Affective norms for Chinese characters

Thomas Schlatter
Graduate Institute of Linguistics, National Taiwan University, Taiwan
https://doi.org/10.36505/ExLing-2021/12/0053/000526

Abstract

A database has been generated which constitutes affective ratings for single Chinese characters, viz. the Affective Norms for Chinese Characters (ANCC). This database enables researchers who study the lexico-semantic properties of Chinese characters to necessitate affective and other psycholinguistic properties in a manner that is independent and unbiased. Previous ratings for single Chinese characters, although extensive, have omitted affective properties such as valence and arousal. These factors are known to significantly influence the performance of participants in word recognition and production tasks. Close examination of the data in this study shows that affective and other psycholinguistic aspects of meaning explain a significant portion of the outcomes of previous experiments conducted on Chinese characters. The database can be accessed via OSF (osf.io/jn538).

Keywords: affective norms, Chinese characters, valence, arousal, dominance

Introduction

The Affective Norms for Chinese Characters (ANCC) is currently the largest available dataset on single Chinese characters. This database was compiled with the goal of providing affective aspects of meaning for an extensive list of single Chinese characters, in addition to already existing dimensions such as concreteness (concrete vs. abstract) and imageability (un-imageable vs. imageable) (Liu et al. 2007). These aspects include the affective meaning aspects of valence (positive vs. negative), arousal (intense vs. calm), dominance (dominating vs. dominated), origin (emotional vs. rational) and significance (important vs. unimportant).

Rather than dealing with words, we deal with the smallest units of meaning, which in the case of Chinese, are comprised of syllables or their visual representation, viz. characters. Chinese characters are a crucial part of the Chinese language that have been shaped by Chinese culture. For example, the Chinese character for "good", 好, is composed of the radicals 女 (woman) and 子 (child), representing the ancient Chinese idea that it is good for a woman to have a child.

Materials and methods

The assessment was conducted using seven Self-Assessment Manikin (SAM) scales adapted from Imbir (2016). In order to make the assessments more easily accessible to the participants, data was collected in the form of an online survey, which could be completed by using either a computer or a mobile phone. Each word was presented in isolation and accompanied by either a combination of valence-arousal-dominance or concreteness-imageability-significance-origin scales. The Chinese characters were separated into groups of 100. Both character order and the affective or psycholinguistic dimensions were randomized. The participants were first introduced to the aim of the study followed by an instruction on how to complete the survey. First, data on age, sex, and education were collected. The SAM scales were then provided, accompanied by a brief explanation of each variable. Examples were avoided to prevent bias.

The choice of the respective Chinese characters was based on frequency (Da 2004). In addition to the 3450 most frequent characters, this study added 149 gendered characters to the dataset, those of which can be found within the word list cited above. These words all contained either a 人 (person) or 女 (woman) radical. Consequently, the final list contained 3599 Chinese characters.

Participants

Each single character in the list was rated by at least 30 participants. The participants were sourced from the Upwork network, using their native tongue (Chinese) and place of origin (China) as defining characteristics for their participation. All participants were given a one-week period to complete the online survey in a self-paced manner, with remuneration provided for their time.

Reliability and data analysis

To assess the reliability of the data, a comparison is made between parts of the dataset in this study and an earlier dataset. This included measurements for the dimensions concreteness and imageability. Correlations between the data in this study and the data collected by Liu et al. (2007) (n=2172) were strong, at 0.6 and 0.71 for imageability and concreteness, respectively.

The two models presented here include the dimensions of the ANCC in addition to the frequency (Da 2004) as the independent variables. The reaction times comprised the dependent variables. Both models were statistically significant for character recognition ($F(1,2244)=1097.194$, $p<0.001$, $R^2= 0.39$) and character naming ($F(1,1951)=690.992$, $p<0.001$, $R^2= 0.33$). The first regression showed that the dimensions of valence, arousal and imageability were all negatively correlated with reaction times, while dominance and significance were positively correlated with reaction times in a character recognition task. The second regression showed that the dimensions of

valence, arousal, significance, imageability were negatively correlated with reaction times, while significance and origin were positively correlated with reaction times in a character naming task. Compared to models that include frequency as the only independent variable, the complete models, including all affective and other psycholinguistic dimensions, were able to explain a further 4% (ΔR^2=0.04 for character recognition) or 5% (ΔR^2=0.05 for character naming) of the variance in the dependent outcomes.

