Bibliography for LDC June 2017 Talk

Online Language Archives: Corpora, Ontologies, and Online Language Learning Resources

Sinica Corpus: Academia Sinica Balanced Corpus for Mandarin Chinese中央研究院現代漢語平衡語料庫. November 1996 (First Web Version).
http://asbc.iis.sinica.edu.tw/

搜文解字 - 語文知識網路SouWenJieZi - A Linguistic KnowledgeNet. August 1999. http://words.sinica.edu.tw/

文國尋寶記—中小學語文知識網路。 Adventures in Wen-Land. November 2000. http://wen.ling.sinica.edu.tw/

「荔鏡姻，河洛緣」- 閩南語第一名著《荔鏡記》數位博物館。 September 2001. http://cls.hs.yzu.edu.tw/LM/

Sinica BOW: Academia Sinica Bilingual Ontological Wordnet中央研究院中英雙語知識本體詞網. October 2003. http://BOW.ling.sinica.edu.tw

Sinica TreeBank中文句結構樹資料庫. April 2004. http://TreeBank.sinica.edu.tw/

Chinese Word Sketch (Prototype) 2005. http://corpora.fi.muni.cz/chinese_all/

Chinese Word Sketch 2006. http://wordsketch.ling.sinica.edu.tw

Chinese Wordnet (Prototype) 2005. http://cwn.ling.sinica.edu.tw

Hantology 漢字知識本體. 2006. http://hantology.ling.sinica.edu.tw

General References

Huang, Chu-Ren, Barbara Meisterernst, and Zhuo Jing-Schmidt. (Eds.). In Preparation (2018). Routledge Handbook on Chinese Applied Linguistics. London: Routledge.

Lu, Qin, Nianwen Xue, and Chu-Ren Huang. In Preparation. Computer Processing of Chinese. Studies in Natural Language Processing Book Series. Cambridge University Press.

Huang, Chu-Ren, Yen-Hwei Lin, and I-Hsuan Chen. In Preparation. (Eds.) Cambridge Handbook of Chinese Linguistics. Cambridge: Cambridge University Press.

Huang, Chu-Ren, Shu-Kai Hsieh, and Keh-Jiann Chen. (2017). Mandarin Chinese Words and Parts of Speech: A corpus-based study. London: Routledge

Huang, Chu-Ren, and Dingxu Shi. (2016). A Reference Grammar of Chinese. Cambridge: Cambridge University Press.
http://crg.cbs.polyu.edu.hk

Emotion/Cause Annotated Corpus

Lee, Sophia YM. In Preparation. A Linguistic Approach to Emotion Detection and Classification. [tentative title]. To Appear in Studies in East Asian Linguistics book series. Springer.Lee, Sophia Yat-Mei, Ying Chen, Chu-Ren Huang and Shoushan Li. 2013. Detecting emotion causes with a linguistic rule-based approach. Computational Intelligence. 29(3). 390-416.

Lee Sophia Yat Mei李逸薇, Shou-Shan Li李壽山, Churen Huang黃居仁, and Wei Gao 高偉. 2013. Detecting emotion cause with sequence labeling model基於序列標注模型的情緒原因識別方法. Journal of Chinese Information Processing《中文信息學報》. 27(5). 93-99.

Chen, Ying, Sophia Yat Mei Lee, and Chu-Ren Huang. 2011. Automatic recognition of emotion based on a cognitively motivated emotion annotation system. Journal of Cognitive Science. 12(3). 277-296.

Lee, Sophia Y.M., Shoushan Li, and Chu-Ren Huang. 2014. Annotating Event in an Emotion Corpus. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14). Reykjavik, Iceland. 19-24 May.Yao, Yao, Jingxia, Lin and Chu-Ren Huang. 2013. Lexicalized emotion? Tonal patterns of emotion words in Mandarin Chinese. The 25th annual conference of North American Conference on Chinese Linguistics (NACCL-25). Ann Arbor, MI. 21 -23 June.

Lee, Sophia Yat Mei, Huarui Zhang, Chu-Ren Huang. 2013. An Event-based Emotion Corpus. The 14th Chinese Lexical Semantics Workshop. 10-12 May. Zhengzhou.

Lee, Sophia Yat Mei, Huarui Zhang, Chu-Ren Huang. 2013. Emotion as a Pivot Event. The 21th Annual Meeting of the International Association of Chinese Linguistics. Taipei: Taiwan Normal University.

Chen, Ying, Sophia Yat Mei Lee, Shoushan Li and Chu-Ren Huang. 2010. Emotion Cause Detection with Linguistic Constructions. Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). August 23-27, Beijing, China, pp179-187.

Huang, Chu-Ren, Ying Chen and Sophia Yat Mei Lee. 2010. Textual Emotion Processing From Event Analysis. 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010). July 11-16, Uppsala, Sweden.

Lee, Sophia Yat Mei, Ying Chen and Chu-Ren Huang. 2010. A Text-driven Rule-based System for Emotion Cause Detection. NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 5, Los Angeles, USA.

