


ARTICLE

# The organization of semantic associations between senses in language

Jorge A. Alvarado<sup>1</sup> , Carlos Velasco<sup>2</sup> and Alejandro Salgado<sup>3</sup>

<sup>1</sup>Department of Industrial Engineering, Pontificia Universidad Javeriana, Bogotá, Colombia; <sup>2</sup>Department of Marketing, Centre for Multisensory Marketing, BI Norwegian Business School, Oslo, Norway and <sup>3</sup>Centro de estudios superiores de administración, CESA, Bogotá, Colombia

**Corresponding author:** Jorge A. Alvarado; Email: [jorge.alvarado@javeriana.edu.co](mailto:jorge.alvarado@javeriana.edu.co)

(Received 28 March 2023; Revised 23 February 2024; Accepted 26 February 2024)

## Abstract

Distributional semantic representations were used to investigate crossmodal correspondences within language, offering a comprehensive analysis of how sensory experiences interconnect in linguistic constructs. By computing semantic proximity between words from different sensory modalities, a crossmodal semantic network was constructed, providing a general view of crossmodal correspondences in the English language. Community detection techniques were applied to unveil domains of experience where crossmodal correspondences were likely to manifest, while also considering the role of affective dimensions in shaping these domains. The study revealed the existence of an architecture of structured domains of experience in language, whereby crossmodal correspondences are deeply embedded. The present research highlights the roles of emotion and statistical associations in the organization of sensory concepts across modalities in language. The domains identified, including food, the body, the physical world and emotions/values, underscored the intricate interplay between the senses, emotion and semantic patterns. These findings align with the embodied lexicon hypothesis and the semantic coding hypothesis, emphasizing the capacity of language to capture and reflect crossmodal correspondences' emotional and perceptual subtleties in the form of networks, while also revealing opportunities for further perceptual research on crossmodal correspondences and multi-sensory integration.

**Keywords:** computational linguistics; concept representation; crossmodal correspondences; distributional semantic representations; emotion; multisensory integration; network science; semantic domains

## 1. Introduction

Crossmodal correspondences involve associations between features across sensory modalities, encompassing various multisensory signals in everyday life experiences (Motoki et al., 2023; Spence, 2011). Well-known examples include the relationship

between pitch and spatial elevation (Parise et al., 2014; Zeljko et al., 2019) and tastes and shapes, for instance, sweetness and roundness (Velasco et al., 2016b). They emerge in part to help with crossmodal binding, and their value resides on helping survival and enhancing the quality of life experience (Stein et al., 2014). However, little has been studied about how these correspondences are encoded in language and how language encoding is related to perceptual mechanisms of crossmodal correspondences. Knowing this allows for a better understanding of the relationship between perception and language and offers new insights into the functioning of crossmodal correspondences. Specifically, it could help us comprehend how associations in crossmodal phenomena are encoded in language, and if the cognitive mechanisms of crossmodal correspondences are transferred to language coding.

### 1.1. How language encodes perception

In a general sense, language encodes perception. Embodied cognition research has shown that perception can shape cognition, and therefore, language (Glenberg et al., 2013). Neuroimaging results also support the connections between perception, motor skills, specific sensory brain regions and linguistic concepts (Kiefer et al., 2008; Kuhnke et al., 2020). Regarding the language encoding of perception, there are some fundamental elements to consider. First, there is a differential language hierarchy of the senses across cultures (i.e. Western cultures systematically have a larger lexicon for hearing and vision, Reilly et al., 2020; Winter et al., 2018). Secondly, certain cultures privilege different senses in their coding (Majid et al., 2018). Additionally, the coding of the senses in language is far from perfect. Languages have a lack of capacity to completely capture the perceived reality, a phenomenon known as ineffability (Levinson & Majid, 2014). However, it is expected that sensory language coding resembles at least partially its perceptual counterpart (Marks, 1996; Speed et al., 2015). The embodied lexicon hypothesis (Winter, 2019) specifically postulates that sensory perceptual asymmetries (how the senses are prioritized) and sensory perceptual associations are encoded in language, as a natural consequence of embodiment and perceptual simulation. In summary, sensory perception is expected to be encoded in language, including associations between the senses (e.g. crossmodal correspondences), with some asymmetries due to the hierarchy of the senses, and with certain limitations that would arise, at least, from the ineffability of language.

