Stranger than paradigms Word Embedding Benchmarks Don't Align With Morphology Timothee Mickus, University of Helsinki Maria Copot, The Ohio State University

What we do

RQ: We use word embeddings for morphology studies, but do we evaluate and train them properly?

- We design benchmarks to evaluate word2vec SGNS embeddings for French morphology
- We conduct an extensive hyperparameter exploration across 1200 models
- We study how hyperparameter choices affect performance on morphology as well as traditional NLP benchmark

Take-home messages

• The negative sampling exponent is a **crucial** hyperparameter: it must be **high** for inflection but **low** for NLP lexical semantic tasks.

A high negative sampling exponent forces us to distinguish all embeddings from frequent (and morphologically ambiguous) words. A low negative sampling exponent forces us to distinguish all embeddings from infrequent (and semantically varied) words.

• The optimal SGNS embeddings for morphology are suboptimal for commonplace NLP benchmarks & vice-versa.

Morphology or semantics?

Performance on morphology tasks anticorrelates with performance on NLP lexical semantics benchmarks

1	0.98	0.77	0.7	0.48	0.37	0.43	0.21	-0.59	-0.57	-0.73	-0.43	-0.69	-0.49
0.98	1	0.73	0.62	0.35	0.23	0.3	0.14	-0.67	-0.65	-0.79	-0.52	-0.72	-0.59
0.77	0.73	1	0.71	0.53	0.47	0.46	0.18	-0.26	-0.22	-0.43	-0.17	-0.52	-0.29
0.7	0.62	0.71	1	0.7	0.65	0.66	0.37	-0.21	-0.12	-0.33	-0.04	-0.39	-0.11
0.48	0.35	0.53	0.7	1	0.97	0.96	0.48	-0.06	-0.02	-0.2	0.12	-0.38	0.2
0.37	0.23	0.47	0.65	0.97	1	0.95	0.41	0.06	0.1	-0.07	0.2	-0.31	0.29
0.43	0.3	0.46	0.66	0.96	0.95	1	0.4	-0.1	-0.05	-0.21	0.06	-0.42	0.2
0.21	0.14	0.18	0.37	0.48	0.41	0.4	1	0.14	0.16	0	0.37	0.25	0.49
-0.59	-0.67	-0.26	-0.21	-0.06	0.06	-0.1	0.14	1	0.94	0.94	0.87	0.79	0.75
-0.57	-0.65	-0.22	-0.12	-0.02	0.1	-0.05	0.16	0.94	1	0.94	0.84	0.78	0.78
-0.73	-0.79	-0.43	-0.33	-0.2	-0.07	-0.21	0	0.94	0.94	1	0.8	0.83	0.72
-0.43	-0.52	-0.17	-0.04	0.12	0.2	0.06	0.37	0.87	0.84	0.8	1	0.73	0.76
-0.69	-0.72	-0.52	-0.39	-0.38	-0.31	-0.42	0.25	0.79	0.78	0.83	0.73	1	0.67
-0.49	-0.59	-0.29	-0.11	0.2	0.29	0.2	0.49	0.75	0.78	0.72	0.76	0.67	1
९० ०	50 (<u>کې</u> (50 0	s 0	ð (<u>کي کر</u>	مي آي آي	SGR	19	50 4	str is	U. O	5
