

# CITATION REPORT

List of articles citing

## The Web as a Parallel Corpus

DOI: 10.1162/089120103322711578

Computational Linguistics, 2003, 29, 349-380.

**Source:** <https://exaly.com/paper-pdf/35928361/citation-report.pdf>

**Version:** 2024-04-20

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
239	Towards the self-annotating web. <b>2004</b> ,		152
238	Internet archive as a source of bilingual dictionary. <b>2004</b> ,		
237	Learning by googling. <b>2004</b> , 6, 24-33		83
236	Statistical machine translation with word- and sentence-aligned parallel corpora. <b>2004</b> ,		15
235	Automatic Identification of Parallel Documents With Light or Without Linguistic Resources. <i>Lecture Notes in Computer Science</i> , <b>2005</b> , 354-365	0.9	2
234	The linguist\$ search engine. <b>2005</b> ,		11
233	Text characteristics of English language university Web sites. <b>2005</b> , 56, 609-619		6
232	Mining key phrase translations from web corpora. <b>2005</b> ,		27
231	Exploiting parallel texts in the creation of multilingual semantically annotated resources: the MultiSemCor Corpus. <b>2005</b> , 11, 247		24
230	A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text. <b>2005</b> , 21, 259-274		43
229	Mining translations of OOV terms from the web through cross-lingual query expansion. <b>2005</b> ,		14
228	Creating and using Web corpora. <b>2005</b> , 10, 517-541		6
227	Improving Machine Translation Performance by Exploiting Non-Parallel Corpora. <i>Computational Linguistics</i> , <b>2005</b> , 31, 477-504	2.8	112
226	Advances in Artificial Intelligence. <i>Lecture Notes in Computer Science</i> , <b>2005</b> ,	0.9	
225	A model for ranking sentence pairs in parallel corpora. <b>2005</b> ,		
224	Multilingual single document keyword extraction for information retrieval.		22
223	.		0

222	Web-based models for natural language processing. <b>2005</b> , 2, 3	56
221	Alignment of bilingual named entities in parallel corpora using statistical models and multiple knowledge sources. <b>2006</b> , 5, 121-145	16
220	. <b>2006</b> ,	3
219	Advances in Information Retrieval. <i>Lecture Notes in Computer Science</i> , <b>2006</b> ,	0.9 12
218	Detecting the Same Text in Different Languages. <b>2006</b> ,	
217	. <b>2006</b> ,	15
216	Transcending the Tower of Babel: Supporting Access to Multilingual Information with Cross-Language Information Retrieval. <b>2006</b> , 299-314	2
215	Improved statistical machine translation using paraphrases. <b>2006</b> ,	51
214	Computers in Lexicography. <b>2006</b> , 783-793	
213	Stemming to improve translation lexicon creation from bitexts. <b>2006</b> , 42, 1003-1016	12
212	Novel association measures using web search with double checking. <b>2006</b> ,	46
211	Open-source Corpora. <b>2006</b> , 11, 435-462	56
210	A study of statistical models for query translation. <b>2006</b> ,	15
209	Detecting translations of the same text and data with common source. <b>2006</b> , 2006, P10009-P10009	5
208	Statistical query translation models for cross-language information retrieval. <b>2006</b> , 5, 323-359	10
207	Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead. <i>Lecture Notes in Computer Science</i> , <b>2006</b> ,	0.9 1
206	Named entity translation matching and learning. <b>2007</b> , 25, 2	6
205	An Intelligent Web Agent to Mine Bilingual Parallel Pages via Automatic Discovery of URL Pairing Patterns. <b>2007</b> ,	2

