

Emotional Significance in Cross-Cultural Semantic Crossmodal Correspondences

Jorge Alvarado (jorge.alvarado@javeriana.edu.co)

Department of Industrial Engineering, Pontificia Universidad Javeriana, Carrera 7 40-62
Bogotá, COLOMBIA

Abstract

Crossmodal correspondences are associations between perceptual features from different senses that aid in crossmodal binding. The semantic coding of these correspondences is expected to capture and mediate the emergence of perceptual crossmodal correspondences. However, the cross-cultural nature of such semantic coding has not been thoroughly studied. This research involved five languages across three different linguistic families (English, Dutch, Turkish, Chinese and Italian). Using distributional semantics, modality exclusivity norms and emotional lexicons, networks were constructed to represent semantic crossmodal correspondences and assess their relationship with Valence, Arousal and Dominance. Results indicate that emotions, particularly Valence and Dominance, play pivotal roles in shaping the structure of semantic crossmodal correspondences networks across languages. Moreover, the findings reveal that some types of semantic crossmodal correspondences might be shared among different languages in various language families, suggesting shared cognitive processes. This supports the significance of emotions as fundamental components in semantic crossmodal correspondences. Additionally, the study provides evidence supporting shared crossmodal correspondences among languages and cultures.

Keywords: crossmodal correspondences; emotion; semantic coding; multilingual.

Introduction

Crossmodal correspondences are systematic cognitive associations between features across different sensory modalities (Spence, 2011). Well-known examples include the association between pitch and spatial elevation (Parise et al., 2014) and between tastes and shapes; for example, people often associate roundness with sweetness (Velasco et al., 2016). Crossmodal correspondences play an important role in crossmodal binding, which is essential for survival and enhances the quality of sensory experiences (Stein et al., 2014). Importantly, crossmodal correspondences relate not to specific objects (e.g., the sound and smell of a dog) but to specific modal features (e.g., a round shape and a sweet taste).

Emotion is one of the proposed mechanisms driving the formation of crossmodal correspondences (Spence & Deroy, 2013). Numerous studies have demonstrated the role of emotional mediation, influencing hedonic and affective responses in associations between different modalities, such as music and color (Spence, 2020a) or tastes and shapes (Velasco et al., 2015). Moreover, emotional valence is

encoded early by sensory and brain receptors, which subsequently influences the identification and naming of emotions (Feinberg & Mallatt, 2016).

Understanding how crossmodal correspondences are encoded in language, and whether and how the emotional mechanisms involved in their formation transfer to this encoding, would enhance our knowledge of the relationships between perception and language, including embodied cognition.

Language differentially encodes the senses in relation to emotion (Speed & Majid, 2020), particularly demonstrating a strong valence component in the semantic representations of taste and smell (Arshamian et al., 2022). The transference of meaning between words from different sensory modalities may be based on evaluative judgments (i.e., emotional valuation) or sensory proximity rather than purely metaphorical descriptions (Winter, 2019b). The encoding of crossmodal correspondences in language is likely constrained by potential sensory hierarchies within a specific language (Reilly et al., 2020) and issues of ineffability (Levinson & Majid, 2014). Despite these limitations, and according to the semantic coding hypothesis (Martino & Marks, 2001), perceptual experience and language form an abstract cognitive crossmodal semantic network of mutual influence, as showed by various researchers (Martino & Marks, 2000; Walker, 2012). Winter (2019a) specifically argues that certain sensory perceptual asymmetries and associations are likely encoded in language due to mechanisms of embodiment and perceptual simulation, indicating that perceptual mechanisms of crossmodal correspondence formation might indeed be transferred to language.

It is important to notice that many of the findings mentioned before about language, emotion and crossmodal correspondences vary across languages and cultures. For example, some crossmodal correspondences differ among languages (Wan et al., 2014); ineffability and the hierarchy of senses are also language-dependent (Majid et al., 2018), and the associations between perceptions and emotions exhibit culture variability, such as with color (Jonaskaite et al., 2020). Investigating these similarities and differences can help us understand to what extent the encoding of crossmodal correspondences in language is shared across cultures and determine if there are pervasive perceptual crossmodal correspondences experienced by segments of the human population.

Distributional semantics produces semantic representations of words as numeric vectors. Such

representations capture language structures by predicting a word based in their surrounding words using neural networks trained in large corpora (Mikolov et al., 2013). Following the linguistic distributional hypothesis (Harris, 1954), these word vectors can measure semantic proximity: the closer the word vectors, the more similar the contexts in which they are used and the more similar their meanings. Additionally, word vectors have been shown to capture, at least partially, perceptual information (Johns & Jones, 2012; Utsumi, 2020). Therefore, word vectors can facilitate the construction of a network of semantic crossmodal relationships between words, using semantic closeness as a metric of relationship between words of different modalities.