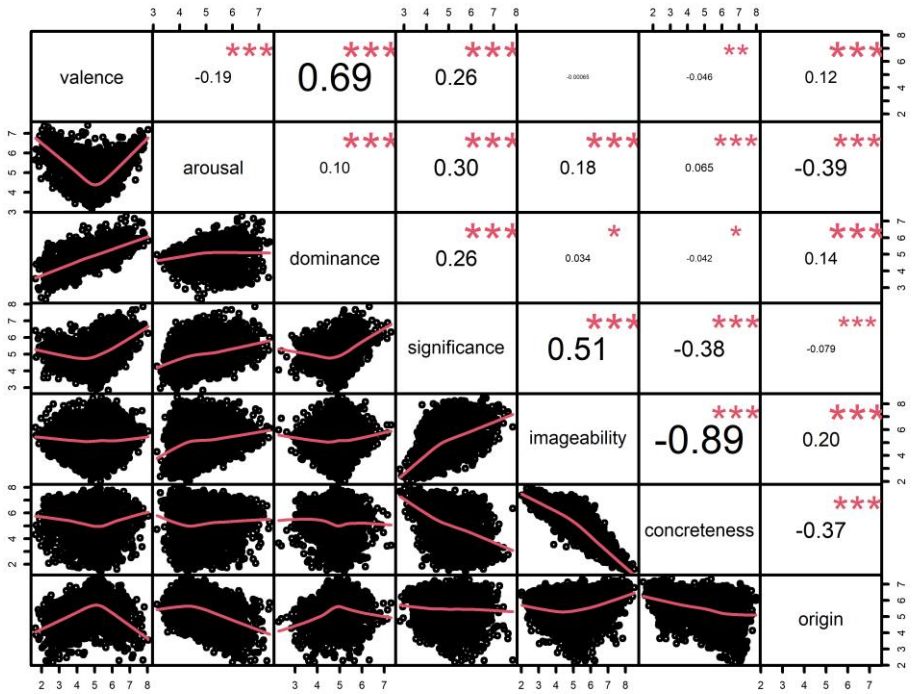

Figure 1. Correlation matrix showing the inter-relationships among the dimensions of the ANCC. The lower triangular matrix is composed by the bivariate scatter plots with a fitted smooth line. The upper triangular matrix shows the Pearson correlation plus significance level. Each significance level is associated to a symbol: p-values 0.001 (***), 0.01 (**), 0.05 (*).

Table 1. Summary outputs of multiple regression analyses with character recognition and naming reaction times as the dependent variables. The IVs were the dimensions of the ANCC, along with frequency. Concreteness was removed from the models for reasons of multicollinearity.

	Recognition. RT (Sze et al. 2014)			Naming.RT (Chang et al. 2016)		
Predictors	*Est.*	*CI*	*p*	*Est.*	*CI*	*p*
(Intercept)	896.74	865.88 – 927.60	<0.001	1041.27	998.31 – 1084.22	<0.001
freq. [log]	-25.01	-26.49 – -23.53	<0.001	-28.05	-30.15 – -25.96	<0.001
valence	-8.32	-11.43 – -5.21	<0.001	-8.76	-13.17 – -4.36	<0.001
arousal	-9.76	-13.43 – -6.10	<0.001	-13.13	-18.34 – -7.91	<0.001
dominance	5.67	0.03 – 11.32	0.049	7.13	-0.84 – 15.09	0.079
significance	8.41	4.45 – 12.36	<0.001	13.14	7.62 – 18.66	<0.001
imageability	-10.23	-12.49 – -7.96	<0.001	-15.65	-18.69 – -12.60	<0.001
origin	0.96	-2.07 – 3.98	0.535	6.28	2.11 – 10.46	0.003
Obs.	2252			1959		
R^2 / R^2 adj.	0.388 / 0.386			0.331 / 0.328		

Conclusions

The collection of emotion and other psycholinguistic norms for 3599 Chinese characters could provide researchers with a corpus that is both independent and unbiased. The ratings presented here were highly correlated with earlier studies and able to explain a significant part of the deviation in both character naming and character recognition studies.

References

Chang, Y.-N., Hsu, C.-H., Tsai, J.-L., Chen, C.-L., Lee, C.-Y. 2016. A Psycholinguistic Database for Traditional Chinese Character Naming. Behavior Research Methods 48 (1): 112–22.

Da, J. 2004. A Corpus-Based Study of Character and Bigram Frequencies in Chinese e-Texts and Its Implications for Chinese Language Instruction. In Proceedings of the Fourth International Conference on New Technologies in Teaching and Learning Chinese, 501–11. Tsinghua University Press, Beijing.

Imbir, K. 2016. Affective Norms for 4900 Polish Words Reload (ANPW_r): Assessments for Valence, Arousal, Dominance, Origin, Significance, Concreteness, Imageability and, Age of Acquisition. Frontiers in Psychology 7: 1081.

Liu, Y., Shu, H., Li, P. 2007. Word Naming and Psycholinguistic Norms: Chinese. Behavior Research Methods 39 (2): 192–98.

Sze, W. P., Liow, S. R., Yap, M. J. 2014. The Chinese Lexicon Project: A Repository of Lexical Decision Behavioral Responses for 2,500 Chinese Characters. Behavior Research Methods 46 (1): 263–73.