What ratings and corpus data reveal about the vividness of Mandarin ABB words

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) (Zhào 2021; Van Hoey 2023).

Reviewing the traditional literature on ABB words, it is clear that they display prototypicality effects from a definitional point of view: there are clear clusters that share family resemblances and the identification of necessary and sufficient conditions for ABB words is not easy. However, it is unclear whether ABB words also show prototypicality effects from the perspective of exemplars, which requires an investigation beyond traditional study of dictionaries and lists in dedicated studies.

To remedy this gap, we combined data from behavioral ratings and corpus work. We collected subjective ratings for 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.

Next, we augmented a previous corpus study on collocate-ideophone constructions (Van Hoey 2023) and obtained lexicostatistical measures about ABB words in written corpora, a social media corpus (van Esch 2012), and in dictionaries and lists; their frequency and dispersion across these data sets; and cue validity on a structural level between the components of A and BB. Here we find higher positive correlations between (r > 0.5) between the different measures, except for the ones involving cue validity (r = 0.1), indicating that this may operate on a different plane than the rest.

In the converging step, we combined rating data and corpus data through principal components analysis (PCA), to explore the different ways in which exemplars of ABB words display prototypicality effects. They indeed do not form a homogeneous group but have clear differences in membership salience and different degrees of typicality. We identify three main clusters to which ABB words can belong: (1) core prototypical (most familiar to people, highly used), (2) average typical (most items are well-known but not necessarily used), (3) novel (new ABB forms, sporadic usage, false positives etc.).

Finally, we use the rating data to probe the differences between vivid ABB words and "normal" prosaic words. Here, a PCA shows that ABB words differ from prosaic words mostly in terms of high imageability and also in lower than average arousal. The high imageability is related to the quality of vividness that is often ascribed to them. The low arousal values may indicate that a decontextualised rating task does not bring about arousing quality—because ABB words (a form of collocate-ideophone expressions, see Van Hoey 2023) rely on a grounding context to let their arousing qualities shine through.

This research provides a comprehensive investigation into the prototypicality effects of Mandarin ABB words. By combining subjective ratings and corpus data, we explore prototypicality effects of exemplars, rather than limiting ourselves to definitional effects, which are based on lists and dictionary material. We thus advocate for an augmentation of such material with actual usage (corpus) or distributed understanding (ratings).

Key references

Cáo, Ruìfāng 曹瑞芳. 1995. Pǔtōnghuà ABB shì xíngróngcí de dìngliàng fēnxī 普通话 ABB 式形容词的定量分析 [Quantitative analysis of Mandarin Chinese ABB adjectives]. *Yǔwén yánjiū 语文研究* 3. 22–25.

Esch, Daan van. 2012. Leiden Weibo Corpus. Database. http://lwc.daanvanesch.nl/index.php. (26 June, 2016).

Huang, Shi-Zhe, Jing Jin & Dingxu Shi. 2016. Adjectives and adjective phrases. In Chu-Ren Huang & Dingxu Shi (eds.), *A reference grammar of Chinese*, 276–296. Cambridge: Cambridge University Press.

T'sou, Benjamin 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. 2023. ABB, a salient prototype of collocate–ideophone constructions in Mandarin Chinese. *Cognitive Linguistics* 34(1). 133–163. https://doi.org/10.1515/cog-2022-0031.

Wang, Zhijun. 2014. The head of the Chinese adjectives and ABB reduplication. *US-China Foreign Language* 12(5). 349–359.

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.

Zhào, Qīngqīng 赵青青. 2021. Tōnggǎn yǐnyù shìjiǎo de Xiàndài Hànyǔ ABB shì zhuàngtài xíngróngcí 通感隐喻视角的现代汉语 ABB 式状态形容词 (ABB-pattern state adjectives in Mandarin: A study from the perspective of synaesthetic metaphor). *Shìjiè Hànyǔ jiāoxué 世界汉语教学* 35(2). 206–219. https://doi.org/10.13724/j.cnki.ctiw.2021.02.005.