Bibliography of Sketch Engine: Papers and Presentations

See How to reference Sketch Engine.

See Adam Kilgarriff's bibliography.

2015

- Using lecture slides to create an academic corpus
  - Simon Smith (2015)
  - In IATEFL 2014 Harrogate Conference Selections, Pattison Tania (ed.), Faversham, Kent: IATEFL, pp. 149--151

- Semantic Word Sketches
  - Diana McCarthy and Adam Kilgarriff (2015)
  - Accepted for Corpus Linguistics (CL2015), United Kingdom, July 2015

- DIACRAN: a framework for diachronic analysis
  - Adam Kilgarriff, Jan Busta and Pavel Rychlý (2015)
  - Accepted for Corpus Linguistics (CL2015), United Kingdom, July 2015

- Longest-commonest match
  - Accepted for Corpus Linguistics (CL2015), United Kingdom, July 2015

- Sketch Engine for English Language Learning
  - Vít Baisa, Vít Suchomel and Adam Kilgarriff (2015)
  - Accepted for Corpus Linguistics (CL2015), United Kingdom, July 2015

2014

- Effective Corpus Virtualization
  - In Challenges in the Management of Large Corpora (CMLC-2), May 2014

- The Sketch Engine: Ten Years On
  - In Lexicography: Journal of ASIALEX, volume 1, issue 1, pp. 7--36

- Finding Terms in Corpora for Many Languages with the Sketch Engine
  - In Proceedings of the Demonstrations at the 14th Conference the European Chapter of the Association for Computational Linguistics, Sweden, April 2014, pp. 53--56

- arTenTen: Arabic Corpus and Word Sketches
  - Tressy Arts, Yonatan Belinkov, Nizar Habash, Adam Kilgarriff and Vít Suchomel (2014)
In *Journal of King Saud University - Computer and Information Sciences*, volume 26, issue 4, December 2014, pp. 381--395

- **Hindi Word Sketches?**
  - Anil Krishna Eragani, Varun Kuchibhotla, Dipti Sharma, Siva Reddy and Adam Kilgarriff (2014)

- **PtTenTen: A corpus for Portuguese lexicography**

- **Text Tokenisation Using unitok**
  - Vít Suchomel, Jan Michelfeit and Jan Pomikálek (2014)
  - In *Proceedings of the Eighth Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2014*, Czech Republic, December 2014, pp. 71--75

- **Extrinsic Corpus Evaluation with a Collocation Dictionary Task**
  - In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, Iceland, May 2014, pp. 545--552

- **Sketching the Dependency Relations of Words in Chinese**
  - Meng-Hsien Shih and Shu-Kai Hsieh
  - In *Proceedings of the 26th Conference on Computational Linguistics and Speech Processing (ROCLING 2014)*, Taiwan, September 2014, pp. 139--152

- **Bilingual Word Sketches: the translate Button**
  - In *Proceedings of the 16th EURALEX International Congress. 15--19 July 2014*, Bolzano, Italy, pp. 505--513

- **Compatible Sketch Grammars for Comparable Corpora**
  - Vladimír Benko

- **Aranuem Nederlandicum Maius, A New Family Member?**
  - Vladimír Benko (2014)
  - *CLIN24 Workshop in Leiden, Sketch examples?*

2013

- **Sketch Engine as a Platform for Providing Corpora?**
  - Presentation at META-FORUM Germany, September 2013 (link to ?YouTube)

- The TenTen Corpus Family
  In Proceedings of the 7th International Corpus Linguistics Conference CL 2013, United Kingdom, July 2013, pp. 125--127

• Web Spam
  Adam Kilgarriff and Vít Suchomel (2013)
  In Proceedings of the 8th Web as Corpus Workshop (WAC-8), United Kingdom, July 2013, pp. 46--52

• arTenTen: a new, vast corpus for Arabic
  Yonatan Belinkov, Nizar Habash, Adam Kilgarriff, Noam Ordan, Ryan Roth and Vít Suchomel (2013)
  In Proceedings of WACL?2 Second Workshop on Arabic Corpus Linguistics, United Kingdom, July 2013, pp. 20