### 1.2. The semantic coding hypothesis

More particularly, there is evidence showing that crossmodal correspondences can also be encoded in language and affect perceptual responses.

Some words have meanings in more than one sensory modality. For instance, *high* and *low* are words with meaning in both the auditory (pitch) and the visual senses (spatial location) in the English language. Simultaneously, there is a well-known perceptual crossmodal association between pitch and spatial location (Spence, 2011). Therefore, some words with meaning in more than one sensorial modality are probably encoding perceptual crossmodal correspondences.

When words replace perceptual stimuli in interference perceptual tasks, similar results are obtained. For instance, subjects presented with sharpness words (*sharp/blunt*) or brightness words (*bright/dull*) and an irrelevant high/low pitched tone,

classified the words faster when the pitch was congruent than when the pitch was incongruent (Walker & Smith, 1984), supporting the idea that semantic encoding of perceptions further influence perceptual tasks.

Moreover, words do not need to be an exact description of the perceptual stimuli if they are semantically related to the stimuli. The speed of classification of a high/low pitched tone was interfered with when congruent/incongruent words like *white/black* were presented, but also when the original words were replaced by the words *day/night* (Martino & Marks, 1999). Subjects presented words within an irrelevant shape outline (angular/curved) classified congruent words faster than incongruent words that were intended to mean hardness (*granite/fur*), pitch (*squeak/drone*) or brightness (*glisten/gloom*) (Walker, 2012). These studies suggest that perceptual encoding of crossmodal correspondences does not occur solely in isolated words but extends to a network of semantically related words.

Perceptual results vary when different languages encode crossmodal correspondences differently. For instance, people using languages with different semantic encoding for pitch such as Dutch (*high/low*) and Farsi (*thin/thick*) perform differently when asked to reproduce the pitch with congruent/incongruent visual stimuli that match/mismatch their language encoding (Dolscheid et al., 2013). Interestingly, after training, Dutch speakers learned to describe pitch in the terms of Farsi language, and their performance in the task started to resemble that of the native speakers of Farsi. Such evidence supports the influence of linguistic input on semantic encoding of perception.

The aforementioned evidence seems to support the semantic coding hypothesis. This hypothesis posits that perceptual experiences from various modalities, along with the language used to describe these perceptions, serve as inputs to produce an abstract (high cognitive level) semantic network that encodes and influences crossmodal correspondences (Martino & Marks, 2001; Melara, 1989).

Two claims derived from the semantic coding hypothesis are important for the present work. First, if there is a clear connection between crossmodal correspondences and their encoding in language, this implies that it is possible for the mechanisms of crossmodal correspondences formation to be transferred to their encoding in language. Second, the encoding of crossmodal correspondences happens in a semantic network where perceptual related concepts are linked accordingly with their relationships and matchings across modalities.

### 1.3. Formation mechanisms of crossmodal correspondences

If there is a clear connection between crossmodal correspondences and their encoding in language, this implies that it is possible for the mechanisms of crossmodal correspondences formation to be transferred to their coding in language (e.g. Saluja & Stevenson, 2018). According to Spence (2011), at least four cognitive mechanisms play a role in crossmodal correspondences. First, structural similarities in the brain might connect the senses; for instance, intensity might be equally represented in the brain as neural firing regardless of the senses involved. Second, statistical co-occurrence in the environment leads to the formation of crossmodal correspondences. For example, pitch and location are associated because higher pitches in natural environments are more frequently produced in higher locations (Parise et al., 2014). Third, emotion appears to mediate certain correspondences, such as between music and

color (Spence, 2020a). Finally, a semantic (or lexical) mechanism has been proposed (Walker, 2016), which is encompassed by the semantic coding hypothesis.

