

Bridging corpus and norm: Mandarin sensory adjectival phrases

Thomas Van Hoey¹, Xiaoyu Yu², Tung-Le Pan³ & Youngah Do⁴

¹KU Leuven, thomas_van_hoey@outlook.com

²University of Hong Kong, xyu97@connect.hku.hk

³Independent scholar, happypan2466@gmail.com

⁴University of Hong Kong, youngah@hku.hk

Keywords: ABB, Chinese, ideophones, norming studies

The well-studied Mandarin ABB construction (T'sou 1978; Cáo 1995; Wang 2014) consists of a content word followed by a reduplicated syllable that makes the whole more vivid (Huang, Jin & Shi 2016), e.g., wù-mángmáng 霧茫茫 'fog-unclear.unclear' "extremely foggy". Because of this vivid quality, ABB constructions have been argued to consist of ideophones (BB) preceded by prosaic collocates (A) (Lu 2011; Van Hoey 2020; 2023; Zhào 2021).

In a corpus-driven study (Van Hoey, 2023), it is concluded that ABB phrases in Mandarin are not a homogeneous group, but are prototypically structured, depending on which lexical statistic is considered, e.g., type and token frequency, dispersion across the corpus, and attraction and repulsion between collocate (A) and ideophone. However, while these different degrees in salience may indeed be observed in the corpus, it is not clear whether these different degrees of salience are also present in Chinese speakers' minds. Can we really rely on findings based on textual data alone, or can such models be improved by turning to behavioral data as well? Previous studies (Dąbrowska 2016; Klavan & Divjak 2016) have argued for cross-validation between data obtained from the corpus and experiments. In other words, we need to study ABB constructions from multiple perspectives to arrive at a comprehensive understanding of their usage and structuring across the lexicon.

To complement the corpus-based perspective of ABB words (Van Hoey 2023), our experiments examined different sets of decontextualized subjective rating variables, namely familiarity, valence, imagery, concreteness, sensory experience ratings (SER), iconicity and arousal (e.g., see Yao et al. 2016). Participants consisted of native speakers of Mandarin Chinese (n = 519 in total, on average n = 74 per variable). Each stimulus was rated at least 30 times. Imagery, concreteness, SER and familiarity are highly correlated with each other (r > 0.85). As expected, familiarity also highly correlates with token frequency (r = 0.59). However, we found strong inversed correlations between the token frequency and dispersion across genre (r = -0.67), highlighting the lexical specificity of sensory phrases like ABB, i.e., they will be used quite frequently, but only in limited contexts.

Bridging textual and behavioral data, it turns out that dispersion is the strongest negative predictor for SER, iconicity, concreteness, imagery, and familiarity, while token frequency (and not type frequency) is the strongest positive predictor for these subjective variables. Surprisingly, we found only weak (r = 0.12) to insignificant correlations between the two lexicostatistic measures that probed compositional aspects of ABB constructions (attraction and repulsion) on the one hand, and the subjective behavioral variables on the other hand. This indicates that ABB items tend to be processed on the compound or phrasal level rather than on the compositional level, i.e., provides support against the compositional fallacy (Langacker 1987). We also conclude that corpus data and behavioral data do not necessarily lead to the same understanding of this linguistic construction.

Key references

- Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). <https://doi.org/10.1515/flin-2016-0014>.
- T'sou, B.K. 1978. Sound symbolism and some socio- and historical linguistic implications of linguistic diversity in Sino-Tibetan languages. *Cahiers de linguistique - Asie orientale* 3(1). 67–76. <https://doi.org/10.3406/clao.1978.1039>.
- Van Hoey, Thomas. In print. ABB, a salient prototype of collocate–ideophone constructions in Mandarin Chinese. *Cognitive Linguistics*. <https://doi.org/10.1515/cog-2022-0031>.
- Yao, Zhao, Jia Wu, Yanyan Zhang & Zhenhong Wang. 2016. Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods* 49(4). 1374–1385. <https://doi.org/10.3758/s13428-016-0793-2>.