

CITATION REPORT

List of articles citing

Introduction to the Special Issue on the Web as Corpus

DOI: 10.1162/089120103322711569

Computational Linguistics, 2003, 29, 333-347.

Source: <https://exaly.com/paper-pdf/35402556/citation-report.pdf>

Version: 2024-04-26

This report has been generated based on the citations recorded by exaly.com for the above article. For the latest version of this publication list, visit the link given above.

The third column is the impact factor (IF) of the journal, and the fourth column is the number of citations of the article.

#	Paper	IF	Citations
331	Webaffix: une boîte à outils d'acquisition lexicale à partir du Web. 2003 , 32, 61-84		
330	Creating multilingual translation lexicons with regional variations using web corpora. 2004 ,		6
329	Translating unknown queries with web corpora for cross-language information retrieval. 2004 ,		34
328	Creating and using Web corpora. 2005 , 10, 517-541		6
327	Automatic training corpora acquisition through Web mining.		3
326	Time to get wired: Using web-based corpora in critical discourse analysis. 2005 , 16, 809-828		91
325	Language is never, ever, ever, random. 2005 , 1,		86
324	A Web-based unsupervised algorithm for learning transliteration model to improve translation of low-frequency proper names.		3
323	Using the Web corpus to translate the queries in cross-lingual information retrieval.		
322	Web-based models for natural language processing. 2005 , 2, 3		56
321	Alignment of bilingual named entities in parallel corpora using statistical models and multiple knowledge sources. 2006 , 5, 121-145		16
320	A Framework for Web Science. 2006 , 1, 1-130		169
319	Bilingual Lexicography. 2006 , 787-796		
318	Computers in Lexicography. 2006 , 783-793		
317	Describing verb-complementational profiles of New Englishes. 2006 , 27, 147-173		53
316	2020 computing: the creativity machine. 2006 , 440, 411		10
315	Familiarity and relational preference in the understanding of noun--noun compounds. 2006 , 34, 1285-97		9

314	Open-source Corpora. 2006 , 11, 435-462		56
313	Collocations in Indian English: A Corpus-Based Sample Analysis. 2006 , 124,		3
312	Development of a New Corpus System for English Writing Support and its Fundamental Study. 2006 ,		
311	. 2006 ,		
310	Natural Language Processing and Information Systems. <i>Lecture Notes in Computer Science</i> , 2006 ,	0.9	
309	Web resources for language modeling in conversational speech recognition. 2007 , 5, 1-25		21
308	Geo-tagging for imprecise regions of different sizes. 2007 ,		17
307	An Intelligent Web Agent to Mine Bilingual Parallel Pages via Automatic Discovery of URL Pairing Patterns. 2007 ,		2
306	Using the Web to Reduce Data Sparseness in Pattern-Based Information Extraction. <i>Lecture Notes in Computer Science</i> , 2007 , 18-29	0.9	11
305	Construction Grammar in the twenty-first century. 2007 , 11, 569-585		8
304	Corpus Linguistics and the Web. 2007 ,		80
303	Automatic Acquisition of Lexical Information and Examples. 2007 , 253-274		1
302	Choosing the content of textual summaries of large time-series data sets. 2007 , 13, 25-49		48
301	Evaluation of Text and Speech Systems. 2007 ,		13
300	Correct your text with Google. 2007 ,		4
299	Automatic Correction of Idiomatic Usage in English Using Web Search. 2007 ,		
298	A statistical model for near-synonym choice. 2007 , 4, 1-17		31
297	An Ontology-Based Exploration of Knowledge Systems for Metaphor. 2007 , 489-517		4