In the present research, we studied how crossmodal correspondences emerge in language, following some of the first three mechanisms presented above. Structural similarities are expected to be innate, and therefore, to be formed before language. Although some of them might be posteriorly encoded in language, given the linguistic nature of our research, it is not possible for us to clearly separate or detect them. On the other hand, we expected to find statistical crossmodal correspondences in language. These are correlations of sensory signals that generate crossmodal correspondences emerging from everyday life experiences, whether natural (Parise, 2016) or cultural, and humans probably use language to refer to these correlated experiences. However, not all kinds of experiences influence, with the same magnitude, the development and elicitation of crossmodal correspondences. We anticipated that a crossmodal semantic network might reveal semantic domains related to everyday life experiences (hereafter, domains of experience) in which crossmodal correspondences are highly prevalent.

#### 1.4. *The role of affect on crossmodal correspondences*

We also expected that affect would be an essential cohesive element to strengthen crossmodal relationships in crossmodal semantic networks. Affect can be conceptualized through the affective domains identified in the literature of valence, potency or arousal and activity or dominance (see also Bakker et al., 2014). These affective domains were constructed based on the semantic differential technique (see Osgood, 1964, for some seminal work on this) and the research presented by Russell and Mehrabian (1977). In their work, Mehrabian and Russell (1974) conceptualized valence as a continuum of responses spanning from positive to negative, which they measured employing polar adjectives such as *joyful* or *unhappy*, while they conceptualized arousal as a spectrum of mental states stretching from low to high excitement, captured by terms like *stimulated-relaxed*. Finally, they associated dominance with notions of control and individual behavioral constraints, reflected in words such as *brave-discouraged*.

Senses and emotion are frequently connected by language and semantics. Previous research has shown that, for instance, there is evidence that valence may be an important part of the semantic representation of taste and smell (Arshamian et al., 2022; Majid et al., 2018; Speed & Majid, 2020). Some semantically related lines and shapes might be easily connected to emotions (Salgado-Montejo et al., 2017). Moreover, there is increasing evidence that sensory receptors and neural circuits code sensory information in terms of emotional valence and that this initial system may be involved in posterior complex cognition such as in experiencing and naming emotions (Feinberg & Mallatt, 2016; Kryklywy et al., 2020).

Recent literature, building on earlier work by researchers such as Kenneth (1923), suggests that hedonics and, more broadly affect (Spence & Deroy, 2013; Velasco et al., 2015) play a role in the mediation of crossmodal correspondences. In this case, perceptual experiences in different contexts that were tied to the same emotion also show crossmodal associations. Two good examples are the emotional mediation between shapes and tastes (Velasco et al., 2015) and the mediation between colors and music (Hauck et al., 2022; Palmer et al., 2013; Spence, 2020a). Our expectation

was that the affective dimensions of crossmodal associations would be visible in the crossmodal semantic network.

### 1.5. *The encoding of perceptual meaning*

How can the language encoding of perception be represented in an accessible and operable way? One possibility is constructing modality norms, where people rate words according to the strength of elicitation of each sense associated with the word. An example of this is the work of Lynott and Connell (2009), which is detailed in Section 2.1. Representations based on human ratings might include other aspects of meaning beyond perception. For instance, Binder et al. (2016) created an encoding of meaning for 535 words using 63 cognitive features neurobiologically motivated and organized in 14 categories, including the senses separated as categories.

An alternative possibility is to base meaning encoding on distributional semantics, i.e. leveraging statistical patterns of language usage to construct a mathematical artifact that convey the meaning of words. Indeed, distributional semantics capture semantic association proximity based on the linguistic distributional hypothesis, which states that the semantic proximity between two linguistic expressions is a function of the similarity of the linguistic contexts where the expressions appear (Harris, 1954).