204	Measuring semantic similarity between words using web search engines. <b>2007</b> ,		53
203	Corpus Linguistics and the Web. <b>2007</b> ,		80
202	Automatic Acquisition of Lexical Information and Examples. <b>2007</b> , 253-274		1
201	Correct your text with Google. <b>2007</b> ,		4
200	Web-based parallel corpora for statistical machine translation. <b>2007</b> ,		
199	Automatic Extending HowNet's Attribute Lexicon on the Web. <b>2007</b> ,		
198	Automatic Construction of English-Vietnamese Parallel Corpus through Web Mining. <b>2007</b> ,		1
197	Sentence alignment using P-NNT and GMM. <b>2007</b> , 21, 594-608		4
196	Image estimation of words based on adjective co-occurrences. <b>2007</b> , 38, 14-24		
195	Automatic extraction of translations from web-based bilingual materials. <i>Machine Translation</i> , <b>2007</b> , 21, 139-163	1.1	
194	Pivot language approach for phrase-based statistical machine translation. <i>Machine Translation</i> , <b>2007</b> , 21, 165-181	1.1	47
193	Quantitative comparisons of search engine results. <b>2008</b> , 59, 1702-1710		53
192	Efficient data selection for machine translation. <b>2008</b> ,		5
191	Study on Web-Page Classification Algorithm Based on Rough Set Theory. <b>2008</b> ,		2
190	Natural Language Processing and the Web. <b>2008</b> , 23, 16-17		5
189	Integrating Cross-Language Hierarchies and Its Application to Retrieving Relevant Documents. <b>2008</b> , 7, 1-22		
188	Statistical machine translation. <b>2008</b> , 40, 1-49		92
187	Mapping geographic coverage of the web. <b>2008</b> ,		5

186	OOV Translation Mining from Mixed-Language Snippets from a Search Engine. <b>2008,</b>		
185	Word sense disambiguation. <b>2008, 3, 4358</b>		4
184	AEON ¶An approach to the automatic evaluation of ontologies. <b>2008, 3, 41-62</b>		21
183	The Acquisition and Sentence Alignment for Academic Bilingual Resources Based on Web Paper Libraries. <b>2009,</b>		
182	Unsupervised Translation Disambiguation Based on Maximum Web Bilingual Relatedness: Web as Lexicon. <b>2009,</b>		
181	Word sense disambiguation. <b>2009, 41, 1-69</b>		689
180	Improving the extraction of bilingual terminology from Wikipedia. <b>2009, 5, 1-17</b>		28
179	. <b>2009, 17, 13-23</b>		14
178	Translating medical terminologies through word alignment in parallel text corpora. <b>2009, 42, 692-701</b>		19
177	Web based translation of Chinese organization name. <b>2009, 26, 279-284</b>		
176	Advanced Information Retrieval. <b>2009, 225, 303-317</b>		27
175	Combining Content-Based and URL-Based Heuristics to Harvest Aligned Bitexts from Multilingual Sites with Bitextor. <i>Prague Bulletin of Mathematical Linguistics</i> , <b>2010, 93,</b>	0.3	8
174	A kernel regression framework for SMT. <i>Machine Translation</i> , <b>2010, 24, 87-102</b>	1.1	1
173	Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. <b>2010, 23, 377-393</b>		22
172	Cross-Language Information Retrieval. <i>Synthesis Lectures on Human Language Technologies</i> , <b>2010, 3, 1-1253</b>		39
171	Designing effective web mining-based techniques for OOV translation. <b>2010,</b>		
170	Extracting Parallel Texts from the Web. <b>2010,</b>		1
169	Building Bilingual Parallel Corpora Based on Wikipedia. <b>2010,</b>		17

168	Information Retrieval Technology. <i>Lecture Notes in Computer Science</i> , <b>2010</b> ,	0.9	2
167	Minimum Normalized Google Distance for Unsupervised Multilingual Chinese-English Word Sense Disambiguation. <b>2010</b> ,		
166	Study on Wikipedia for translation mining for CLIR. <b>2010</b> ,		
165	Data selection for statistical machine translation. <b>2010</b> ,		3
164	Approaches to improving corpus quality for statistical machine translation. <b>2010</b> ,		1
163	A method of mining bilingual resources from Web Based on Maximum Frequent Sequential Pattern. <b>2010</b> ,		
162	Automatic Construction of Web-Based English/Chinese Parallel Corpora. <b>2010</b> ,		0
161	Mining Parallel Data from Comparable Corpora via Triangulation. <b>2011</b> ,		1
160	DOCUMENT TRANSLATION RETRIEVAL BASED ON STATISTICAL MACHINE TRANSLATION TECHNIQUES. <b>2011</b> , 25, 329-340		3
159	Optimal distance metric function with trigram features for case based word sense disambiguation using artificial neural network. <b>2011</b> ,		0
158	. <b>2011</b> ,		4
157	Bitext Alignment. <i>Synthesis Lectures on Human Language Technologies</i> , <b>2011</b> , 4, 1-165	2.3	16
156	Ontologies and the Semantic Web. <b>2011</b> , 507-579		14
155	Information Retrieval Technology. <i>Lecture Notes in Computer Science</i> , <b>2011</b> ,	0.9	1
154	Machine Translation from Text. <b>2011</b> , 133-397		1
153	Parallel corpora and WordSpace models: using a third language as an interlingua to enrich multilingual resources. <b>2011</b> , 3, 299		
152	Towards a Specialized Search Engine for Language Learners [Point of View]. <b>2011</b> , 99, 1462-1465		3
151	Constructing specialised corpora through analysing domain representativeness of websites. <i>Language Resources and Evaluation</i> , <b>2011</b> , 45, 209-241	1.8	5