296	Googling for opposites—a web-based study of antonym canonicity. 2007 , 2, 129-155	30
295	Efficient corpus development for lexicography: building the New Corpus for Ireland. 2007 , 40, 127-152	6
294	A web-based Bengali news corpus for named entity recognition. 2008 , 42, 173-182	30
293	Asian language resources: the state-of-the-art. 2008 , 42, 109-116	1
292	A method for words combinability estimation based on internet texts. 2008 , 32, 234-238	
291	Enhancing portability with multilingual ontology-based knowledge management. 2008 , 45, 567-584	18
290	System Search for learning. 2008 ,	
289	Testing search engine frequencies: Patterns of inconsistency. 2008 , 4,	5
288	A Corpus-based evaluation of lexical components of a domain-specific text to Knowledge Mapping prototype. 2008 ,	2
287	Modelling vague places with knowledge from the Web. 2008 , 22, 1045-1065	249
286	Seeing as though1. 2008 , 12, 103-139	7
285	Linguistics Tools. 2008 ,	
284	Mapping geographic coverage of the web. 2008 ,	5
283	Search advertising using web relevance feedback. 2008 ,	57
282	Information Extraction, Search, Interaction and Collaboration on the Web in Mexico. 2008 ,	
281	Word sense disambiguation. 2008 , 3, 4358	4
280	Identification of Complex Named-Entities in Chinese Queries Using WWW. 2008 ,	
279	Unsupervised Translation Disambiguation Based on Maximum Web Bilingual Relatedness: Web as Lexicon. 2009 ,	

278	Word sense disambiguation. 2009 , 41, 1-69	689
277	Towards Executable Acceptable Use Policies (execAUPs) for email clouds. 2009 ,	1
276	The BE06 Corpus of British English and recent language change. 2009 , 14, 312-337	68
275	Category invasion and Web data. 2009 , 31, 447-476	
274	Using the Web as corpus for self-training text categorization. 2009 , 12, 400-415	9
273	The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. 2009 , 43, 209-226	353
272	Refining the use of the web (and web search) as a language teaching and learning resource. 2009 , 22, 249-268	17
271	. 2009 ,	6
270	Exploring English Phraseology with Two Tools: NSM Semantic Methodology and Google. 2009 , 37, 101-129	11
269	Utilizing the Web for Automatic Word Spacing. 2009 , E92-D, 2553-2556	1
268	An English and/or Japanese writing support tool based on a web search engine. 2010 , 38, 324	4
267	References. 2010 ,	
266	The Effect of Corpus Size on Case Frame Acquisition for Predicate-Argument Structure Analysis. 2010 , E93-D, 1361-1368	1
265	Automatic word sense disambiguation based on document networks. 2010 , 36, 11-18	2
264	Word sense disambiguation methods. 2010 , 36, 309-326	9
263	Challenges to issues of balance and representativeness in African lexicography. 2010 , 16,	
262	Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. 2010 , 22, 83-102	21
261	Der Einfluss der Dependenzgrammatik auf die Computerlinguistik. 2010 , 38,	

260	Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. 2010 , 23, 377-393	22
259	Trends in Multiple Criteria Decision Analysis. 2010 ,	49
258	What helps where and why? Semantic relatedness for knowledge transfer. 2010 ,	162
257	Mining Parallel Data from Comparable Corpora via Triangulation. 2011 ,	1
256	Metaphor Aptness and Conventionality: A Processing Fluency Account. 2011 , 26, 206-226	52
255	DOCUMENT TRANSLATION RETRIEVAL BASED ON STATISTICAL MACHINE TRANSLATION TECHNIQUES. 2011 , 25, 329-340	3
254	Babouk: Focused Web Crawling for Corpus Compilation and Automatic Terminology Extraction. 2011 ,	10
253	References. 259-291	
252	La lexicographie et l'analyse de corpus : nouvelles perspectives. 2011 , 56, 247-265	
251	Resources for Turkish morphological processing. 2011 , 45, 249-261	28
250	Constructing specialised corpora through analysing domain representativeness of websites. 2011 , 45, 209-241	5
249	Automatic extraction of acronym definitions from the Web. 2011 , 34, 311-327	32
248	Linguistic evidence and grammatical theory. 2011 , 2, 206-221	16
247	Translation students' use and evaluation of online resources for Chinese-English translation at the word level. 2011 , 6, 62-86	8
246	A Validation Study of the CEFR Levels of Phrasal Verbs in the English Vocabulary Profile. 2012 , 3,	2
245	Corpora and Language Education. 2012 ,	54
244	109. Semantics in corpus linguistics.	
243	References.	