Methods

The present research involved eight key steps designed to assess how crossmodal correspondences emerge in language, as proposed by Alvarado et al.(2023)

Word and Modality Selection

Modality exclusivity norms were searched for in various languages, including only those languages with a sizable number of adjectives/object property words, as these words are closest to the sensorial features of crossmodal correspondences. Selected languages, modality exclusivity norm sources, and applied filters are showed in Table 1.

Table 1: Languages and their emotional and modality sources

Language/ Modality norms source	Modifications	Emotional complementary source
Dutch: Speed & Brybaert (2021)	Only adjectives filtered and Maximum perceptual strength >2.5	(Verheyen et al., 2020)
Chinese: Chen et al. (2019)	Only single characters were included	(Zhao et al., 2019) and back translation from VAD lexicon
Turkish: Kumcu (2021)	Only adjectives filtered	Sağlam et al.(2019)
Italian: Morucci et al. (2019)	Only adjectives filtered	(Basile & Nissim, 2013; Montefinese et al., 2014)
English: Lynott & Connell (2009)	All words included	N/A

Dutch norms have the largest number of adjectives, with diverse perceptual semantic content. Fortunately, all words were rated according to the perceptual strength of their semantics for each sense. Therefore, adjectives with high

perceptual strength were filtered out to prioritize perceptual adjectives over other types.

Chinese norms only included single characters words, given that two-character words might already include some crossmodal combinations (Chen et al., 2019). At this stage, for each language, sensorial adjectives were retrieved with their respective dominant modality corresponding to the highest perceived modality of the word. Only the five Aristotelian senses were included, reflecting the modalities present in all norms found.

Modality Labeling and Emotional Values Attachment

Emotional valence, arousal, and dominance values were included for each word in each language, extracted from the NRC VAD multilingual Lexicon (Mohammad, 2018). The VAD lexicon has values from zero (lowest) to one (highest) for each emotional dimension and word. Due to the low number of matched words with the VAD lexicon and its reliance on translations from English, additional sources directly generated from each language that included at least the valence dimension were incorporated. Complementary emotional sources are detailed in Table 1, and their values where standardized to values from zero to one where necessary to make them comparable with the NRC VAD multilingual lexicon. Values of Valence, Arousal, or Dominance were then averaged for words that had more than one emotional source.

Distance Calculations among Crossmodal Words

For each word selected in each language 300-dimensional word vectors were extracted using fastText from extensive texts from Wikipedia and Common Crawl(Grave et al., 2018). Distances were calculated among vectors across each pair of modality word lists in each language (ten possible combinations of the five modalities in each language). Distances were standardized by the mean distance between each word with all the words of the paired modality, in order to detect words of one modality closer to a word of other modality rather than close to all words of the other modality.

Network of Crossmodal Correspondences Extraction

Each standardized distance is a relationship between two words from different modalities. The pairs of words with the lowest scaled distance between each pair combination of senses were selected as candidate crossmodal correspondences. The threshold selected was the proposed by Tukey (1977) for the selection of extreme points in a univariate distribution. Each pair combination of senses had its own threshold, thus controlling for potential effects of the hierarchy of senses. The result is a sparse graph (network) for each language whose vertices are words and whose edges are associations between two words of different modalities with

a scaled cosine distance below the selected threshold for each pair of modalities in each language.

Community Extraction from the Network

In each language network of crossmodal correspondences, communities were detected using Newman's leading eigenvector method (Newman, 2006). This technique divides the complete network in smaller subnetworks of vertices (words) with high density of interconnections among them, called communities.

Communities' Emotional Assessment of Differences

ANOVAs were conducted for each dimension of emotions (Valence, Arousal and Dominance) as dependent variables with communities for each language (only the communities with more than five words were selected) as independent variables. This allows assessing if communities differ in their emotional average values for each language.

Shared Crossmodal Associations

Crossmodal relationships from other languages were translated into English using Google Translate, except for Chinese, whose norms already included an English translation. Only translations whose dominant modalities agreed between English Lancaster Sensorimotor norms (Lynott et al., 2020) and the source language were retained. Shared crossmodal relationships among languages in at least two language families were selected, and a network with relationships shared by two or more languages in more than one language family was created, aiming to reduce bias from the inclusion of three Indo-European languages.