150	Exploring the sawa corpus: collection and deployment of a parallel corpus English↔wahili. <i>Language Resources and Evaluation</i> , <b>2011</b> , 45, 331-344	1.8	25
149	Parallel sentence generation from comparable corpora for improved SMT. <i>Machine Translation</i> , <b>2011</b> , 25, 341-375	1.1	20
148	Automatic extraction of acronym definitions from the Web. <b>2011</b> , 34, 311-327		32
147	Measuring semantic similarity between words by removing noise and redundancy in web snippets. <b>2011</b> , 23, 2496-2510		30
146	No free lunch. <b>2011</b> ,		19
145	Approaches to Improving Corpus Quality for Statistical Machine Translation. <b>2011</b> , 23, 327-348		2
144	Enabling users to create their own web-based machine translation engine. <b>2012</b> ,		1
143	Finding translations in scanned book collections. <b>2012</b> ,		2
142	Mining a large Chinese-English corpus from web. <b>2012</b> ,		
141	Exploring the further integration of machine translation in English-Chinese cross language information access. <b>2012</b> , 46, 429-457		7
140	Context similarity measure using Fuzzy Formal Concept Analysis. <b>2012</b> ,		0
139	KnoE: A Web Mining Tool to Validate Previously Discovered Semantic Correspondences. <b>2012</b> , 27, 1222-1232		2
138	Document and sentence alignment in comparable corpora using bipartite graph matching. <b>2012</b> ,		
137	Automatically Mining Parallel Corpora for Minority Languages from Web Pages. <b>2012</b> ,		0
136	Global Trends in Computing and Communication Systems. <i>Communications in Computer and Information Science</i> , <b>2012</b> ,	0.3	2
135	Smart combination of web measures for solving semantic similarity problems. <b>2012</b> , 36, 724-738		2
134	Translation techniques in cross-language information retrieval. <b>2012</b> , 45, 1-44		33
133	Mining Text Data. <b>2012</b> ,		460

132	Computational Linguistics and Intelligent Text Processing. <i>Lecture Notes in Computer Science</i> , <b>2013</b> ,	0.9
131	Collaboratively built semi-structured content and Artificial Intelligence: The story so far. <b>2013</b> , 194, 2-27	81
130	Cross-language information retrieval models based on latent topic models trained with document-aligned comparable corpora. <b>2013</b> , 16, 331-368	30
129	A unified alignment algorithm for bilingual data. <b>2013</b> , 19, 33-60	3
128	Mining Parallel Corpus from Sina Microblog. <b>2013</b> ,	
127	Creating Indonesian-Javanese parallel corpora using wikipedia articles. <b>2014</b> ,	1
126	Computational Linguistics and Intelligent Text Processing. <i>Lecture Notes in Computer Science</i> , <b>2014</b> ,	0.9 1
125	Natural Language Processing and Information Systems. <i>Lecture Notes in Computer Science</i> , <b>2014</b> ,	0.9 1
124	Construction of a large-scale Sino-Vietnamese bilingual parallel corpus. <b>2014</b> ,	
123	Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. <i>Lecture Notes in Computer Science</i> , <b>2014</b> ,	0.9
122	Generating Chinese named entity data from parallel corpora. <b>2014</b> , 8, 629-641	4
121	Parallel corpora make sense. <b>2014</b> , 19, 333-367	1
120	How much is said in a microblog?. <b>2015</b> ,	3
119	Machine translation from Japanese and French to Vietnamese, the difference among language families. <b>2015</b> ,	0
118	Extracting bilingual terms from the Web. <b>2015</b> , 21, 205-236	0
117	English-Arabic Statistical Machine Translation: State of the Art. <i>Lecture Notes in Computer Science</i> , <b>2015</b> , 520-533	0.9 2
116	Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. <b>2015</b> , 51, 111-147	49
115	Romanized urdu Corpus development (RUCD) model: Edit-distance based most frequent unique unigram extraction approach using real-time interactive dataset. <b>2016</b> ,	