• Intrinsic Methods for Comparison of Corpora
  Vít Baisa and Vít Suchomel (2013)
  In Proceedings of the Seventh Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2013, Czech Republic, December 2013, pp. 51--58

• Compatible Sketch Grammar Experiment
  Vladimír Benko (2013)

• User-friendly interface of error/correction-annotated corpus for both teachers and researchers (The Sketch Engine interface for a learner corpus annotated with errors and corrections)
  In In Book of Abstracts LCR 2013, Norway, September 2013, pp. 82--84

• Corpus-based vocabulary lists for language learners for nine languages
  In Language Resources and Evaluation, volume 48, issue 1, March 2014, pp. 121--163

• esTenTen, a vast web corpus of Peninsular and American Spanish?
  Adam Kilgarriff and Irene Renau (2013)
  In Procedia - Social and Behavioral Sciences, volume 55, Elsevier, October 2013, pp. 12--19

• Using corpora as data sources for dictionaries?
  Adam Kilgarriff (2013)
  In The Bloomsbury Companion to Lexicography, Howard Jackson (ed.), Bloomsbury, London. Chapter 4.1, pp. 77--96

• Terminology finding, parallel corpora and bilingual word sketches in the Sketch Engine??
  Adam Kilgarriff (2013)

• ?????????????????????????????
  (Japanese Language Lexical and Grammatical Profiling Using the Web Corpus JpTenTen)
• Irena Srdanović, Vít Suchomel, Toshinobu Ogiso and Adam Kilgarriff (2013)
  • ???3??????????????????????????????????????????????? (In Proceeding of the 3rd Japanese corpus linguistics workshop, Department of Corpus Studies, Center for Corpus Development, NINJAL), pp. 229--238

• ?Quantifying Lexical Usage: Vocabulary pertaining to Ecosystems and the Environment
  • Kate Wild, Andrew Church, Diana McCarthy and Jacqueline Burgess
  • In Corpora, volume 8, May 2013, pp. 53--79

2012

• New Learner Corpus Functionality in the Sketch Engine?
  • Vojtěch Kovář and Diana McCarthy (2012)
  • In Proceedings of the 2012 Asia Pacific Corpus Linguistics Conference (APCLC), Auckland, February 2012

• ?Performance stylistics: Deleuze and Guattari, poetry and (corpus) linguistics
  • Kieran O’Halloran (2012)

• ?Genres across the Disciplines Student Writing in Higher Education
  • Hilary Nesi and Sheena Gardner, 2012
  • Cambridge University Press

• ?Word Sense Induction for Novel Sense Detection
  • Jey Han Lau, Paul Cook, Diana McCarthy, David Newman and Timothy Baldwin (2012)
  • In 13th Conference of the European Chapter of the Association for computational Linguistics (EACL 2012), France, April 2012, pp. 591--601

• Efficient Web Crawling for Large Text Corpora
  • Vít Baisa, Vít Suchomel and Jan Pomikálek (2012)
  • In Proceedings of the seventh Web as Corpus Workshop (WAC7), France, April 2012, pp. 39--43

• ?Getting to know your corpus
  • Adam Kilgarriff (2012)
  • In Proceedings of The 15th International Conference on Text, Speech and Dialogue (TSD), Petr Šojka, Aleš Horák, Ivan Kope?ek and Karel Pala (eds.), Czech Republic, September 2012, pp. 3--15

• ?Detecting Spam in Web Corpora
  • Vít Baisa and Vít Suchomel (2012)
  • In Proceedings of the Sixth Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2012, Czech Republic, December 2012, pp. 69--76

• ?Recent Czech Web Corpora
  • Vít Suchomel (2012)
  • In Proceedings of the Sixth Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2012, Czech Republic, December 2012, pp. 77--83

• ?Finding Multiwords of More Than Two Words
• Large Corpora for Turkic Languages and Unsupervised Morphological Analysis
  ◆ Vít Baisa and Vít Suchomel (2012)

• *Word Sketches for Turkish?*
  ◆ Bharat Ram Ambati, Siva Reddy and Adam Kilgarriff (2012)

• *Managing Ambiguity in Reference Generation: The Role of Surface Structure*
  ◆ Imtiaz H. Khan, Kees van Deemter and Graeme Ritchie

• *Learner corpora and second language acquisition*
  ◆ Meng Huat Chau (2012)

• Setting up for corpus lexicography?

