

# Chinese Word Sketches

*Adam Kilgarriff, Chu-Ren Huang, Pavel Rychly, Simon Smith, David Tugwell*

Word sketches are one-page automatic, corpus-based summaries of a word's grammatical and collocational behaviour. They were first used in the production of the Macmillan English Dictionary (2002). At that point, they only existed for English. Now, we have developed the Sketch Engine, a corpus tool which takes as input a corpus of any language and corresponding grammar patterns and which generates word sketches for the words of that language. It also automatically generates a thesaurus and 'sketch differences', which specify similarities and differences between near-

synonyms. We have recently prepared word sketches for Chinese, based on the Chinese Gigaword corpus. The corpus was tokenized and part-of-speech tagged in collaboration with Academia Sinica. We shall demonstrate the Chinese word sketches, and show how they can be used in lexicography, to enrich and improve existing definitions, to quickly add further information on collocates, and for selecting typical illustrative examples, as well as for more ambitious projects involving electronic and web dictionaries and translation aids. We shall also describe the software and principles behind it, differentiating it from other corpus query systems such as Wordsmith and the Stuttgart tools. We shall outline the issues arising in the preparation of the Chinese data.