242	N-grams corpus generation from inverted index for query refinement in information retrieval applications. 2012,	
241	Use of Google Scholar in corpus-driven EAP research. 2012, 11, 319-331	10
240	Corpus Analysis of the World Wide Web. 2012,	12
239	Heterogeneity in Word-Formation Patterns: A Corpus-Based Analysis of Suffixation with -ee and Its Productivity in English. 2012, 29, 132-143	
238	Natural language interaction with the web of data by mining its textual side. 2012, 6, 121-133	0
237	Web query disambiguation using PageRank. 2012, 63, 1581-1592	8
236	Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. 2013, 14, 10	46
235	Automatic building of Arabic multi dialect text corpora by bootstrapping dialect words. 2013,	9
234	Web comment-based trend analysis on deep geothermal energy. 2013,	6
233	Morphological query expansion and language-filtering words for improving Basque web retrieval. 2013, 47, 425-448	6
232	A brief survey on corpus uses. 2013,	
231	Instability in Search Engine Results: Lessons Learned in the Context of Horizon Scanning Applications. 2013,	1
230	Collaboratively built semi-structured content and Artificial Intelligence: The story so far. 2013, 194, 2-27	81
229	Phraseology and frequency of occurrence on the web: native speakers' perceptions of Google-informed second language writing. 2013, 26, 144-157	21
228	The People's Web Meets NLP. 2013,	3
227	Likelihood calculation classification for Indonesian language news documents. 2013,	1
226	The Knowledge-Remixing Bottleneck. 2013, 28, 44-48	2
225	Using the Web to Create Dynamic Dictionaries in Handwritten Out-of-Vocabulary Word Recognition. 2013,	4

224	Web Corpus Construction. 2013 , 6, 1-145	36
223	99. The World Wide Web as a resource for lexicography. 2013 , 1365-1374	
222	An Emergent Approach to Text Analysis Based on a Connectionist Model and the Web. 2013 , 6, 565-590	
221	Assessing web search engines: a webometric approach. 135-146	
220	References. 274-286	
219	Ne jetons pas le Wiktionnaire avec l'oripeau du Web ! Etudes et réalisations fondées sur le dictionnaire collaboratif. 2014 , 8, 663-680	1
218	Community, tools, and practices in web archiving: The state-of-the-art in relation to social science and humanities research needs. 2014 , 65, 2195-2209	14
217	Roland Schfer & Felix Bildhauer, Web Corpus Construction (Synthesis Lectures on Human Language Technologies 22). Morgan & Claypool, 2013. Pp. xv + 129.. 2014 , 37, 457-463	
216	Review of Flowerdew (2012): Corpora and Language Education. 2014 , 19, 147-155	
215	Miros'awa Podhajecka. 2013. Russian Borrowings in English: A Dictionary and Corpus Study.. 2014 , 27, 319-324	1
214	Physiological Mandarin Speech Corpus for the International Preparatory Students in China: Design and Initial Application. 2014 , 989-994, 5336-5340	
213	Exploring and Visualizing Differences in Geographic and Linguistic Web Coverage. 2014 , 18, 852-876	1
212	Researching Language and Social Media. 2014 ,	70
211	Considering a Wider Web?. 2015 ,	
210	Bibliography. 629-724	
209	Machine translation from Japanese and French to Vietnamese, the difference among language families. 2015 ,	0
208	Corpus Annotation. 2015 ,	0
207	Corpus Linguistics in Trademark Cases. 2015 , 36, 100-114	1