Distributional semantics do not rule out embodied cognition in meaning; in fact, they can capture and explain it (Louwerse, 2011). For instance, when people had to switch perceptual modalities in a property verification task, processing speed was slower than when they stayed in the same modality (Pecher et al., 2003). Accordingly, latent semantic analysis distances (a distributional semantics technique) are closer when words are from the same modality (Louwerse, 2011). Johns and Jones (2012) tested that words form the same modality but otherwise not related (typewriter, piano) are also closer when comparing first-order co-occurrences statistics. That evidence suggests that information on perception modalities and their differences is captured in distributional semantics approaches.

Based on latent semantic analysis, it was possible to replicate maps and distances between cities with some accuracy (Louwerse & Zwaan, 2009). Such a result has been replicated for Word Embeddings (Konkol et al., 2017) and large language models (Gurnee & Tegmark, 2023), showing that even information not primarily linguistic can, anyway, be encoded in distributional semantic approaches.

Louwerse and Connell (2011) showed that the modality of words based on human ratings can be predicted from first order co-occurrences. Utsumi (2020) tried to predict Binder et al. (2016) meaning representations using multiple distributional semantic approaches based on Word Embeddings. Although results showed an overall higher predictability for abstract knowledge, results for concrete and perceptual knowledge were good enough to suggest that perceptual knowledge is likely encoded in Word Embeddings more than expected. In summary, perceptual meaning can be inferred, at least in part, from distributional semantic approaches.

Word Embeddings are a distributional semantic approach that produce vector encodings of words. Such vectors are the result of training neural network architectures on word co-occurrences in large text corpora (Mikolov et al., 2013). Therefore, the closer the word vectors, the more similar are the contexts where these words are

used and more closely are the semantics of the words, revealing an association between them.

Word Embeddings have been proven to capture diverse types of semantic relationships, including topic relatedness (Das et al., 2015), semantic hierarchies such as hypernymy and hyponymy (Fu et al., 2014), context-dependent world knowledge (Grand et al., 2022) and even biases present in culture and language, or across languages (Bolukbasi et al., 2016; Lewis & Lupyán, 2020). Importantly, Word Embeddings have been proved to capture emotional information (Passaro et al., 2017). As a result, Word Embeddings allow combining perceptual, emotional and topical meaning in a single encoding.

### 1.6. *The present work*

In the present research, we used Word Embeddings to study how crossmodal correspondences emerge in language. If semantic proximity is calculated between words of different senses, we expected the semantic encoding of crossmodal correspondences in such proximity. We evaluated the closeness of semantic associations of words across senses to create a crossmodal semantic network as it appears in the English language. Then, we developed mechanisms to explore the network, namely, seeking domains of experience where semantic crossmodal correspondences are more likely to emerge through community detection, and assessing the role of affective functions (particularly, valence, arousal and dominance) of semantic crossmodal correspondences in these domains of experience.

The contributions of the present work are twofold. First, we evaluate a prediction of the embodied lexicon hypothesis, specifically, that associations between senses (in the context of this research, crossmodal correspondences) are encoded in the lexicon. Second, drawing upon the semantic coding hypothesis, we assess whether the examination of an abstract semantic network of crossmodal correspondences will reveal the statistical and emotional mechanisms underlying the formation of crossmodal correspondences.

## 2. Methods

The present research involved eight key steps designed to assess how crossmodal correspondences emerge in language. First, we selected words and their modalities based on sensory adjectives, focusing on object properties. Second, we retrieved 300-dimensional Word Embeddings for these words and calculated distances between embeddings of different sensory modalities. Third, we generated a crossmodal correspondence network by identifying closely related word pairs based on scaled cosine distances. Fourth, we detected communities within this network using Newman's leading eigenvector method (Newman, 2006). Fifth, we assessed the robustness of community selection by varying thresholds and community detection methods. Sixth, we identified dominant sensory modalities across communities using Cramer's V and chi-square statistics. Seventh, we matched words with emotional valence, arousal and dominance values obtained from the NRC VAD Lexicon (Mohammad, 2018) and compared these values across communities. Finally, we identified domains of experience for each community based on semantic domains from the Intercontinental Dictionary Series (IDS) (Key & Comrie, 2021) and the Summer Institute of

Linguistics (SIL) (Moe, 2003) classifications, as well as other factors such as emotional attributes and modalities represented within the community. Below, we elaborate on the details associated with each step.

