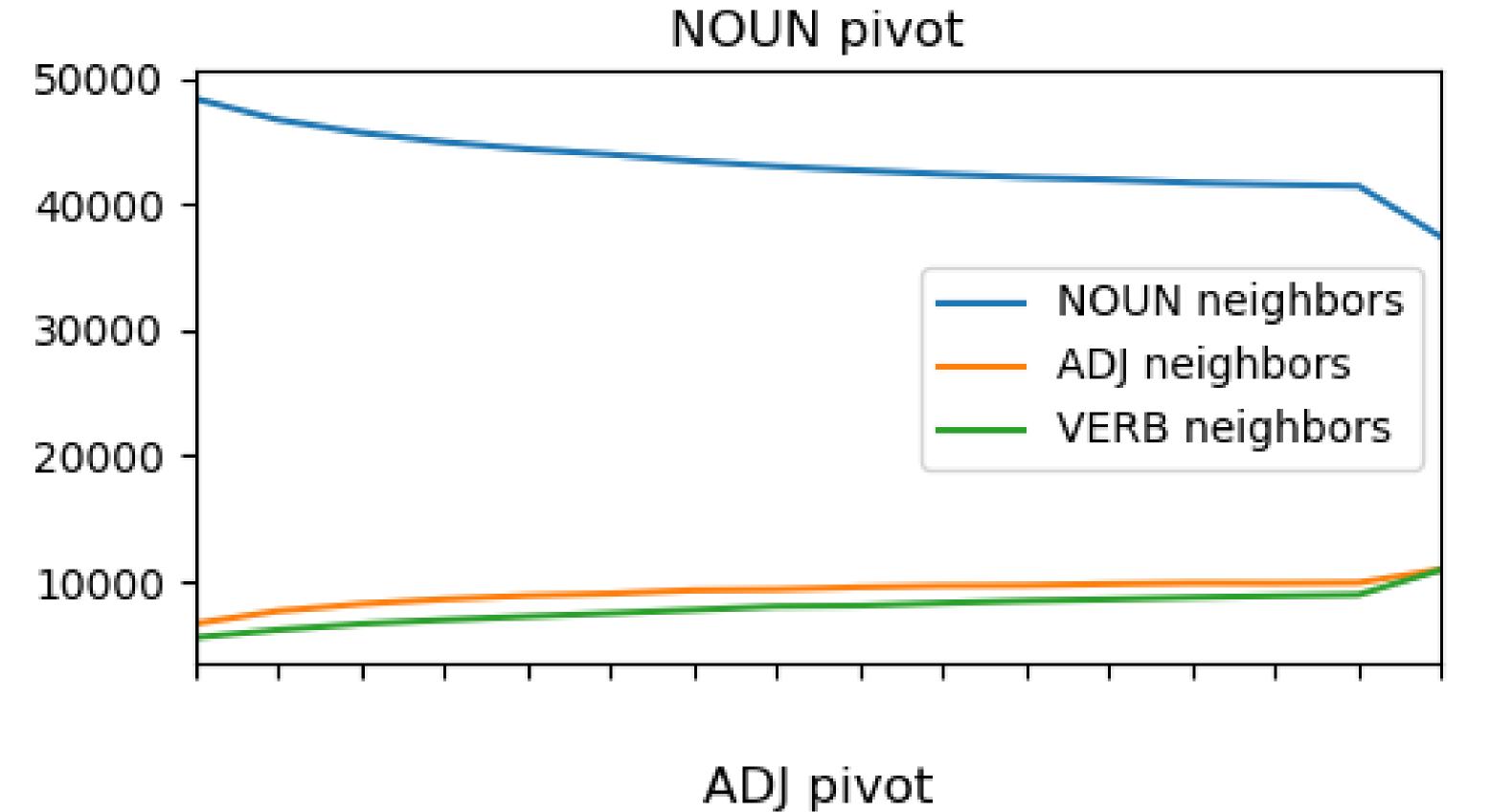
Context window size is negatively correlated with interchangeability.