Robustness Assessment

The robustness of community selection was assessed by varying community detection methods. Specifically, Infomap (Rosvall & Bergstrom, 2008) and Louvain (Blondel et al., 2008) were employed as community detection alternatives to Newman's eigenvector method. It was also tested whether emotional differences are maintained when using single sources of emotional dimensions instead of the average of all sources available.

Results

The number of original words for each language, and the number of vertices (words) finally selected with at least one candidate crossmodal correspondence for each language are shown in Table 2, along with the percentage of words that have assigned emotional values in the emotional sources.

The percentage of words with emotional values for Dutch and Italian was low. These are the two largest lexicons of the study. Further analyses showed that words without emotional values have lower frequency in the language for Dutch and Italian. In addition, words without emotional values were not

statistically associated with modality or perceptual strength. Thus, these words likely did not have emotional values due to their low frequency in the languages.

Table 2: Number of selected words and crossmodal correspondences /emotional coverage

Language	Initial words/ selected vertices	# of crossmodal correspondences	% emotion coverage
Dutch	1523/1176	4750	57.3%
Chinese	171/80	110	87.5%
Turkish	258/155	211	85.2%
Italian	643/518	1775	61.2%
English	423/378	1206	92.7%

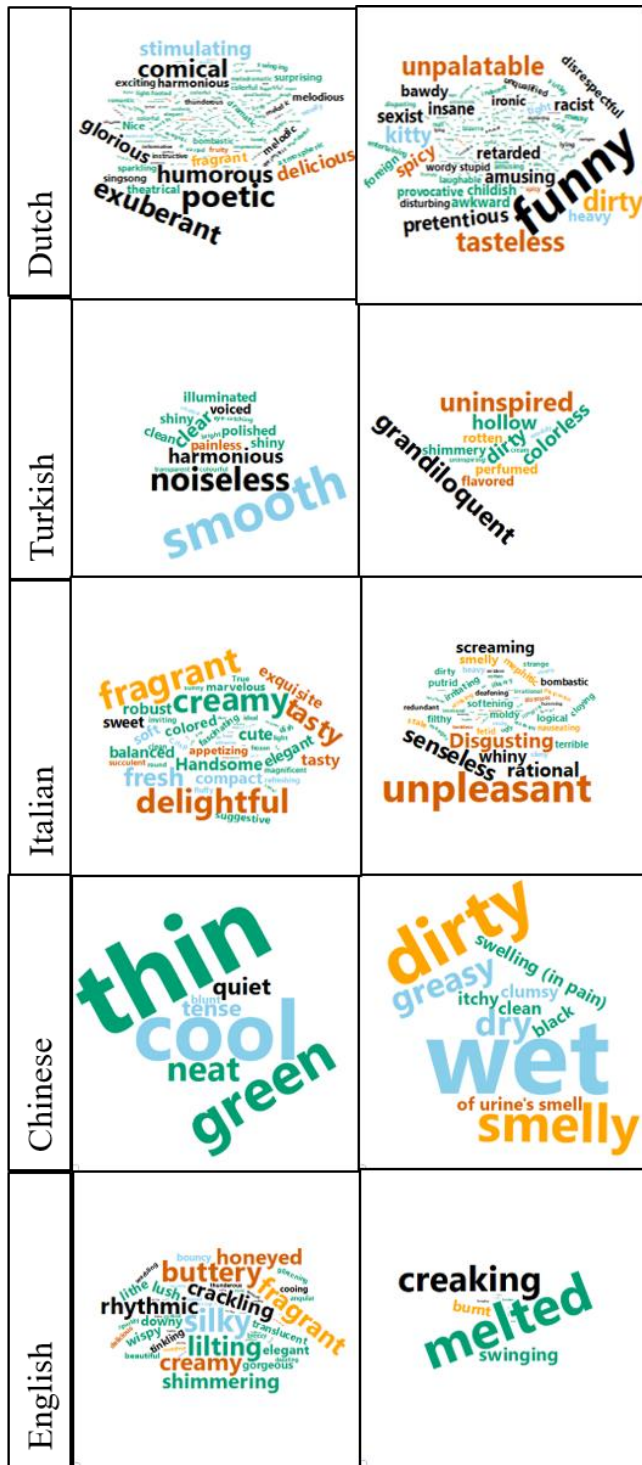
Tests were conducted to determine if selected vertices were more associated with a specific modality than expected. Dutch and Italian had more haptic and auditory words than expected and fewer visual words than expected. The effect was marginal in the same direction for Chinese and English, and non-existent for Turkish. In each language, all of the ten possible relationships had at least one crossmodal correspondence.

Table 3 displays the number of communities in each language, and the effect sizes of Valence, Arousal and Dominance for each one. Figure 1 shows words belonging to the community with the highest and lowest average valence respectively, for each language (translated to English), with different colors representing each modality.