### 2.1. Selection of words and their modalities

We used the list of 423 object properties from Lynott and Connell (2009), extensively employed in various psycholinguistic analyses (Connell & Lynott, 2012, 2014). Lynott and Connell (2009) extracted a list of sensory adjectives describing object properties in the English language from diverse sources, such as dictionaries and thesauruses, with the aim of a wide coverage of the senses. They established modality exclusivity norms for these adjectives based on the average rating given for each word in each sensory modality on a scale from zero to five. For each word, the modality that, on average, received the highest rating was selected as the dominant modality. Additionally, a modality exclusivity index was obtained by dividing the range of average ratings across the senses by the sum of these ratings. For instance, *harsh* is a word with an auditory dominant modality and a low modality exclusivity rating (0.12) due to its potential meaning in other senses, whereas *deafening* is also auditory dominant but much more exclusive (modality exclusivity rating: 0.77).

Two critical reasons supported the selection of Lynott and Connell exclusivity norms (2009) over more recent alternatives such as the affective ratings of the Lancaster sensorimotor norms (Lynott et al., 2020). First, Lynott and Connell norms (2009) focus on object properties, which is relevant given that we are specifically exploring associations between sensory features. These associations are better represented by the sensory properties of objects rather than by the extensive list of commonly used lemmas in English across all syntactic categories within the Lancaster sensorimotor norms. For instance, using the Lancaster sensorimotor norms, we might find an association between *dog* (higher in its visual rating) and *bark* (higher in its auditory rating). Such association is known as crossmodal semantic congruence (Laurienti et al., 2004), because it emerges from a shared identity or meaning, not from crossmodally corresponding features as those involved in crossmodal correspondences (Knöferle et al., 2016). Second, the object properties by Lynott and Connell (2009) were specifically chosen for their potential to capture the semantics of the senses. Consequently, Lynott's norms exhibit greater modality exclusivity compared to norms based on nouns (Lynott & Connell, 2013) or those using randomly selected word sets (Winter, 2019, Ch. 12). Greater modality exclusivity implies that this set of words is less experienced in multiple senses simultaneously, reducing the influence exerted by the multisensory nature of the words on the relationships found between different senses.

A potential limitation is that the selected norms are constrained to the Aristotelian senses (i.e. sight, hearing, touch, smell and taste). It is worth noting that our senses are not limited to these traditional five (Velasco & Obrist, 2020). However, we focused on these considering the available linguistic tools to approach them in text analysis, their frequent use in everyday scenarios and their utility as a basic model for research (Winter, 2019, Ch. 2).

The result of this first stage of the research is the list of words, their dominant modality and their modality exclusivity index.

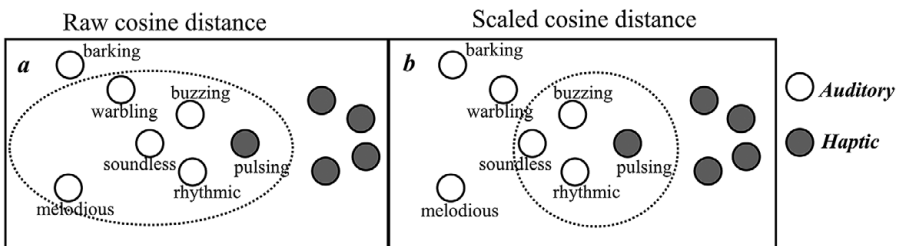
## 2.2. Word Embedding retrievals and calculations of distances between embeddings of different modalities

We retrieved Word Embeddings that are associated with the English language from the Google News vectors database. These vectors are the results of training a neural network on the Google News dataset of about 100 billion words (Mikolov, 2013). We then extracted 300-dimensional Word Embedding vectors (i.e. a list of 300 scores in a vector) for each of the 423 words of the Lynott and Connell norms (2009), mentioned in the Section 3.1 and already classified by their dominant sensory modality.