114	Extracting parallel phrases from comparable data for machine translation□ <b>2016</b> , 22, 549-573		7
113	Building and using multimodal comparable corpora for machine translation□ <b>2016</b> , 22, 603-625		2
112	Sentence alignment using local and global information. <b>2016</b> , 39, 88-107		4
111	The effects of data-driven learning activities on EFL learnersSwriting development. <b>2016</b> , 5, 1255		11
110	Using the wayback machine to mine websites in the social sciences: A methodological resource. <b>2016</b> , 67, 1904-1915		33
109	Statistical Machine Translation System for Indian Languages. <b>2016</b> ,		4
108	Mining Parallel Corpora from Sina Weibo and Twitter. <i>Computational Linguistics</i> , <b>2016</b> , 42, 307-343	2.8	6
107	Identification of the parallel documents from multilingual news websites. <b>2016</b> ,		0
106	Twitter corpus creation: The case of a Malay Chat-style-text Corpus (MCC). <b>2016</b> , 31, 227-243		5
105	Capturing the Semantics of Key Phrases Using Multiple Languages for Question Retrieval. <b>2016</b> , 28, 888-900		20
104	Integrated Parallel Sentence and Fragment Extraction from Comparable Corpora. <i>ACM Transactions on Asian and Low-Resource Language Information Processing</i> , <b>2016</b> , 15, 1-22	1.1	12
103	On the feasibility of character n-grams pseudo-translation for Cross-Language Information Retrieval tasks. <b>2016</b> , 36, 136-164		5
102	State of the art in statistical methods for language and speech processing. <b>2016</b> , 35, 163-184		21
101	Building and evaluating web corpora representing national varieties of English. <i>Language Resources and Evaluation</i> , <b>2017</b> , 51, 643-662	1.8	5
100	Crawl and crowd to bring machine translation to under-resourced languages. <i>Language Resources and Evaluation</i> , <b>2017</b> , 51, 1019-1051	1.8	4
99	EMT. <b>2017</b> ,		
98	A Comprehensive Analysis of Bilingual Lexicon Induction. <i>Computational Linguistics</i> , <b>2017</b> , 43, 273-310	2.8	7
97	Hiencor: On mining of a hi-en general purpose parallel corpus from the web. <b>2017</b> ,		