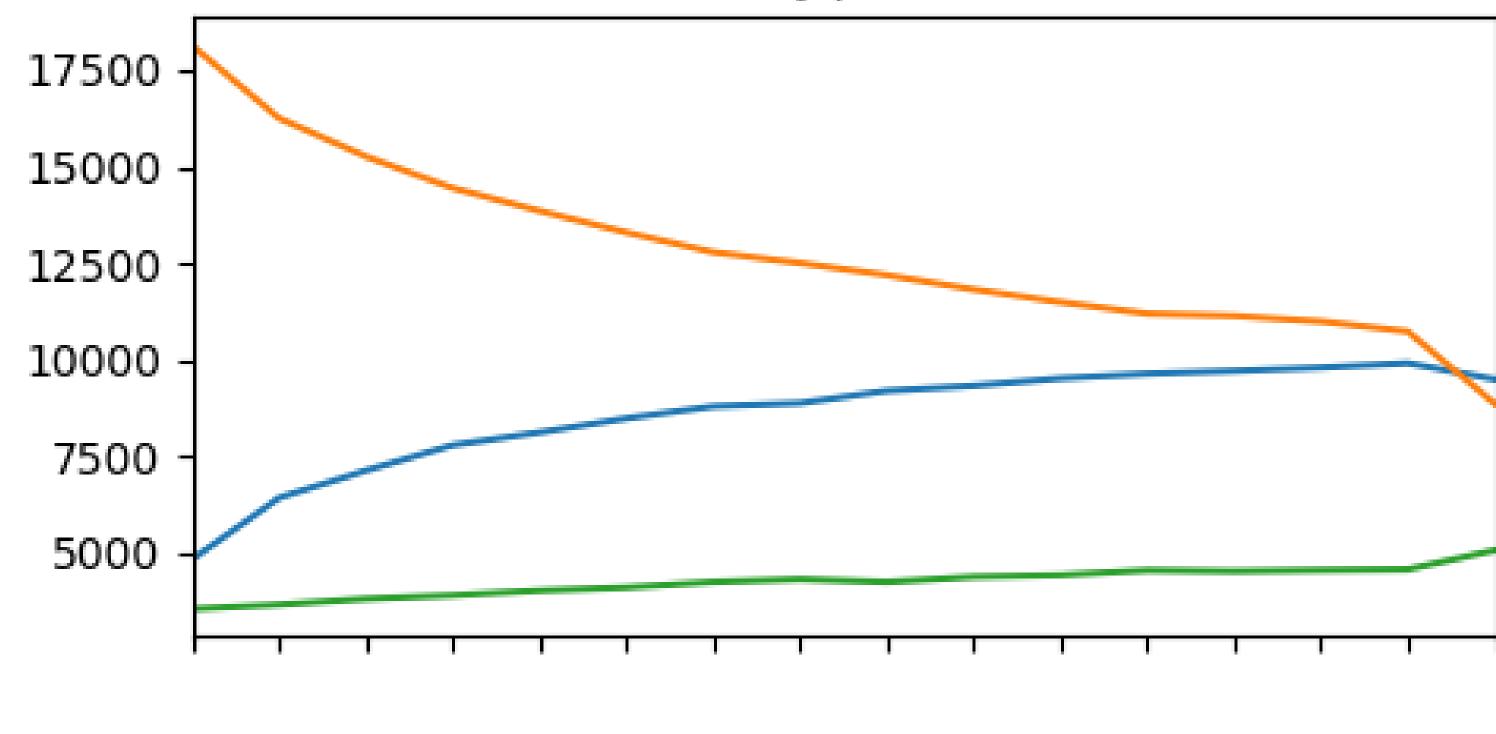


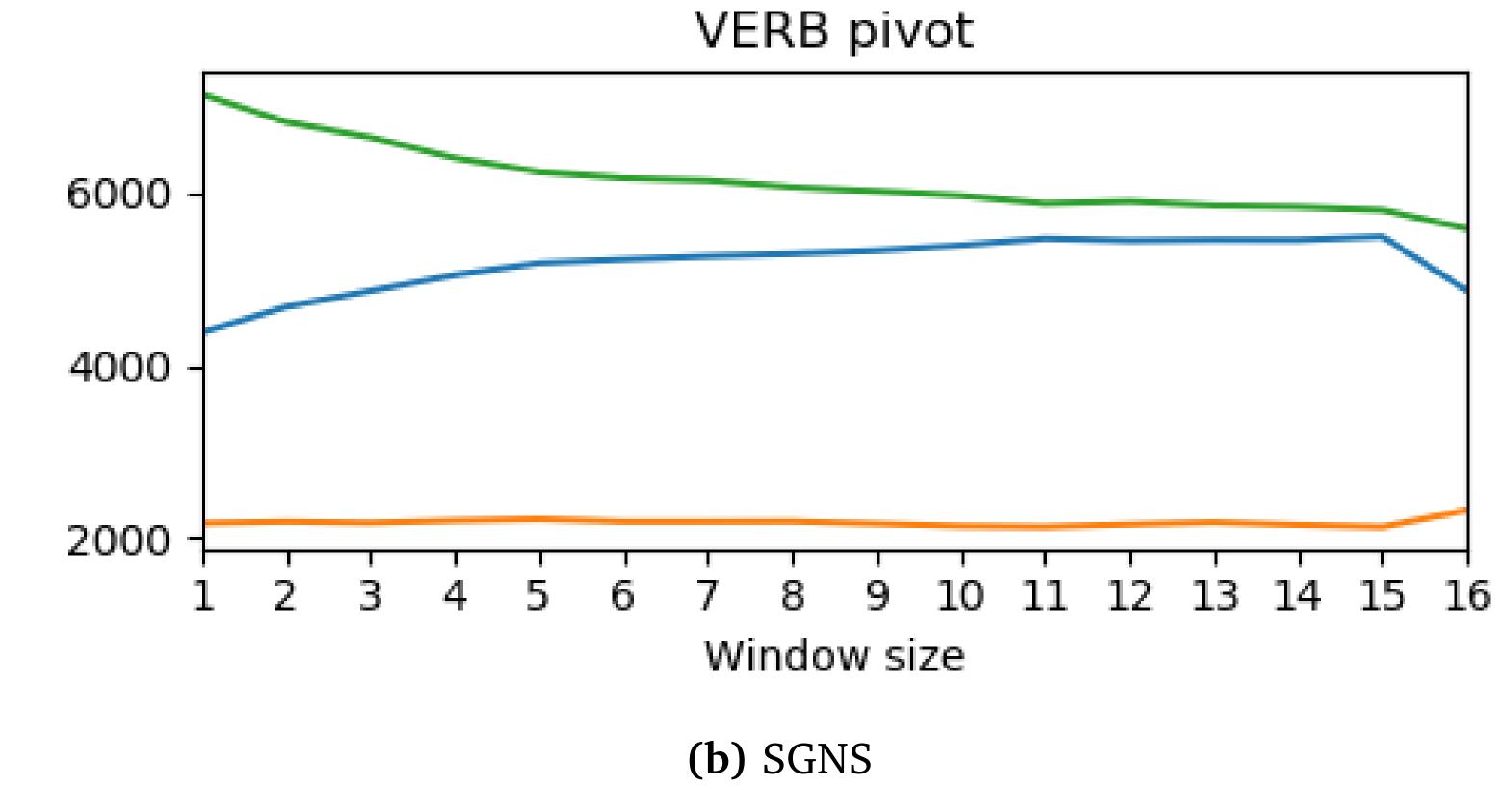












Interchangeability Enrichment in Benchmarks and in Word Embedding Models

Words are approximately syntactically interchangeable if they have the same POS (part-of-speech).

		Δ win $=$	$2 \rightarrow 15(\%)$	#	Related	# T	Jnrelated	
Benchmark	Size	CBOW	SGNS	All	Same-POS	All	Same-POS	p-value
WordSim353	353	24	-3	122	107	53	40	0.038
WordSim353-S	203	13	-6	60	53	53	40	0.061
WordSim353-R	252	42	4	104	89	39	31	0.26
SimLex999	999	-1	-20	234	199	334	295	0.897
RW	2034	37	-12	944	555	262	144	0.149
MEN	3000	9	-2	791	564	781	439	$3\cdot 10^{-10}$
MTurk287	287	8	-5	49	39	119	68	0.004
MTurk771	771	12	-5	204	153	200	146	0.365
SimVerb3500	3500	6	-30	633	265	1217	566	0.974

Algo-	NOUN	ADJ	VERB		
rithm	1 15 r	1 15 r	1 15 r		
CBOW	79 70 -0.96	72 48 -0.93	55 41 -0.91		
SGNS	78 66 -0.95	66 39 -0.94	51 41 -0.92		