

Quantitative Analyses of Chinese Poetry of Tang and Song Dynasties: Using Changing Colors and Innovative Terms as Examples

Chao-Lin Liu, National Chengchi University (Taiwan), chaolin@nccu.edu.tw

**This file is a supplement for the paper, with the above title, that appeared in the
*Proceedings of the 2016 International Conference on Digital Humanities.***

We analyzed the collocation of Tang poetry at two different levels. Ordinary collocation analysis in linguistics considers neighboring words in a given context. The composition of Tang poetry requires a special level of constraint. Words at the corresponding positions of certain paired sentences in a poem must follow some linguistic and even artistic constraints, i.e., “對仗” (dui4 zhang4). Both types of collocation are informative and deserve detailed studies.

Colors convey different meanings in different cultures and at different times. In modern Chinese societies, red color carries the notion of auspiciousness, and white color suggests depletion and sometimes even mournfulness. In Tibet, white is a color of blessing. Hence, studying changing colors in representative collections of poetry like QTS and QSC can lead to interesting discoveries.

“白髮” (bai2 fa3, gray hair), “白頭” (bai2 tou2, white head), and “白首” (bai2 shou3, white head) are three of the most frequent white words for describing depressive situations in QTS. “紅塵” (hong2 chen2, the world [Buddhism]) and “殘紅” (can2 hong2, remaining red) are the most frequent red words in QSC, and both carry non-positive sentiment in the poetry.

The proportions of these five words appeared in poetry in QTS are, respectively, 1.21%, 0.97%, 0.61%, 0.35%, and 0.09%. The proportions changed to 1.06%, 0.45%, 0.30%, 0.95%, and 0.59%, respectively, in QSC. While “白髮” remained to be the leader, the proportions of “紅塵” and “殘紅” burgeoned to exceed other white words. The changing ranks of these words suggest that more thorough reading and comparison of the poetry which used these colored words should be rewarding.

We did not claim that we solved the WSD challenge. Using contextual information like word co-occurrences is minimal, but reading cannot be avoided.

We wish to discuss more technical details and more actual examples when there is no word limit during the conference.