96	Language and translation on the Web. <b>2017</b> , 357-373		
95	Automatically Extracting Parallel Sentences from Wikipedia Using Sequential Matching of Language Resources. <b>2017</b> , E100.D, 405-408		0
94	Revolt. <b>2017</b> ,		73
93	Survey and evaluation of web search engine hit counts as research tools in computational linguistics. <b>2018</b> , 73, 50-60		12
92	A resource-light method for cross-lingual semantic textual similarity. <b>2018</b> , 143, 1-9		18
91	Word Embedding based Semantic Cross-Lingual Document Alignment in Comparable Corpora. <b>2018</b> ,		
90	Explaining the complexities of cross-linguistic features using comparable Arabic and English corpora. <b>2018</b> , 13, 135-168		
89	PC-Corpus: A Persian-Chinese Parallel Corpora. <b>2019</b> , 1176, 022002		
88	Efficient document alignment across scenarios. <i>Machine Translation</i> , <b>2019</b> , 33, 205-237	1.1	1
87	Zero-Shot Language Transfer for Cross-Lingual Sentence Retrieval Using Bidirectional Attention Model. <i>Lecture Notes in Computer Science</i> , <b>2019</b> , 523-538	0.9	0
86	A Hybrid of Sentence-Level Approach and Fragment-Level Approach of Parallel Text Extraction from Comparable Text. <b>2019</b> , 161, 406-414		1
85	Arabic word sense disambiguation: a review. <b>2019</b> , 52, 2475-2532		7
84	On the Pronunciation Dictionaries of Contemporary German: Principles of Selection and Lemmatization of Lexical Material. <b>2020</b> , 33, 337-357		
83	An Automatic and a Machine-assisted Method to Clean Bilingual Corpus. <i>ACM Transactions on Asian and Low-Resource Language Information Processing</i> , <b>2020</b> , 19, 1-19	1.1	0
82	Fostering intuitive competence in L2 for a better performance in EAP writing through fraze.it in a Turkish context. <b>2020</b> , 25, 5405-5426		
81	Document Alignment for Generation of English-Punjabi Comparable Corpora from Wikipedia. <b>2020</b> , 12, 42-51		1
80	Preparation of Sentiment tagged Parallel Corpus and Testing Its Effect on Machine Translation. <b>2021</b> , 131-145		
79	Parallel sentence extraction to improve cross-language information retrieval from Wikipedia. <b>2021</b> , 47, 281-293		2

78	Deep learning approach for Translating Arabic Holy Quran into Italian language. <b>2021</b> ,		0
77	Augmenting Training Data for Low-Resource Neural Machine Translation via Bilingual Word Embeddings and BERT Language Modelling. <b>2021</b> ,		
76	Automatic Evaluation of Ontologies (AEON). <i>Lecture Notes in Computer Science</i> , <b>2005</b> , 716-731	0.9	19
75	Automatic Acquisition of ChineseEnglish Parallel Corpus from the Web. <i>Lecture Notes in Computer Science</i> , <b>2006</b> , 420-431	0.9	25
74	Construct Trilingual Parallel Corpus on Demand. <i>Lecture Notes in Computer Science</i> , <b>2006</b> , 760-767	0.9	5
73	Translingual Mining from Text Data. <b>2012</b> , 323-359		3
72	Paragraph Alignment for English-Vietnamese Parallel E-Books. <i>Advances in Intelligent Systems and Computing</i> , <b>2014</b> , 251-259	0.4	1
71	Constructing a Large Scale Text Corpus Based on the Grid and Trustworthiness. <b>2007</b> , 56-65		1
70	Learning Medical Ontologies from the Web. <b>2007</b> , 32-45		6
69	Overviewing Important Aspects of the Last Twenty Years of Research in Comparable Corpora. <b>2013</b> , 1-17		3
68	Extracting Parallel Phrases from Comparable Data. <b>2013</b> , 191-204		4
67	Extracting Parallel Paragraphs and Sentences from English-Persian Translated Documents. <i>Lecture Notes in Computer Science</i> , <b>2011</b> , 574-583	0.9	2
66	Improved Text Extraction from PDF Documents for Large-Scale Natural Language Processing. <i>Lecture Notes in Computer Science</i> , <b>2014</b> , 102-112	0.9	4
65	Unsupervised Bitext Mining and Translation via Self-Trained Contextual Embeddings. <b>2020</b> , 8, 828-841		2
64	High-Precision Sentence Alignment by Bootstrapping from Word Standard Annotations. <i>Prague Bulletin of Mathematical Linguistics</i> , <b>2013</b> , 99, 5-16	0.3	2
63	Unsupervised Translation Disambiguation Based on Web Indirect Association of Bilingual Word. <b>2010</b> , 21, 575-585		1
62	The Use of MSVM and HMM for Sentence Alignment. <b>2012</b> , 8, 301-314		7
61	An appraisal of recent breakthroughs in machine translation: the case of past participle-based compound adjectives in ESP. <b>2017</b> , 29-48		1

