Anna Korhonen

List of Publications by Year in Descending Order

55 papers	**1,182** citations	**17** h-index	**32** g-index
56 ext. papers	**1,500** ext. citations	**4.1** avg, IF	**4.77** L-index

#	Paper	IF	Citations
55	Subcategorization frame identification for learner English. *International Journal of Corpus Linguistics*, **2021**, 26, 187-218	0.8	
54	Semantic Data Set Construction from Human Clustering and Spatial Arrangement. *Computational Linguistics*, **2021**, 47, 69-116	2.8	1
53	Application of Text Mining in Risk Assessment of Chemical Mixtures: A Case Study of Polycyclic Aromatic Hydrocarbons (PAHs). *Environmental Health Perspectives*, **2021**, 129, 67008	8.4	2
52	Multi-SimLex: A Large-Scale Evaluation of Multilingual and Crosslingual Lexical Semantic Similarity. *Computational Linguistics*, **2021**, 46, 847-897	2.8	4
51	Parameter Space Factorization for Zero-Shot Learning across Tasks and Languages. *Transactions of the Association for Computational Linguistics*, **2021**, 9, 410-428	5.6	1
50	A systematic literature review of automatic Alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association: JAMIA*, **2020**, 27, 1784-1797	8.6	23
49	Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing. *Computational Linguistics*, **2019**, 45, 559-601	2.8	13
48	LION LBD: a literature-based discovery system for cancer biology. *Bioinformatics*, **2019**, 35, 1553-1561	7.2	23
47	Investigating the cross-lingual translatability of VerbNet-style classification. *Language Resources and Evaluation*, **2018**, 52, 771-799	1.8	4
46	Bio-SimVerb and Bio-SimLex: wide-coverage evaluation sets of word similarity in biomedicine. *BMC Bioinformatics*, **2018**, 19, 33	3.6	11
45	Dependency parsing of learner English. *International Journal of Corpus Linguistics*, **2018**, 23, 28-54	0.8	15
44	Adversarial Propagation and Zero-Shot Cross-Lingual Transfer of Word Vector Specialization **2018**,		8
43	On the Relation between Linguistic Typology and (Limitations of) Multilingual Language Modeling **2018**,		8
42	Isomorphic Transfer of Syntactic Structures in Cross-Lingual NLP **2018**,		6
41	Language Modeling for Morphologically Rich Languages: Character-Aware Modeling for Word-Level Prediction. *Transactions of the Association for Computational Linguistics*, **2018**, 6, 451-465	5.6	10
40	Link prediction in drug-target interactions network using similarity indices. *BMC Bioinformatics*, **2017**, 18, 39	3.6	69
39	Cancer Hallmarks Analytics Tool (CHAT): a text mining approach to organize and evaluate scientific literature on cancer. *Bioinformatics*, **2017**, 33, 3973-3981	7.2	63