Example Database of A Reference Grammar of Chinese

Xu, Hongzhi, Helen Kaiyun Chen, Chu-Ren Huang, Qin Lu, Tin-Shing Chiu, and Dingxu Shi. 2012. A Grammar-informed Corpus-based Sentence Database for Linguistic and Computational Studies. The 2012 Language Resources and Evaluation Conference (LERC). Istanbul.

Semantic Transparency

Wang, Shichang. (2016). Crowdsourcing Method in Empirical Linguistic Research: Chinese Studies Using Mechanical Turk-Based Experimentation. Assistant Professor, Shandong University. PolyU PhD Thesis.

Huang, Chu-Ren, and Shichag Wang. 2016. 众包策略在语言资源建设中的应用 The Application of Crowdsourcing Strategy in Utilizing Language Resources. Chinese Journal of Language Policy and Planning (语言战略研究). 2016.6.36.46.

Wang, S.; Huang, C.-R.; Yao, Y. & Chan, A. (2015), Create a Manual Chinese Word Segmentation Dataset Using Crowdsourcing Method, in 'Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing ', Association for Computational Linguistics, Beijing, China, pp. 7--14.

Wang, S.; Huang, C.-R.; Yao, Y. & Chan, A. (2014), Building a Semantic Transparency Dataset of Chinese Nominal Compounds: A Practice of Crowdsourcing Methodology, in 'Proceedings of Workshop on Lexical and Grammatical Resources for Language Processing', Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pp. 147–156.

Wang, Shichang, Chu-Ren Huang, Yao Yao and Angel Chan. 2014. Exploring mental lexicon in an efficient and economic way: crowdsourcing for linguistic experiments. Proceedings of the Workshop on Cognitive Aspects of the Lexicon (CogALex-IV), the 25th International Conference on Computational Linguistics. Dublin, Ireland. 24 August.

Relata: Semantic Relations

Santus, Enrico, Anna Gladkova, Stefan Evert, and Alessandro Lenci. "The CogALex-V Shared Task on the Corpus-Based Identification of Semantic Relations." COLING 2016 (2016): 69.

Liu, Hongchao, Karl Neergaard, Enrico Santus and Chu-Ren Huang. 2016.EVALution-MAN: A Chinese Dataset for the Training and Evaluation of DSMs. Proceedings of LREC 2016. 4583-4587. The 10th Language Resources and Evaluation Conference, 23-28 May 2016. Portorož, Slovenia.

Santus, Enrico, Alessandro Lenci, , Tin-Shing Chiu, Qin Lu and Chu-Ren Huang. 2016. Unsupervised Measure of Word Similarity: How to Outperform Co-Occurrence and Vector Cosine in VSMs., in Dale Schuurmans & Michael P. Wellman, ed., Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16). pp. 4260-4261 . AAAI Press,

Santus, Enrico, Alessandro Lenci, , Tin-Shing Chiu, Qin Lu and Chu-Ren Huang. 2016. ROOT13: Spotting Hypernyms, Co-Hyponyms and Randoms. in Dale Schuurmans & Michael P. Wellman, edProceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16). pp. 4262-4263 . AAAI Press.

Liu, Hongchao Liu, Karl Neergaard, Enrico Santus and Chu-Ren Huang. 2016. EVALution-MAN 2.0: 扩展用于训练和评价DSM的数据集. Presented at the 2016 Chinese Lexical Semantica Workshop. 19-21 May, Singapore.

Santus, Enrico, Frances Yung, Alessandro Lenci and Chu-Ren Huang. 2015. EVALution 1.0: an Evolving Semantic Dataset. the Fourth Workshop on Linked Data in Linguistics (LDL-2015), In conjunction with ACL-IJCNLP 2015. July 31, Beijing.

Phonological Neighbourhood Density

Neergaard, Karl and Chu-Ren Huang. 2016. Database of Mandarin Neighborhood Statistics. Proceedings of LREC 2016. Pp. 2270-2277. The 10th Language Resources and Evaluation Conference, 23-28 May 2016. Portorož, Slovenia.

Neergaard, Karl and Chu-Ren Huang. 2016. Model Selection Across Multiple Segmentation Schemas Reveals Mandarin Chinese Phonological Neighborhood Effects. Presented at the 2016 International Meeting of the Psychonmic Society.

Neergaard, Karl and Chu-Ren Huang. 2016. Phonological neighborhood density in a tonal language: Mandarin neighbor generation task. The 90th Annual Meeting of the Linguistic Society of America. 7-9 January. Washington DC.

Neergaard, Karl, Hongzhi Xu, and Chu-Ren Huang. 2015.Database of Word Level Statistics from Multiple Movie Subtitle Corpora. International Conference on Corpus Linguistics and Technological Advancement (CoLTA). 16-18 December. Hong Kong Institute of Education.

Neergaard, Karl, and Chu-Ren Huang. 2015. Graph Theoretic Features of the Adult Mental lexicon Predict Language Production in Mandarin: Clustering Coefficient. PACLIC-29. 30 October-1 November. Shanghai Jiaotung University.