

# Taiwan Corpora of Chinese Emotions and Relevant Psychophysiological Data -- A Study on the Norm of Chinese Emotional Words

Shu-Ling Cho,<sup>1</sup> Hsueh-Chih Chen,<sup>2</sup> and Chao-Ming Cheng<sup>3</sup>

<sup>1</sup>Department of Clinical Psychology, Fu Jen Catholic University

<sup>2</sup>Department of Educational Psychology and Counseling, National Taiwan Normal University

<sup>3</sup>Department of Psychology, Fo Guang University

This study attempts to construct a Chinese emotion word database in three stages. It started first with a collection of emotion words. The words were collected mainly from four sources: (1) an analysis of the Sinica Balanced Corpus; (2) emotion words used in theses for master degrees; (3) analyses of Chinese dictionaries by emotion types defined across dimensions of Chinese culture such as Confucianism, Buddhism, and Traditional Chinese Medicine; and (4) free associations of the six basic emotion words from 120 college students. As an emotion word can either describe an emotional state or induce it, all collected words were categorized into emotion-describing and emotion-inducing words. Our analyses yielded a total of 218 emotion-describing words and 395 emotion-inducing words. In the second stage, on a 9-point Likert scale, 353 participants rated each of the emotion-describing words for valence, arousal, continuance, dominance, frequency, and typicality; 1,673 participants rated each of the emotion-inducing words for valence, continuance, dominance and frequency. As significant gender differences were noted in the subjective feeling for certain emotion words, the means and standard deviations of ratings in the database are presented respectively for the entire sample and separately for both genders. The third stage of the study analyzed the two types of words on their clustering in various dimensions. The main findings were: (1) for emotion-describing words, valence had the highest correlations with dominance and frequency, arousal had higher correlations with typicality, but had a small negative correlation with continuance and dominance. Frequency had medium to high positive correlation with all of the dimensions except for arousal, with which no significant correlation was found. (2) Each of all four dimensions rated for emotion-inducing words had medium to high positive correlation with the other three, in which valence and dominance had the strongest correlation. (3) Two methods of K-means cluster analysis (classify-only and classify with-iteration) were used to categorize the emotion words. (a) The first method (classify-only) with the six basic emotion words as centroids revealed that 21.1%, 21.8% of the words had a distance greater than 2 from the centroids. (b) When classified with iteration, 1.8%, 1.5% of the words had a distance greater than 2 from the final centroids. Further, after iteration the six basic emotion words were classified into three to four groups. A possible implication is that we need more research for basic emotional types in Chinese culture.

**Keywords:** *valence, frequency, emotion-describing words, emotion-inducing words, arousal*