38	Text mining for improved exposure assessment. *PLoS ONE*, **2017**, 12, e0173132	3.7	8
37	HyperLex: A Large-Scale Evaluation of Graded Lexical Entailment. *Computational Linguistics*, **2017**, 43, 781-835	2.8	14
36	Semantic Specialization of Distributional Word Vector Spaces using Monolingual and Cross-Lingual Constraints. *Transactions of the Association for Computational Linguistics*, **2017**, 5, 309-324	5.6	29
35	Morph-fitting: Fine-Tuning Word Vector Spaces with Simple Language-Specific Rules **2017**,		7
34	Decoding Sentiment from Distributed Representations of Sentences **2017**,		3
33	Automatic semantic classification of scientific literature according to the hallmarks of cancer. *Bioinformatics*, **2016**, 32, 432-40	7.2	27
32	Grouping chemicals for health risk assessment: A text mining-based case study of polychlorinated biphenyls (PCBs). *Toxicology Letters*, **2016**, 241, 32-7	4.4	9
31	SimVerb-3500: A Large-Scale Evaluation Set of Verb Similarity **2016**,		47
30	Anchoring and Agreement in Syntactic Annotations **2016**,		6
29	Is "Universal Syntax" Universally Useful for Learning Distributed Word Representations? **2016**,		2
28	Learning to Understand Phrases by Embedding the Dictionary. *Transactions of the Association for Computational Linguistics*, **2016**, 4, 17-30	5.6	25
27	Gender differences in cancer susceptibility: role of oxidative stress. *Carcinogenesis*, **2016**, 37, 985-992	4.6	15
26	Unsupervised discovery of information structure in biomedical documents. *Bioinformatics*, **2015**, 31, 1084-92	7.2	3
25	SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation. *Computational Linguistics*, **2015**, 41, 665-695	2.8	259
24	Exploring big educational learner corpora for SLA research. *International Journal of Learner Corpus Research*, **2015**, 1, 96-129	0.6	15
23	Automatic extraction of property norm-like data from large text corpora. *Cognitive Science*, **2014**, 38, 638-82	2.2	3
22	Multi-Modal Models for Concrete and Abstract Concept Meaning. *Transactions of the Association for Computational Linguistics*, **2014**, 2, 285-296	5.6	8
21	Evaluation of carcinogenic modes of action for pesticides in fruit on the Swedish market using a text-mining tool. *Frontiers in Pharmacology*, **2014**, 5, 145	5.6	7

#	Reference	IF	Cites
20	A quantitative empirical analysis of the abstract/concrete distinction. *Cognitive Science*, **2014**, 38, 162-77	2.2	11
19	Probabilistic Distributional Semantics with Latent Variable Models. *Computational Linguistics*, **2014**, 40, 587-631	2.8	6
18	Improving Multi-Modal Representations Using Image Dispersion: Why Less is Sometimes More **2014**,		11
17	Conceptual metaphor theory meets the data: a corpus-based human annotation study. *Language Resources and Evaluation*, **2013**, 47, 1261-1284	1.8	8
16	Active learning-based information structure analysis of full scientific articles and two applications for biomedical literature review. *Bioinformatics*, **2013**, 29, 1440-7	7.2	9
15	Statistical Metaphor Processing. *Computational Linguistics*, **2013**, 39, 301-353	2.8	41
14	Text mining for literature review and knowledge discovery in cancer risk assessment and research. *PLoS ONE*, **2012**, 7, e33427	3.7	45
13	Exocrine pancreatic carcinogenesis and autotaxin expression. *PLoS ONE*, **2012**, 7, e43209	3.7	12
12	A comparison and user-based evaluation of models of textual information structure in the context of cancer risk assessment. *BMC Bioinformatics*, **2011**, 12, 69	3.6	12
11	Weakly supervised learning of information structure of scientific abstracts--is it accurate enough to benefit real-world tasks in biomedicine?. *Bioinformatics*, **2011**, 27, 3179-85	7.2	9
10	Automatic lexical classification: bridging research and practice. *Philosophical Transactions Series A, Mathematical, Physical, and Engineering Sciences*, **2010**, 368, 3621-32	3	2
9	The first step in the development of Text Mining technology for Cancer Risk Assessment: identifying and organizing scientific evidence in risk assessment literature. *BMC Bioinformatics*, **2009**, 10, 303	3.6	17
8	Towards Unrestricted, Large-Scale Acquisition of Feature-Based Conceptual Representations from Corpus Data. *Research on Language and Computation*, **2009**, 7, 137-170		11
7	Improving verb clustering with automatically acquired selectional preferences **2009**,		19
6	Unsupervised and constrained Dirichlet process mixture models for verb clustering **2009**,		21
5	A large-scale classification of English verbs. *Computers and the Humanities*, **2008**, 42, 21-40		106
4	Verb Class Discovery from Rich Syntactic Data **2008**, 16-27		9
3	Zone analysis in biology articles as a basis for information extraction. *International Journal of Medical Informatics*, **2006**, 75, 468-87	5.3	49

List of Publications