	1 0.98 0.77 0.7 0.48 0.43 0.43 0.21 -0.59 -0.57 -0.57 -0.57 -0.69 -0.43 -0.43	1 0.98 0.98 1 0.77 0.73 0.77 0.73 0.77 0.62 0.48 0.35 0.37 0.23 0.43 0.34 0.43 0.34 0.43 0.34 0.43 0.34 0.43 0.34 0.43 0.34 0.43 0.34 0.43 0.34 0.43 0.44 0.43 0.44 0.43 0.44 0.43 0.44 0.43 0.44 -0.57 -0.65 -0.43 -0.72 -0.43 -0.52 -0.44 -0.59	1 0.988 0.771 0.988 1 0.731 0.777 0.732 0.711 0.77 0.622 0.711 0.748 0.325 0.471 0.377 0.622 0.471 0.373 0.233 0.472 0.371 0.231 0.471 0.373 0.323 0.471 0.431 0.335 0.472 0.432 0.343 0.461 0.433 0.343 0.462 0.443 0.343 0.463 0.443 0.414 0.183 -0.559 -0.663 -0.223 -0.433 -0.732 -0.433 -0.433 -0.742 -0.433 -0.439 -0.752 -0.433 -0.439 -0.559 -0.433	10.980.770.70.9810.730.620.770.7310.710.70.620.7110.480.350.530.70.370.230.470.650.430.30.460.660.430.140.180.370.530.670.220.120.570.650.220.120.430.520.170.430.440.550.630.330.550.520.170.430.640.550.550.550.640.550.250.11	1 0.988 0.777 0.73 0.73 0.622 0.355 0.777 0.733 1 0.71 0.73 0.77 0.622 0.71 1 0.73 0.77 0.622 0.71 1 0.73 0.77 0.622 0.71 1 0.74 0.48 0.355 0.533 0.74 1 0.37 0.233 0.47 0.65 0.97 0.37 0.233 0.47 0.65 0.97 0.37 0.23 0.47 0.65 0.97 0.37 0.23 0.47 0.65 0.97 0.43 0.33 0.47 0.63 0.97 0.43 0.14 0.18 0.37 0.48 0.059 -0.26 -0.22 0.12 0.02 -0.43 -0.52 -0.17 0.04 0.12 -0.43 -0.55 -0.55 -0.39 -0.38 -0.44 -0.55 -0.55 -0.55 -0.55	1 0.98 0.77 0.7 0.48 0.37 0.98 1 0.73 0.62 0.35 0.23 0.77 0.73 1 0.71 0.53 0.47 0.77 0.73 1 0.71 0.53 0.47 0.70 0.62 0.71 1 0.73 0.47 0.71 0.62 0.71 1 0.77 0.65 0.48 0.35 0.53 0.7 1 0.97 0.48 0.35 0.53 0.7 1 0.97 0.37 0.23 0.47 0.65 0.97 1 0.43 0.33 0.46 0.66 0.97 1 0.43 0.14 0.18 0.37 0.48 0.41 -0.57 0.67 0.22 0.12 0.02 0.11 -0.43 0.59 0.12 0.39 0.39 0.31 0.21 -0.43 0.59 0.12 0.39 0.39 0.39 0.31 0.21 -0.49 0.59 <t< td=""><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.98 1 0.73 0.62 0.35 0.23 0.3 0.77 0.73 1 0.71 0.53 0.47 0.46 0.77 0.73 1 0.71 0.53 0.47 0.46 0.7 0.62 0.71 1 0.73 0.46 0.67 0.46 0.7 0.62 0.71 1 0.77 0.65 0.66 0.48 0.35 0.53 0.7 1 0.97 0.96 0.37 0.23 0.47 0.65 0.97 1 0.95 0.37 0.23 0.47 0.65 0.97 1 0.96 0.43 0.3 0.46 0.66 0.96 0.91 0.16 0.50 0.67 0.26 0.21 0.06 0.01 0.01 0.02 0.01 0.02 0.01 0.21 0.02 0.02 0.11 0.24 0.14 0.42 0.24 0.24 0.24 0.24 <t< td=""><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.77 0.62 0.71 1 0.77 0.46 0.37 0.48 0.35 0.71 1 0.77 0.46 0.37 0.48 0.35 0.71 1 0.77 0.46 0.37 0.48 0.35 0.53 0.7 1 0.97 0.99 0.48 0.37 0.23 0.47 0.65 0.97 1 0.99 0.41 0.43 0.33 0.46 0.66 0.96 0.97 1 0.99 0.41 0.43 0.31 0.46 0.46 0.96 0.91 0.41 0.41 0.43 0.47 0.43 0.47 0.43 0.41 0.41 0.41 0.43 0.47 0.43 0.43 0.43 0.42</td><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.59 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.67 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.26 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.26 0.77 0.62 0.71 1 0.77 0.66 0.37 0.26 0.77 0.62 0.71 1 0.77 0.66 0.37 0.26 0.44 0.35 0.77 1 0.97 0.96 0.48 0.06 0.43 0.33 0.47 0.66 0.97 1 