- 206 In Search of Oblivion? How the Right to be Forgotten Could Undermine Web-based Corpora. **2015**, 198, 95-102
- 205 55. Syntax and Corpora.
- 204 Domain adaptation of statistical machine translation with domain-focused web crawling. **2015**, 49, 147-193 5
- 203 [Big data, medical language and biomedical terminology systems]. **2015**, 58, 844-852
- 202 The grammatical features of English in a Chinese Internet discussion forum. **2015**, 34, 211-230 8
- 201 Processing Conventional Conceptual Metaphors in Persian: A Corpus-Based Psycholinguistic Study. **2015**, 44, 495-518 2
- 200 Enrichir le balisage de corpus footballistiques pour en augmenter le pouvoir documentaire. **2016**, 27, 11004
- 199 Corpora. **2016**,
- 198 A Novel Approach to Managing the Dynamic Nature of Semantic Relatedness. **2016**, 27, 1-26 2
- 197 Constructing a Collocation Learning System from the Wikipedia Corpus. **2016**, 6, 18-35 7
- 196 Bibliography. 497-602
- 195 Frequency and Corpora. 685-709 4
- 194 Big Data in Cognitive Science. **2016**, 9
- 193 Inverse local context analysis. **2016**, 34, 405-418
- 192 Towards a user-based taxonomy of web registers. **2016**, 19-42
- 191 Register Variation on the Searchable Web: A Multi-Dimensional Analysis. **2016**, 44, 95-137 63
- 190 Adaptive memory: Animacy enhances free recall but impairs cued recall. **2016**, 42, 186-201 33
- 189 Theoretical Background. **2016**, 23-62

188	A Template Independent Approach for Web News and Blog Content Extraction. 2016,	0
187	Twitter corpus creation: The case of a Malay Chat-style-text Corpus (MCC). 2016, 31, 227-243	5
186	Cognitive Mechanisms Underlying Fake Reflexive Resultatives** We would like to thank two anonymous reviewers for their feedback. This work was supported by the Spanish Ministry of Economy and Competitiveness [Grants numbers FFI2011-29798-C02-01 and FFI2013-43593-P].View all notes. 2016, 36, 502-541	5
185	Introduction: Using Multilingual Written Internet Data in Code-Switching and Language Contact Research. 2016, 9, 5-22	4
184	Estimating search engine index size variability: a 9-year longitudinal study. 2016, 107, 839-856	31
183	Nenek: a cloud-based collaboration platform for the management of Amerindian language resources. 2017, 51, 897-925	0
182	Corpus Use in Language Learning: A Meta-Analysis. 2017, 67, 348-393	137
181	Crawl and crowd to bring machine translation to under-resourced languages. 2017, 51, 1019-1051	4
180	Effects of context and individual differences on the processing of taboo words. 2017, 178, 73-86	5
179	Evaluating a 12-Million-Word Corpus as a Source of Dictionary Data. 2017,	1
178	Patterns of Google use in language reference and learning: a user survey. 2017, 4, 419-439	0
177	MC4WEPS: a multilingual corpus for Web people search disambiguation. 2017, 51, 805-832	2
176	Literaturverzeichnis. 2017,	
175	Adam Kilgarriff Legacy to Computational Linguistics and Beyond. <i>Lecture Notes in Computer Science</i> , 2018, 3-25	0.9 7
174	Big Data: Deep Learning for financial sentiment analysis. 2018, 5,	108
173	Corpus Pragmatics and Second Language Pragmatics: A Mutualistic Entente in Theory and Practice. 2018, 2, 113-127	7
172	The animacy advantage for free-recall performance is not attributable to greater mental arousal. 2018, 26, 89-95	29
171	Testing a methodological framework for retrieving parallel texts in the domain of business translation. 2018, 26, 39-53	2