• *Corpus Tools for Lexicographers?*
  ◆ Adam Kilgarriff and Iztok Kosem (2012)

• *Tools for historical corpus research, and a corpus of Latin??*
  ◆ Barbara McGillivray and Adam Kilgarriff (2012)
  ◆ In *New Methods in Historical Corpus Linguistics 3*, Germany, 2013, pp. 247–255

• *Vietnamese Word Sketches?*
  ◆ Adam Kilgarriff and Phuong Le-Hong (2012)
  ◆ In *Workshop on Vietnamese Language and Speech Processing (IEEE-RIVF 9)*, Vietnam, February 2012

2011

• *Corpora and Language Education*
  ◆ Lynne Flowerdew (2011)
  ◆ Palgrave Macmillan, December 2011

• *Comparable Corpora BootCaT*
  ◆ Adam Kilgarriff, Avinesh PVS and Jan Pomikálek (2011)
  ◆ In *Proceedings of eLEX 2011*, Slovenia, November 2011, pp. 122–128
• **GDEX for Slovene**
  ♦ Iztok Kosem, Miloš Husák and Diana McCarthy (2011)
  ♦ In *Proceedings of eLEX 2011*, Slovenia, November 2011, pp. 151--159

• Dynamic and Static Prototype Vectors for Semantic Composition
  ♦ Siva Reddy, Ioannis P. Klapaftis, Diana McCarthy and Suresh Manandhar (2011)
  ♦ Best paper award at the 5th International Joint Conference on Natural Language Processing.

• Learner construction of corpora for general English in Taiwan
  ♦ Simon Smith (2011)
  ♦ In *Computer Assisted Language Learning*, volume 24, Issue 4, pp. 291--316

• **Corpus-based Disambiguation for Machine Translation**
  ♦ Vít Baisa (2011)
  ♦ In *Proceedings of the Fifth Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2011*, Czech Republic, December 2011, pp. 81--87

• Corpus-based tasks for learning Chinese: a data-driven approach
  ♦ Simon Smith and Xuanying Shen (2011)
  ♦ In *The Asian Conference on Technology in the Classroom 2011* Japan, August 2011, pp. 48--59

• **Large Web Corpora for Indian Languages**
  ♦ Adam Kilgarriff and Girish Duvuru (2011)
  ♦ In *Proceedings of International Conference on Information Systems for Indian Languages (ICISIL)*, India, 2011 pp. 312--313

• **Polish Word Sketches?**
  ♦ Adam Radziszewski, Adam Kilgarriff and Robert Lew (2011)
  ♦ In *Proceedings of the 5th Language & Technology Conference (LTC)*, Poland, November 2012, pp. 237--242

• Japanese Word Sketches: Advances and Problems
  ♦ In *Acta Linguistica Asiatica*, University of Ljubljana, Slovenia 2011, pp. 63--82

• **The Pearson International Corpus of Academic English (PICAE)? (PICAE poster)**
  ♦ In *International Corpus Linguistics Conference (ICLIC)*, Birmingham, UK, 2011

### 2010

• **Database of ANalysed Texts of English (DANTE): the NEID database project**
  ♦ Sue Atkins, Adam Kilgarriff, Michael Rundell (2010)
  ♦ In *Proceedings of the 14th EURALEX International Congress*. The Netherlands, July 2010, pp. 549--556
  ◊ *(sample entry here)*

• **Usi e pratiche della comprensione attraverso la lente dei verba reciproci? (in Italian)**
  ♦ Isabella Chiari (2010)
• Bollettino di Italianistica, I, pp. 30--70