0.95 0.41 0.48 0.43 0.33 0.46 0.96 0.95 1 0.44 0.14 0.53 0.67 0.62 0.12 0.06 0.14 0.41 0.4 0.14 0.54 0.56 0.22 0.12 0.06 0.14 0.44</td><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.59 0.57 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.67 0.65 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 -0.22 0.70 0.62 0.71 1 0.77 0.65 0.66 0.37 -0.21 -0.12 0.48 0.35 0.53 0.7 1 0.97 0.96 0.48 -0.22 -0.22 0.48 0.35 0.53 0.7 1 0.97 0.96 0.48 -0.21 -0.12 0.43 0.33 0.47 0.65 0.97 1 0.95 0.48 -0.21 -0.02 0.43 0.33 0.46 0.99 0.95 1 0.49 0.11 -0.49 0.43 0.41 0.18 0.47 0.48 0.41 0.41 0.41 0.41 0.59 0.67 0.22 0.12 0.01 0.14 0.</td><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 -0.59 -0.73 -0.73 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 -0.67 -0.65 -0.79 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 -0.22 -0.43 0.77 0.62 0.71 1 0.77 0.65 0.66 0.37 -0.21 -0.12 -0.33 0.48 0.35 0.53 0.7 1 0.77 0.66 0.37 -0.21 -0.12 -0.33 0.48 0.35 0.53 0.7 1 0.77 0.46 0.48 -0.06 -0.12 -0.12 -0.33 0.43 0.23 0.47 0.65 0.47 0.48 0.41 0.48 -0.41 -0.49 -0.41 -0.43 -0.41 0.45 -0.41 -0.41 -0.41 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 <</td><td> 1. 0.98 0.77 0.70 0.48 0.37 0.43 0.44 0.45 0.45 0.45 0.46 0.44 0.47 0.45 0.47 0.43 0.44 0.47 0.45 0.47 0.43 0.44 0.44 0.44 0.44 0.45 0.45 0.45 0.46 0.47 0.48 0.41 0.44 0.45 0.45 0.45 0.44 0.45 0.45 0.46 0.47 0.48 0.49 0.49 0.41 0.49 0.41 0.44 0.4</td><td>1 0.98 0.77 0.48 0.37 0.43 0.21 0.59 0.73 0.43 0.49 0.98 1 0.73 0.62 0.62 0.35 0.23 0.43 0.14 0.67 0.65 0.79 0.73 0.43 0.47 0.48 0.18 0.26 0.22 0.43 0.17 0.52 0.77 0.53 0.17 0.53 0.47 0.46 0.18 0.26 0.22 0.43 0.17 0.53 0.70 0.62 0.71 1 0.77 0.65 0.66 0.37 0.21 0.12 0.33 0.04 0.33 0.48 0.35 0.53 0.77 1 0.95 0.48 0.06 0.10 0.12 0.33 0.04 0.33 0.43 0.35 0.47 0.46 0.97 1 0.95 0.41 0.46 0.41 0.46 0.41 0.46 0.41 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41</td></t<></td></t<>	1 0.98 0.77 0.7 0.48 0.37 0.43 0.98 1 0.73 0.62 0.35 0.23 0.3 0.77 0.73 1 0.71 0.53 0.47 0.46 0.77 0.73 1 0.71 0.53 0.47 0.46 0.7 0.62 0.71 1 0.73 0.46 0.67 0.46 0.7 0.62 0.71 1 0.77 0.65 0.66 0.48 0.35 0.53 0.7 1 0.97 0.96 0.37 0.23 0.47 0.65 0.97 1 0.95 0.37 0.23 0.47 0.65 0.97 1 0.96 0.43 0.3 0.46 0.66 0.96 0.91 0.16 0.50 0.67 0.26 0.21 0.06 0.01 0.01 0.02 0.01 0.02 0.01 0.21 0.02 0.02 0.11 0.24 0.14 0.42 0.24 0.24 0.24 0.24 <t< td=""><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.77 0.62 0.71 1 0.77 0.46 0.37 0.48 0.35 0.71 1 0.77 0.46 0.37 0.48 0.35 0.71 1 0.77 0.46 0.37 0.48 0.35 0.53 0.7 1 0.97 0.99 0.48 0.37 0.23 0.47 0.65 0.97 1 0.99 0.41 0.43 0.33 