60	Mining Parallel Knowledge from Comparable Patents. <b>2011</b> , 247-271	2
59	Parallel Sentence Extraction Based on Unsupervised Bilingual Lexicon Extraction from Comparable Corpora. <b>2015</b> , 22, 139-170	2
58	Multilingual Sentence Hunter. <i>Lecture Notes in Computer Science</i> , <b>2005</b> , 84-93	0.9
57	Extracting English-Korean Transliteration Pairs from Web Corpora. <i>Lecture Notes in Computer Science</i> , <b>2006</b> , 222-233	0.9
56	Automatically Extracting Unknown Translations Using Phrase Alignment. <b>2007</b> , 14B, 231-240	
55	Query Classification and Expansion for Translation Mining Via Search Engines. <i>Lecture Notes in Computer Science</i> , <b>2008</b> , 1121-1126	0.9
54	Cross-Language Information Retrieval on the Web. <b>2009</b> , 704-719	1
53	Utilizing Web Directories for Translation Disambiguation in Cross-Language Information Retrieval. <b>2009</b> , 95-107	
52	Mining Parenthetical Translations for Polish-English Lexica. <i>Lecture Notes in Computer Science</i> , <b>2010</b> , 464-472	0.9
51	Mining Parallel Documents across Web Sites. <i>Lecture Notes in Computer Science</i> , <b>2010</b> , 552-563	0.9
50	Unsupervised Translation Disambiguation by Using Semantic Dictionary and Mining Language Model from Web. <b>2010</b> , 20, 1292-1300	0
49	Cross-language Ontology Learning. <b>2011</b> , 272-297	
48	Performance Analysis of Case Based Word Sense Disambiguation with Minimal Features Using Neural Network. <i>Communications in Computer and Information Science</i> , <b>2012</b> , 687-696	0.3
47	QRpac: User-Driven Archiving of Parallel and Comparable Documents from the Web. <i>Lecture Notes in Computer Science</i> , <b>2012</b> , 321-324	0.9
46	Parallel Texts Extraction from Multimodal Comparable Corpora. <i>Lecture Notes in Computer Science</i> , <b>2012</b> , 40-51	0.9
45	Mining Parallel Documents Using Low Bandwidth and High Precision CLIR from the Heterogeneous Web. <b>2013</b> , 21-49	
44	Comparable Multilingual Patents as Large-Scale Parallel Corpora. <b>2013</b> , 167-187	
43	Exploiting Comparable Corpora. <b>2013</b> , 205-222	

42	Leveraging Arabic-English Bilingual Corpora with Crowd Sourcing-Based Annotation for Arabic-Hebrew SMT. <i>Lecture Notes in Computer Science</i> , <b>2013</b> , 297-310	0.9	
41	Bilingual Terminology Mining from Language for Special Purposes Comparable Corpora. <b>2013</b> , 265-284		1
40	Automatic Collection of the Parallel Corpus with Little Prior Knowledge. <i>Lecture Notes in Computer Science</i> , <b>2014</b> , 95-106	0.9	1
39	Extracting Parallel Fragments from Comparable Documents Using a Feature-Based Method. <i>Communications in Computer and Information Science</i> , <b>2014</b> , 288-298	0.3	
38	Enhancing Multilingual Biomedical Terminologies via Machine Translation from Parallel Corpora. <i>Lecture Notes in Computer Science</i> , <b>2014</b> , 9-20	0.9	0
37	Cross-Language Mining and Retrieval. <b>2016</b> , 1-7		
36	FaDA: Fast Document Aligner using Word Embedding. <i>Prague Bulletin of Mathematical Linguistics</i> , <b>2016</b> , 106, 169-179	0.3	0
35	Encyclopedia of Database Systems. <b>2017</b> , 1-7		
34	Simple and Accurate Method for Parallel Web Pages Detection. <i>Lecture Notes in Computer Science</i> , <b>2017</b> , 238-247	0.9	
33	Harvest Uyghur-Chinese Aligned-Sentences Bitexts from Multilingual Sites Based on Word Embedding. <i>Lecture Notes in Computer Science</i> , <b>2017</b> , 449-460	0.9	
32	Encyclopedia of Database Systems. <b>2018</b> , 667-672		
31	English-Cebuano Parallel Language Resource for Statistical Machine Translation System. <i>Advances in Intelligent Systems and Computing</i> , <b>2018</b> , 49-60	0.4	
30	Knowledge-Enhanced Bilingual Textual Representations for Cross-Lingual Semantic Textual Similarity. <i>Communications in Computer and Information Science</i> , <b>2019</b> , 425-440	0.3	
29	The TransBank Aligner: Cross-Sentence Alignment with Deep Neural Networks. <i>Lecture Notes in Computer Science</i> , <b>2019</b> , 185-196	0.9	
28	Integrating Local and Global Data View for Bilingual Sense Correspondences. <i>Communications in Computer and Information Science</i> , <b>2019</b> , 224-241	0.3	
27	Working with parallel corpora. <i>Studies in Corpus Linguistics</i> , <b>2019</b> , 57-78	0.3	0
26	Pretrained Transformers for Text Ranking: BERT and Beyond. <i>Synthesis Lectures on Human Language Technologies</i> , <b>2021</b> , 14, 1-325	2.3	24
25	Recent advances of low-resource neural machine translation. <i>Machine Translation</i> , 1	1.1	0