• Helping Our Own
  ♦ Robert Dale and Adam Kilgarriff (2010)
  ♦ In *International Natural Language Generation Conference*, Dublin, Ireland

• Fast syntactic searching in very large corpora for many languages
  ♦ In *Proceedings of Workshop on Advanced Corpus Solutions*, PACLIC 24, Japan, November 2010

• Studying Word Sketches for Russian
  ♦ Maria Khokhlova and Victor Zakharov (2010)
  ♦ In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’12)* Malta, May 2010

• Building Russian Word Sketches as Models of Phrases
  ♦ Maria Khokhlova (2010)
  ♦ In *Proceedings of the 14th EURALEX International Congress*. The Netherlands, July 2010, pp. 364--371

• A Case Study in Word Sketches - Czech Verb vid?t?
  ♦ Karel Pala and Pavel Rychlý (2010)

• Google The Verb
  ♦ Adam Kilgarriff (2010)

• Tickbox Lexicography?

• Semi-automatic Dictionary Drafting?
  ♦ Adam Kilgarriff and Pavel Rychlý (2010)
  ♦ In *A Way with Words: Recent Advances in Lexical Theory and Analysis*, Uganda: Menha Publishers Ltd., 2010, 299--312

• Corpora by Web Services
  ♦ Adam Kilgarriff (2010)
  ♦ In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’12)* Malta, May 2010

• A Corpus Factory for Many Languages
  ♦ Adam Kilgarriff, Siva Reddy, Jan Pomikálek and Avinesh PVS (2010)
  ♦ LREC workshop on Web Services and Processing Pipelines, Malta, May 2010

• The RoWaC Corpus and Romanian Word Sketches
  ♦ Monica Macoveiciuc and Adam Kilgarriff (2010)
In *Multilinguality and Interoperability in Language Processing with Emphasis on Romanian* Edited by Dan Tufis and Corina Forasucu. Romanian Academy

- **DANTE: A Detailed, Accurate, Extensive, Available English Lexical Database**
  - Adam Kilgarriff (2010)
  - In *Proceedings of the NAACL HLT 2010 Demonstration Session*, Los Angeles, June 2010, pp. 21--24

- **A Quantitative Evaluation of Word Sketches**
  - In *Proceedings of the 14th EURALEX International Congress*. The Netherlands, July 2010

- **Comparable Corpora Within and Across Languages, Word Frequency Lists and the Kelly Project**
  - Adam Kilgarriff (2010)
  - Invited talk at the LREC workshop on Building and Using Comparable Corpora, Malta, May 2010

- **SemEval-2010 Task 7: Argument Selection and Coercion**

### 2009

- **Scaling to Billion-plus Word Corpora**
  - Jan Pomikálek, Pavel Rychlí and Adam Kilgarriff
  - In *Advances in Computational Linguistics*, Instituto Politécnico Nacional, volume 41, Mexico, 2009, pp. 3--13

- **Simple maths for keywords**
  - Adam Kilgarriff

- **The metalanguage of impoliteness: Using Sketch Engine to explore the Oxford English Corpus**
  - Jonathan Culpeper (2009)
  - In *Contemporary Corpus Linguistics* P. Baker (ed.), London: Continuum, pp. 64?-86

- **Extracting distant collocations of adverbs and modality forms using web corpus and query system**
  - (Extracting distant collocations of adverbs and modality forms using web corpus and query system, Journal of Natural Language Processing), 16/4, pp. 29--46

- **Towards creation of lexical syllabus based on corpora - on suppositional adverbs and clause-final modality collocations**
  - (Towards creation of lexical syllabus based on corpora - on suppositional adverbs and clause-final modality collocations, Journal of Japanese Language Education, 142, pp. 69--79)

- **Chapter III - WebCorp: the Web as Corpus**
• Maristella Gatto (2009) (email ?m.gatto@lingue.uniba.it)
  ♦ In From Body to Web. An Introduction to the Web as Corpus, Laterza University Press Online (Bari-Roma) 2009, pp. 77--100