0.46 0.66 0.96 0.97 1 0.99 0.41 0.43 0.31 0.46 0.46 0.96 0.91 0.41 0.41 0.43 0.47 0.43 0.47 0.43 0.41 0.41 0.41 0.43 0.47 0.43 0.43 0.43 0.42</td><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.59 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.67 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.26 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.26 0.77 0.62 0.71 1 0.77 0.66 0.37 0.26 0.77 0.62 0.71 1 0.77 0.66 0.37 0.26 0.44 0.35 0.77 1 0.97 0.96 0.48 0.06 0.43 0.33 0.47 0.66 0.97 1 0.95 0.41 0.48 0.43 0.33 0.46 0.96 0.95 1 0.44 0.14 0.53 0.67 0.62 0.12 0.06 0.14 0.41 0.4 0.14 0.54 0.56 0.22 0.12 0.06 0.14 0.44</td><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.59 0.57 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.67 0.65 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 -0.22 0.70 0.62 0.71 1 0.77 0.65 0.66 0.37 -0.21 -0.12 0.48 0.35 0.53 0.7 1 0.97 0.96 0.48 -0.22 -0.22 0.48 0.35 0.53 0.7 1 0.97 0.96 0.48 -0.21 -0.12 0.43 0.33 0.47 0.65 0.97 1 0.95 0.48 -0.21 -0.02 0.43 0.33 0.46 0.99 0.95 1 0.49 0.11 -0.49 0.43 0.41 0.18 0.47 0.48 0.41 0.41 0.41 0.41 0.59 0.67 0.22 0.12 0.01 0.14 0.</td><td>1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 -0.59 -0.73 -0.73 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 -0.67 -0.65 -0.79 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 -0.22 -0.43 0.77 0.62 0.71 1 0.77 0.65 0.66 0.37 -0.21 -0.12 -0.33 0.48 0.35 0.53 0.7 1 0.77 0.66 0.37 -0.21 -0.12 -0.33 0.48 0.35 0.53 0.7 1 0.77 0.46 0.48 -0.06 -0.12 -0.12 -0.33 0.43 0.23 0.47 0.65 0.47 0.48 0.41 0.48 -0.41 -0.49 -0.41 -0.43 -0.41 0.45 -0.41 -0.41 -0.41 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 <</td><td> 1. 0.98 0.77 0.70 0.48 0.37 0.43 0.44 0.45 0.45 0.45 0.46 0.44 0.47 0.45 0.47 0.43 0.44 0.47 0.45 0.47 0.43 0.44 0.44 0.44 0.44 0.45 0.45 0.45 0.46 0.47 0.48 0.41 0.44 0.45 0.45 0.45 0.44 0.45 0.45 0.46 0.47 0.48 0.49 0.49 0.41 0.49 0.41 0.44 0.4</td><td>1 0.98 0.77 0.48 0.37 0.43 0.21 0.59 0.73 0.43 0.49 0.98 1 0.73 0.62 0.62 0.35 0.23 0.43 0.14 0.67 0.65 0.79 0.73 0.43 0.47 0.48 0.18 0.26 0.22 0.43 0.17 0.52 0.77 0.53 0.17 0.53 0.47 0.46 0.18 0.26 0.22 0.43 0.17 0.53 0.70 0.62 0.71 1 0.77 0.65 0.66 0.37 0.21 0.12 0.33 0.04 0.33 0.48 0.35 0.53 0.77 1 0.95 0.48 0.06 0.10 0.12 0.33 0.04 0.33 0.43 0.35 0.47 0.46 0.97 1 0.95 0.41 0.46 0.41 0.46 0.41 0.46 0.41 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41</td></t<>	1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.77 0.62 0.71 1 0.77 0.46 0.37 0.48 0.35 0.71 1 0.77 0.46 0.37 0.48 0.35 0.71 1 0.77 0.46 0.37 0.48 0.35 0.53 0.7 