170	Towards Accurate Detection of Offensive Language in Online Communication in Arabic. 2018 , 142, 315-320	25
169	Exaggerating and mitigating through metonymy: The case of situational and CAUSE FOR EFFECT/EFFECT FOR CAUSE metonymies. 2018 , 62, 51-65	6
168	Introduction. 1-11	
167	Corpus and Methods. 12-25	
166	A Survey of the Registers on the Publicly Searchable Web. 26-43	
165	Overall Patterns of Register Variation on the Searchable Web: A Multidimensional Analysis. 44-72	
164	Narrative Registers. 73-105	
163	Opinion, Advice, and Persuasion Registers. 106-134	
162	Informational Descriptions, Explanations, and Procedures. 135-176	
161	Oral Registers. 177-195	
160	The Web as a Continuous Space of Register Variation. 196-216	
159	Appendix A: Linguistic Features Included in the Multidimensional Analysis. 217-218	0
158	Appendix B: Lexico-grammatical Features Included in the Key Feature Analysis. 219-220	
157	Appendix C: Descriptive Statistics for the Key Feature Analyses. 221-242	
156	Index. 247-252	
155	Framing Big Data. 2019 ,	5
154	CLAD: A corpus-derived Chinese Lexical Association Database. 2019 , 51, 2310-2336	6
153	Introduction. 2019 , 1-16	

152 Corpus and Methodology. **2019**, 17-45

151 Delimiting the Scope of the Study. **2019**, 46-106

150 The Relationship between Phrasal Verbs and the Processes of Grammaticalisation, Lexicalisation, and Idiomatisation. **2019**, 107-152

149 Phrasal Verbs 1650-1990. **2019**, 153-214

148 Phrasal Verbs 1650-1990. **2019**, 215-277

147 Conclusion. **2019**, 278-286

146 List of Phrasal Verb Constructions with a Fixed Syntactic Order. **2019**, 287-288

145 Raw and Normalised Frequencies of Phrasal Verb Particles in ARCHER and the OBC. **2019**, 289-289

144 List of the Twenty Most Common Combinations Distributed across Genres. **2019**, 290-291

143 Raw Frequencies of Particles Distributed across Genres. **2019**, 292-292

142 Bibliography. **2019**, 293-318

141 Index. **2019**, 319-322

140 Characterising postgraduate students' corpus query and usage patterns for disciplinary data-driven learning. **2019**, 31, 255-275 13

139 From a Smoking Gun to Spent Fuel: Principled Subsampling Methods for Building Big Language Data Corpora from Monitor Corpora. **2019**, 4, 48 0

138 Online corpus of spoken Ilokano language. **2019**, 482, 012034

137 Enhanced news sentiment analysis using deep learning methods. **2019**, 2, 33-46 30

136 14. Semantics in corpus linguistics. **2019**, 409-443

135 The Trade-off between Quantity and Quality. Comparing a Large Crawled Corpus and a Small Focused Corpus for Medical Terminology Extraction. **2019**, 20, 197-211

- 134 Digital, digitized, and numerical humanities. **2019**, 34, 616-632 2
- 133 Semantic association computation: a comprehensive survey. **2020**, 53, 3849-3899 1
- 132 Fostering intuitive competence in L2 for a better performance in EAP writing through fraze.it in a Turkish context. **2020**, 25, 5405-5426
- 131 From semantics to pragmatics: where IS can lead in Natural Language Processing (NLP) research. **2020**, 1-22 2
- 130 On Some Pragmatic Effects of Event Metonymies. **2020**, 35, 266-284
- 129 9. Language Corpora. **2020**, 421-434
- 128 El uso de la traducci3n autom3tica en las cuentas de Twitter de los principales peri3dicos de Latinoam3rica: una aproximaci3n. **2021**, 14, 167-194
- 127 Duration as Length Vs Amount in English and Spanish: A Corpus Study. **2021**, 36, 74-84 1
- 126 Empirical Research in Linguistics. **2021**, 3-42
- 125 Language Typology. **2021**, 106-132
- 124 References. **2021**, 281-308
- 123 Basic Research Methods for Data Collection. **2021**, 43-76
- 122 Specific Research Approaches of Linguistic Subdisciplines. **2021**, 77-260
- 121 Corpus Linguistics. **2021**, 133-165
- 120 Index. **2021**, 309-312
- 119 Introducing Linguistic Research. **2021**,
- 118 Sociolinguistics and Anthropological Linguistics. **2021**, 166-194
- 117 Preface. **2021**, xi-xvi