24	Improved Cross-Lingual Document Similarity Measurement. <b>2020</b> ,		
23	The suffix -ee: history, productivity, frequency and violation of st. <b>2020</b> ,		
22	Semantic-Aware Deep Neural Attention Network for Machine Translation Detection. <i>Communications in Computer and Information Science</i> , <b>2021</b> , 63-76	0.3	
21	Crawling Chinese-Myanmar Parallel Corpus: Automatic Collection, Screening and Cleaning Corpus. <i>IOP Conference Series: Materials Science and Engineering</i> , 646, 012046	0.4	
20	Parallel Corpora. <b>2020</b> , 257-282		2
19	Web Corpora. <b>2020</b> , 329-351		
18	Mining Scientific and Technical Literature. <i>Advances in Data Mining and Database Management Book Series</i> , <b>2020</b> , 61-87	0.6	
17	A Refinement Framework for Cross Language Text Categorization. <b>2008</b> , 401-411		
16	Zero Anaphora Resolution in Chinese and Its Application in Chinese-English Machine Translation. <i>Lecture Notes in Computer Science</i> , <b>2007</b> , 364-375	0.9	
15	QRselect: A User-Driven System for Collecting Translation Document Pairs from the Web. <b>2007</b> , 131-140		2
14	Impact assessment indicators for the UK Web Archive. <i>Performance Measurement and Metrics</i> , <b>2021</b> , ahead-of-print,	0.9	
13	Refined Urdu Lexicon Development K-Means Clustering Based Computational Model Using Colloquial Romanized Urdu Dataset. <b>2020</b> ,		
12	Sinhala and English Document Alignment using Statistical Machine Translation. <b>2020</b> ,		
11	Mining Japanese-Vietnamese multi-level parallel text corpus from Wikipedia data resource. <b>2021</b> ,		
10	Making the most of comparable corpora in Neural Machine Translation: a case study. <i>Language Resources and Evaluation</i> , 1	1.8	
9	Leveraging Multilingual News Websites for Building a Kurdish Parallel Corpus. <i>ACM Transactions on Asian and Low-Resource Language Information Processing</i> ,	1.1	0
8	Simplification of English and Bengali Sentences for Improving Quality of Machine Translation. <i>Neural Processing Letters</i> , 1	2.4	1
7	Design of New Word Retrieval Algorithm for Chinese-English Bilingual Parallel Corpus. <i>Mathematical Problems in Engineering</i> , <b>2022</b> , 2022, 1-9	1.1	

- 6 Automatic Acquisition of Lexical Information and Examples. **2007**, 253-274
- 5 Exploiting bilingual lexicons to improve multilingual embedding-based document and sentence alignment for low-resource languages. ○
- 4 Tailoring and evaluating the Wikipedia for in-domain comparable corpora extraction. ○
- 3 Adapting the Tesseract Open-Source OCR Engine for Tamil and Sinhala Legacy Fonts and Creating a Parallel Corpus for Tamil-Sinhala-English. **2022**, ○
- 2 Loanword identification based on web resources: A case study on wikipedia. **2023**, 81, 101517 ○
- 1 Addressing the Issue of Unavailability of Parallel Corpus Incorporating Monolingual Corpus on PBSMT System for English-Manipuri Translation. **2023**, 299-319 ○