• The Sketch Engine for Dutch with the ANW corpus
  ♦ Carole Tiberius and Adam Kilgarriff (2009)
  ♦ In Fons Verbhorum, Festschrift for Fons Moerdijk. Instituut voor Nederlandse Lexicologie, the Netherlands, pp. 273--255

• Classifying corpora based on adverbs distribution
  ♦ International Quantitative Linguistics Conference (Qualico), Austria, September 2009

2008

• A Lexicographer-Friendly Association Score
  ♦ Pavel Rychlý (2008)

• Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine
  ♦ Michael Pearce (2008)
  ♦ Corpora. Volume 3, DOI 10.3366/E174950320800004X, ISSN 1749-5032, pp. 1--29

• GDEX: Automatically finding good dictionary examples in a corpus
  ♦ Adam Kilgarriff, Milo? Husák, Katy McAdam, Michael Rundell and Pavel Rychlý (2008)

• Finding the words which are most X.
  ♦ Adam Kilgarriff and Pavel Rychlý (2008)

• Comparing Lexical Relationships Observed within Japanese Collocation Data and Japanese Word Association Norms.
  ♦ Terry Joyce and Irena Srdanovi? (2008)

• A web corpus and word sketches for Japanese

• The Sketch Engine corpus query tool for Japanese and its possible applications (in Japanese)
  ♦ ?????????????Sketch Engine?????????????????????????????????????????(The Sketch Engine corpus query tool for Japanese and its possible applications, Japanese Linguistics) 23, pp. 59--80
2007 and 2006

• *Manatee/bonito? a modular corpus manager*
  ♦ Pavel Rychlé (2007)

• *An efficient algorithm for building a distributional thesaurus (and other Sketch Engine developments)*
  ♦ Pavel Rychlé and Adam Kilgarriff (2007)
  ♦ In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*. Czech Republic, June 2007, pp. 41--44

• *Slovene Word Sketches*
  ♦ In *Proceedings 5th Slovenian/First International Languages Technology Conference*, Slovenia, October 2006

• *Learning noun-modifier semantic relations with corpus-based and "WordNet"-based features*
  ♦ Vivi Nastase, Jelber Sayyad-Shirabad, Marina Sokolova and Stan Szpakowicz (2006)
  ♦ In *Proceedings of the National Conference on Artificial Intelligence*. Massachusetts, July 2006. Published by The AAAI Press, Menlo Park, California, pp. 781--786

• *WebBootCaT: a web tool for instant corpora*
  ♦ Marco Baroni, Kilgarriff, Jan Pomikálek and Pavel Rychlé (2006)

• WebBootCaT: instant domain-specific corpora to support human translators
  ♦ Marco Baroni, Kilgarriff, Jan Pomikálek and Pavel Rychlé (2006)
  ♦ Proceedings of EAMT. 11th Annual Conference of the European Association for Machine Translation. Oslo, Norway

2005 and earlier

• *Peta poletna delavnica za leksikografijo in leksikalno ra?unalni?tvo*
  ♦ Irena Črdanovi? (2005)

• *Disambiguating Coordinations Using Word Distribution Information*
  ♦ Francis Chantree, Adam Kilgarriff, Anne de Roeck and Alistair Willis
  ♦ In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, Bulgaria, September 2005

• *Chinese Sketch Engine and the Extraction of Grammatical Collocations*
  ♦ Chu-Ren Huang, Adam Kilgarriff, Yiching Wu, Chih-Ming Chiu, Simon Smith, Pavel Rychlé, Ming-Hong Bai and Keh-Jiann Chen (2005)
  ♦ In *Fourth SIGHAN Workshop on Chinese Language Processing*, Korea, October 2005, pp. 48--55

• *Putting the corpus into the dictionary*
♦ Adam Kilgarriff (2005)
♦ In *Proceedings of the MEANING-2005 workshop*, Italy, February 2005

• The sketch engine

If you have any Sketch Engine related paper please do send the details and if possible a link to the document to us (email: ?support@sketchengine.co.uk)