116 Neurolinguistics. **2021**, 229-260

115 Research Basics. **2021**, 1-76

114 Language Documentation and Descriptive Linguistics. **2021**, 79-105

113 Linguistic Research across the Discipline. **2021**, 261-280

112 Insights from Linguistic Research. **2021**, 263-280

111 Cognitive Linguistics and Psycholinguistics. **2021**, 195-228

110 Detection and Correction of Malapropisms in Spanish by Means of Internet Search. *Lecture Notes in Computer Science*, **2005**, 115-122 0.9 6

109 Using the Structure of a Conceptual Network in Computing Semantic Relatedness. *Lecture Notes in Computer Science*, **2005**, 767-778 0.9 27

108 Encyclopedia of Complexity and Systems Science. **2009**, 8084-8104 1

107 Argumentation Theory and Decision Aiding. **2010**, 177-208 21

106 Comparing Neural Lexical Models of a Classic National Corpus and a Web Corpus: The Case for Russian. *Lecture Notes in Computer Science*, **2015**, 47-58 0.9 4

105 Web as a Corpus: Going Beyond the n-gram. *Communications in Computer and Information Science*, **2015**, 185-228 0.3 1

104 Question Classification in Spanish and Portuguese. *Lecture Notes in Computer Science*, **2005**, 612-619 0.9 2

103 Constructing a Large Scale Text Corpus Based on the Grid and Trustworthiness. **2007**, 56-65 1

102 Web-Based Measure of Semantic Relatedness. *Lecture Notes in Computer Science*, **2008**, 136-150 0.9 57

101 Semi-supervised Word Sense Disambiguation Using the Web as Corpus. *Lecture Notes in Computer Science*, **2009**, 256-265 0.9 2

100 Chapter 2: Next Generation Web Search. *Lecture Notes in Computer Science*, **2010**, 11-23 0.9 29

99 Reliability Verification of Search Engines Hit Counts: How to Select a Reliable Hit Count for a Query. *Lecture Notes in Computer Science*, **2010**, 114-125 0.9 4

98	Linguistisches Text Mining [Neue Wege ff die Marktforschung. 2015 , 167-185	1
97	Hypertext Types and Markup Languages. 2010 , 143-164	1
96	The English Phrasal Verb, 1650Present: History, Stylistic Drifts, and Lexicalisation. 2019 ,	8
95	Register Variation Online. 2018 ,	54
94	Overthrowing Yesterday ICM: (Re)focusing of Meaning in a Hong Kong Chinese (Cantonese) Constructional Idiom. 2014 , 68-95	2
93	Using eye-tracking to examine the reading of texts containing taboo words. 2015 , 213-238	2
92	Is 𠄎 fish in a river equivalent to 𠄎 fish a river? 2018 , 5, 208-229	0
91	Cognitive Modeling. 2014 ,	67
90	Chapter 2. Construing and constructing hyperbole. 42-73	7
89	Contrastive Pragmatics and Translation. 2016 ,	17
88	Applying data-driven learning to the web. <i>Studies in Corpus Linguistics</i> , 2015 , 267-296	0.3 5
87	Quantifying lexical usage: vocabulary pertaining to ecosystems and the environment. 2013 , 8, 53-79	8
86	Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. 2015 , 10, 11-45	50
85	Making corpus data visible: visualising text with research intermediaries. 2017 , 12, 459-482	2
84	Unsupervised Translation Disambiguation Based on Web Indirect Association of Bilingual Word. 2010 , 21, 575-585	1
83	OntoWiktionary. 131-161	3
82	WebT-IDC: A Web Tool for Intelligent Dataset Creation A Use Case for Forums and Blogs. 2021 , 192, 1051-1060	1
81	Interpretations based on delayed-domain (dis)appearance in printed advertising. 2021 , 19, 299-331	