1 0.97 0.99 0.48 0.37 0.23 0.47 0.65 0.97 1 0.99 0.41 0.43 0.33 0.46 0.66 0.96 0.97 1 0.99 0.41 0.43 0.31 0.46 0.46 0.96 0.91 0.41 0.41 0.43 0.47 0.43 0.47 0.43 0.41 0.41 0.41 0.43 0.47 0.43 0.43 0.43 0.42	1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.59 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.67 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.26 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 0.26 0.77 0.62 0.71 1 0.77 0.66 0.37 0.26 0.77 0.62 0.71 1 0.77 0.66 0.37 0.26 0.44 0.35 0.77 1 0.97 0.96 0.48 0.06 0.43 0.33 0.47 0.66 0.97 1 0.95 0.41 0.48 0.43 0.33 0.46 0.96 0.95 1 0.44 0.14 0.53 0.67 0.62 0.12 0.06 0.14 0.41 0.4 0.14 0.54 0.56 0.22 0.12 0.06 0.14 0.44	1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 0.59 0.57 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 0.67 0.65 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 -0.22 0.70 0.62 0.71 1 0.77 0.65 0.66 0.37 -0.21 -0.12 0.48 0.35 0.53 0.7 1 0.97 0.96 0.48 -0.22 -0.22 0.48 0.35 0.53 0.7 1 0.97 0.96 0.48 -0.21 -0.12 0.43 0.33 0.47 0.65 0.97 1 0.95 0.48 -0.21 -0.02 0.43 0.33 0.46 0.99 0.95 1 0.49 0.11 -0.49 0.43 0.41 0.18 0.47 0.48 0.41 0.41 0.41 0.41 0.59 0.67 0.22 0.12 0.01 0.14 0.	1 0.98 0.77 0.7 0.48 0.37 0.43 0.21 -0.59 -0.73 -0.73 0.98 1 0.73 0.62 0.35 0.23 0.3 0.14 -0.67 -0.65 -0.79 0.77 0.73 1 0.71 0.53 0.47 0.46 0.18 -0.22 -0.43 0.77 0.62 0.71 1 0.77 0.65 0.66 0.37 -0.21 -0.12 -0.33 0.48 0.35 0.53 0.7 1 0.77 0.66 0.37 -0.21 -0.12 -0.33 0.48 0.35 0.53 0.7 1 0.77 0.46 0.48 -0.06 -0.12 -0.12 -0.33 0.43 0.23 0.47 0.65 0.47 0.48 0.41 0.48 -0.41 -0.49 -0.41 -0.43 -0.41 0.45 -0.41 -0.41 -0.41 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 0.44 <	 1. 0.98 0.77 0.70 0.48 0.37 0.43 0.44 0.45 0.45 0.45 0.46 0.44 0.47 0.45 0.47 0.43 0.44 0.47 0.45 0.47 0.43 0.44 0.44 0.44 0.44 0.45 0.45 0.45 0.46 0.47 0.48 0.41 0.44 0.45 0.45 0.45 0.44 0.45 0.45 0.46 0.47 0.48 0.49 0.49 0.41 0.49 0.41 0.44 0.4	1 0.98 0.77 0.48 0.37 0.43 0.21 0.59 0.73 0.43 0.49 0.98 1 0.73 0.62 0.62 0.35 0.23 0.43 0.14 0.67 0.65 0.79 0.73 0.43 0.47 0.48 0.18 0.26 0.22 0.43 0.17 0.52 0.77 0.53 0.17 0.53 0.47 0.46 0.18 0.26 0.22 0.43 0.17 0.53 0.70 0.62 0.71 1 0.77 0.65 0.66 0.37 0.21 0.12 0.33 0.04 0.33 0.48 0.35 0.53 0.77 1 0.95 0.48 0.06 0.10 0.12 0.33 0.04 0.33 0.43 0.35 0.47 0.46 0.97 1 0.95 0.41 0.46 0.41 0.46 0.41 0.46 0.41 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41 0.45 0.41







Comparison of all tasks, pairwise. Colors correspond to different value of the negative sampling exponent hyperparameter.





Negative sampling drives much of performance