Table 3: Emotional differences (Valence, Arousal and Dominance) across languages.

Language N(n)*	Valence	Arousal	Dominance
Dutch 10(6)	$\eta^2=.27$ $F(5,664)=$ 49.55	$\eta^2=.09$ $F(5,664)=$ 12.73	$\eta^2=.23$ $F(5,559)=$ 32.66
Chinese 16(5)	$\eta^2=.33$ $F(4,42)=$ 5.126	$F(4,42)=$ 0.345	$\eta^2=.35$ $F(4,41)=$ 5.481
Turkish 11(8)	$\eta^2=.27$ $F(7,108)=$ 5.606	$F(7,66)=$ 0.419	$\eta^2=.26$ $F(7,66)=$ 3.364
Italian 17(9)	$\eta^2=.28$ $F(8,336)=$ 16.45	$F(8,295)=$ 2.119	$\eta^2=.17$ $F(8,295)=$ 7.807
English 13(8)	$\eta^2=.18$ $F(7,334)=$ 10.41	$\eta^2=.06$ $F(7,334)=$ 3.11	$\eta^2=.07$ $F(7,334)=$ 3.72

*N= number of communities; n=communities > 5 words
Bold italics are significant associations (p<.01)



- Auditory
- Gustatory
- Haptic
- Olfactory
- Visual

Figure 1- Words of the highest and lowest average valence community in each language.

Communities extracted with alternative methods (i.e., Infomap and Louvain) showed a moderate purity metric of agreement with original communities, ranging from .28 to .64. Notably, languages with lower number of words (Turkey and Chinese) had lower purity results.

Results indicate a strong difference in valence across the communities of words in different modalities in all languages. Dominance also show significant differences across all languages. However, this might be attributed to the strong correlation usually found between valence and dominance. In contrast, evidence for arousal differences is unclear and weak. Nonetheless, it is clear that emotional components such as valence and dominance are crucial in the formation of networks of semantic crossmodal correspondences, some of which may be coded from the emotional formation mechanism of perceptual crossmodal correspondences. These results are robust even when the data sources of emotional values are evaluated independently and different community detection methods are applied.

Although there is a larger presence of visual words in the depicted communities in Figure 1, the crossmodal nature of the communities is evident. The larger presence of visual words, along with the differences in the number of words, is explained by the imbalance in the original datasets.

From an initial dataset of 3247 potential crossmodal relationships across all languages, only 106 were shared by more than one language. From these, only 25 crossmodal correspondences were shared across two or more language families and are depicted in Figure 2. Three languages, as indicated by the numbers on the edges, share seven out of the 25 relationships.

At least three clear subnetworks emerge in Figure 2. In the lower right, a network related to weather and outdoors is visible, primarily joining haptic and visual modalities (*sunny-warm-misty*). These results underscore the importance of weather and outdoors in crossmodal correspondences in different cultures, with some previous influences found in eating experiences in Norway and Colombia (Tran et al., 2023) and color-emotion relationships globally (Jonaskaite et al., 2020). In this evidence, weather appears as a strong candidate to be a shared cognitive source of crossmodal correspondences in different languages and cultures.

At the upper left, a hedonic network of positive-valenced words across all senses emerge. Although the main line of connections is mostly visual-haptic, some key language terms appear in the other senses (*pure, melodic, and spicy* conceding that *melodic* and *melodious* are synonyms). This network is a strong candidate as a multicultural cognitive mechanism of crossmodal correspondences, clearly linked with emotion, particularly with the brain rewards system (Kühn & Gallinat, 2012) that connect pleasures such as food (*spicy*) and music (*melodious*) (Spence, 2019).