80	Linguistes et ?talistes' peuvent coopfer : rep?age et analyse des gloses. 2004 , IX, 63		2
79	An Experiment in Detection and Correction of Malapropisms Through the Web. <i>Lecture Notes in Computer Science</i> , 2005 , 803-815	0.9	2
78	Web-Assisted Detection and Correction of Joint and Disjoint Malapropos Word Combinations. <i>Lecture Notes in Computer Science</i> , 2005 , 126-137	0.9	1
77	Measurements of Lexico-Syntactic Cohesion by Means of Internet. <i>Lecture Notes in Computer Science</i> , 2005 , 790-799	0.9	4
76	Creating Ontologies for Content Representation The OntoSeed Suite. <i>Lecture Notes in Computer Science</i> , 2005 , 1296-1313	0.9	1
75	Automatic Construction of a Japanese Onomatopoeic Dictionary Using Text Data on the WWW. <i>Lecture Notes in Computer Science</i> , 2006 , 209-215	0.9	1
74	What is the Semantic Web and What Will it Do for eScience?.		1
73	Identifying the Multiple Contexts of a Situation. <i>Lecture Notes in Computer Science</i> , 2006 , 118-133	0.9	3
72	Automatic search of information about cancer in the Web. 2007 , 1157-1159		
71	Automatically Extracting Unknown Translations Using Phrase Alignment. 2007 , 14B, 231-240		
70	Que peuvent nous apprendre en syntaxe des corpus oraux « anciens »?. 2007 , 15-25		1
69	Concordancing 2.0. 2009 , 411-431		
68	Unsupervised Translation Disambiguation by Using Semantic Dictionary and Mining Language Model from Web. 2010 , 20, 1292-1300		0
67	Large Web Corpora for Indian Languages. <i>Communications in Computer and Information Science</i> , 2011 , 312-313	0.3	1
66	Computational Complexity. 2012 , 2810-2829		
65	Analyzing the Sense Distribution of Concordances Obtained by Web as Corpus Approach. <i>Lecture Notes in Computer Science</i> , 2013 , 355-367	0.9	
64	TWORPUS An Easy-to-Use Tool for the Creation of Tailored Twitter Corpora. <i>Lecture Notes in Computer Science</i> , 2013 , 23-34	0.9	0
63	Building a Linked Open Data Cloud of Linguistic Resources: Motivations and Developments. 2013 , 315-348		

- 62 Susanne Mlleisen, Heterogeneity in word-formation patterns. A corpus-based analysis of suffixation with -ee and its productivity in English. (= Studies in Language Companion Series 118). Amsterdam and Philadelphia: Benjamins, 2010. xiv + 245pp. ISBN 978-90-272-0585-8 (hardback).. **2013**, 6, 214-225
- 61 Chapter 2. Theoretical pre-requisites. **2014**, 17-57
- 60 Chapter 6. Conclusions. **2014**, 223-226
- 59 Name index. **2014**, 245-247
- 58 Chapter 4. Cognitive operations. **2014**, 85-145
- 57 Acknowledgements. **2014**, ix-ix
- 56 Chapter 3. Cognitive models. **2014**, 59-84
- 55 References. **2014**, 227-244
- 54 Subject index. **2014**, 249-250
- 53 Chapter 1. Introduction. **2014**, 1-15
- 52 Chapter 5. Content operations across levels of representation. **2014**, 147-221
- 51 A new approach to the creation of web-based core dictionaries. **2017**, 45-56
- 50 Developing Linguistic Theories Using Annotated Corpora. **2017**, 411-438 1
- 49 Changes in the Use of Neologisms Based on a Large-scale Web-crawled Corpus ? an Investigation of the Current State and Related Issues ?. **2017**, null, 72-106
- 48 Chapter 8. Automatic detection of syntactic patterns from texts with application to Spanish clitic doubling. 169-188
- 47 Development of a Brazilian Portuguese Hotel Reviews Corpus. *Lecture Notes in Computer Science*, **2018**, 353-361 0.9 0
- 46 Mining the Web for Collocations: IR Models of Term Associations. *Lecture Notes in Computer Science*, **2018**, 177-194 0.9
- 45 Towards an Investigation of Language Learners (Non)Use of Online Technology. **2018**, 19-44

26	INVESTIGATING ARABIC CORPUS (KorSA) OF INDONESIAN UNDERGRADUATE THESIS ABSTRACTS. <i>Humanities and Social Sciences Reviews</i> , 2020 , 8, 920-927	0.2	
25	The suffix -ee: history, productivity, frequency and violation of st. 2020 ,		
24	Exploration of a Balanced Reference Corpus with a Wide Variety of Text Mining Tools. 2020 ,		
23	Web Corpora. 2020 , 329-351		
22	Types of Computer Corpora. <i>Advances in Educational Technologies and Instructional Design Book Series</i> , 2020 , 1-21	0.3	
21	A Novel Approach to Managing the Dynamic Nature of Semantic Relatedness. 2020 , 1085-1114		
20	Improving Text Classification by Web Corpora. 2007 , 154-159		1
19	Creating Ontologies for Content Representation in the OntoSeed Suite. <i>Lecture Notes in Computer Science</i> , 2007 , 141-166	0.9	
18	Taking Advantage of the Web for Text Classification with Imbalanced Classes. 2007 , 831-838		
17	Various Criteria of Collocation Cohesion in Internet: Comparison of Resolving Power. 2008 , 64-72		2
16	Lexical and Semantic Resources for NLP: From Words to Meanings. <i>Lecture Notes in Computer Science</i> , 2008 , 277-284	0.9	1
15	Chapter 19. Orality on the searchable web. <i>Studies in Corpus Linguistics</i> , 2020 , 318-336	0.3	
14	Web crawling dla celów lingwistycznych. Wybrane aspekty gromadzenia i analizy danych tekstowych na przykładzie rosyjskojęzycznych newsów internetowych. <i>Prace Językoznawcze</i> , 2021 , 23, 87-104		2
13	The effect of editing techniques on machine translation-informed academic foreign language writing. <i>The EUROCALL Review</i> , 2021 , 29, 33-43	0.2	
12	Social changes through the lens of language: A big data study of Chinese modal verbs.. <i>PLoS ONE</i> , 2022 , 17, e0260210	3.7	0
11	Mining Japanese-Vietnamese multi-level parallel text corpus from Wikipedia data resource. 2021 ,		
10	Corpus linguistics for low-density varieties. Minority languages and corpus-based morphological investigations. <i>Corpus - Nice</i> , 2022 ,	0.4	
9	Deep Learning vs Compression-Based vs Traditional Machine Learning Classifiers to Detect Hadith Authenticity. <i>Communications in Computer and Information Science</i> , 2022 , 206-222	0.3	1

8	Modelling the lexical complexity of homogenous texts: a time series approach. <i>Quality and Quantity</i>	2.4
7	Understanding videos with face recognition: a complete pipeline and applications. <i>Multimedia Systems</i> ,	2.2
6	Web text corpus extraction system for linguistic tasks. 2009 , 29, 54-60	○
5	Identification of metadiscourse markers in bachelor's degree theses in Spanish.	○
4	The Academic Discourse of Mechanical Engineering. 2023 ,	○
3	Analyzing the unrestricted web: The finnish corpus of online registers. 1-31	○
2	Applying corpus-based learning to develop communicative competence among documentary linguistics students. 2023 , 61-76	○
1	Competing Views of Word Meaning: Word Embeddings and Word Senses.	○