- associations are further shaped by linguistic and geographic proximity. *Psychological Science*, 31(10), 1245-1260.
- Kumcu, A. (2021). Linguistic synesthesia in Turkish: A corpus-based study of crossmodal directionality. *Metaphor and Symbol*, 36(4), 241-255.
- Kühn, S., & Gallinat, J. (2012). The neural correlates of subjective pleasantness. *Neuroimage*, 61(1), 289-294.
- Levinson, S. C., & Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, 29(4), 407-427.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties [Article]. *Behavior Research Methods*, 41(2), 558-564. <https://doi.org/10.3758/BRM.41.2.558>
- Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, 52, 1271-1291.
- Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., . . . Levinson, S. C. (2018). Differential coding of perception in the world's languages [Article]. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45), 11369-11376. <https://doi.org/10.1073/pnas.1720419115>
- Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence. *Perception*, 29(6), 745-754.
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and Weak. *Current Directions in Psychological Science*, 10(2), 61-65. <https://doi.org/10.1111/1467-8721.00116>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings,
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior research methods*, 46, 887-903.
- Morucci, P., Bottini, R., & Crepaldi, D. (2019). Augmented modality exclusivity norms for concrete and abstract Italian property words. *Journal of Cognition*, 2(1).
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3), 036104.
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing [Article]. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 6104-6108. <https://doi.org/10.1073/pnas.1322705111>
- Reilly, J., Flurie, M., & Peelle, J. E. (2020). The English lexicon mirrors functional brain activation for a sensory hierarchy dominated by vision and audition: Point-counterpoint [Article]. *Journal of Neurolinguistics*, 55, Article 100895. <https://doi.org/10.1016/j.jneuroling.2020.100895>
- Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the national academy of sciences*, 105(4), 1118-1123.
- Saluja, S., & Stevenson, R. J. (2018). Cross-modal associations between real tastes and colors. *Chemical Senses*, 43(7), 475-480.
- SAGLAM, F., Genc, B., & Sever, H. (2019). Extending a sentiment lexicon with synonym--antonym datasets: SWNetTR++. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(3), 1806-1820.
- Speed, L. J., & Brybaert, M. (2021). Dutch sensory modality norms. *Behavior research methods*, 1-13.
- Speed, L. J., & Majid, A. (2020). Grounding language in the neglected senses of touch, taste, and smell [Article]. *Cognitive Neuropsychology*, 37(5-6), 363-392. <https://doi.org/10.1080/02643294.2019.1623188>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review [Review]. *Attention, Perception, and Psychophysics*, 73(4), 971-995. <https://doi.org/10.3758/s13414-010-0073-7>
- Spence, C. (2019). Multisensory experiential wine marketing [Review]. *Food Quality and Preference*, 71, 106-116. <https://doi.org/10.1016/j.foodqual.2018.06.010>
- Spence, C. (2020a). Assessing the Role of Emotional Mediation in Explaining Crossmodal Correspondences Involving Musical Stimuli [Article]. *Multisensory Research*, 33(1), 1-29. <https://doi.org/10.1163/22134808-20191469>
- Spence, C. (2020b). Simple and complex crossmodal correspondences involving audition [Article]. *Acoustical Science and Technology*, 41(1), 6-12. <https://doi.org/10.1250/ast.41.6>
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and Cognition*, 22(1), 245-260.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15(8), 520-535. <https://doi.org/10.1038/nrn3742>
- Tran, H., Veflen, N., Reinoso-Carvalho, F., Tabassum, F., & Velasco, C. (2023). Seasonal multisensory eating experiences in Norway and Colombia. *Food Quality and Preference*, 109, 104873.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Utsumi, A. (2020). Exploring what is encoded in distributional word vectors: A neurobiologically motivated analysis. *Cognitive Science*, 44(6), e12844.

- Velasco, C., Woods, A. T., Deroy, O., & Spence, C. (2015). Hedonic mediation of the crossmodal correspondence between taste and shape [Article]. *Food Quality and Preference*, *41*, 151-158. <https://doi.org/10.1016/j.foodqual.2014.11.010>
- Velasco, C., Woods, A. T., Marks, L. E., Cheok, A. D., & Spence, C. (2016). The semantic basis of taste-shape associations [Article]. *PeerJ*, *2016*(2), Article 1644. <https://doi.org/10.7717/peerj.1644>
- Verheyen, S., De Deyne, S., Linsen, S., & Storms, G. (2020). Lexicosemantic, affective, and distributional norms for 1,000 Dutch adjectives. *Behavior research methods*, *52*, 1108-1121.
- Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics*, *74*, 1792-1809.
- Wan, X., Woods, A. T., Van Den Bosch, J. J. F., McKenzie, K. J., Velasco, C., & Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between basic tastes and visual features [Article]. *Frontiers in Psychology*, *5*(DEC), Article 1365. <https://doi.org/10.3389/fpsyg.2014.01365>
- Winter, B. (2019a). Sensory linguistics. *Sensory Linguistics*, 1-303.
- Winter, B. (2019b). Synaesthetic metaphors are neither synaesthetic nor metaphorical. *Perception metaphors*, *19*, 105-126.
- Yiannakas, A., & Rosenblum, K. (2017). The insula and taste learning. *Frontiers in Molecular Neuroscience*, *10*, 335.
- Zhao, J.-L., Li, M.-Z., Yao, J., & Qin, G.-H. (2019). The development of the Chinese sentiment lexicon for Internet. *Frontiers in Psychology*, *10*, 2473.