The next step was to determine how similar or closely related the words previously classified into different sensory modalities were (e.g. the words within the taste and smell modalities). Given a pair of different dominant modality word sets (A,B), we calculated the cosine distance between each pair of Word Embedding vectors ( $\vec{W}_A \in A, \vec{W}_B \in B$ ) and scaled it by the average cosine distance of  $\vec{W}_A$  with all the other word vectors of modality B. Consequently, we detected word vectors of A that are closer to or further from a specific word vector of B, rather than simply closer to or further from all word vectors of B. For instance, the word *pulsing* has high scores in both the auditory and haptic modalities, although its dominant modality is haptic. However, due to its high auditory score, its raw cosine distance is close to the vast majority of auditory words (see Figure 1a). In this way, the raw cosine distance is only capturing the multimodality of the word *pulsing*, but not its relative proximity to specific auditory words or features. By dividing the cosine distance of *pulsing* to each auditory word by the overall average cosine distance of *pulsing* to all auditory words, we determined which auditory words are genuinely close to *pulsing*, beyond the fact that *pulsing* is close to all auditory words in general (see Figure 1b). Notice that the aforementioned process can generate different scaled cosine distances for pair ( $\vec{W}_A, \vec{W}_B$ ) and ( $\vec{W}_B, \vec{W}_A$ ), yielding 20 crossed sets between the five senses.

Scaled cosine distances are positive real numbers. A value below one means a shorter distance among words than the average across all words for this pair of modalities, and a value above one means a larger distance among words than the average across all words for this pair of modalities. For instance, the words *audible* (auditory) and *bronze* (visual) have a large distance (1.20), whereas the words *reverberating* (auditory) and *rippling* (visual) have a short distance (0.48) and might be candidates for a semantic crossmodal correspondence.

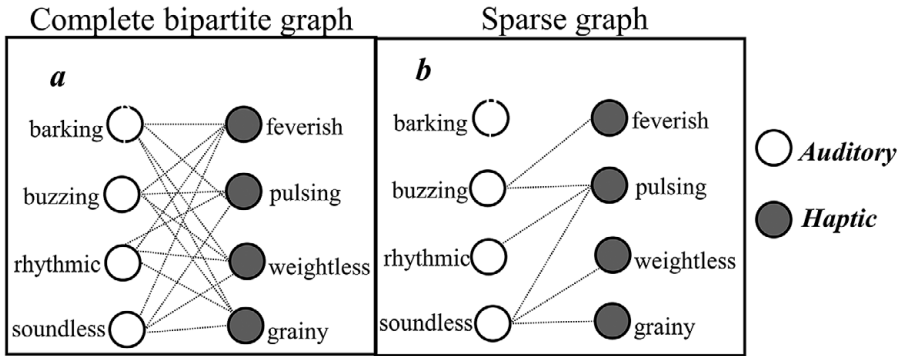
Once calculated, the scaled cosine distances configured a complete bipartite graph between pairs of modalities, including all pairs of modalities. To illustrate, Figure 2a



**Figure 1.** Effects of raw cosine distance vs. scaled cosine distance.

Note: White circles represent auditory words; dark grey circles represent haptic words. Dotted oval depicts words that are close to *pulsing* depending on the selected distance metric. In panel a, raw cosine distance was used; in panel b, scaled cosine distance.





**Figure 2.** Example of complete bipartite graph and sparse graph.

White circles represent selected auditory words; dark grey circles represent selected haptic words. In panel (a), a complete bipartite graph is formed with scaled cosine distance among words; in panel (b), after selecting a threshold for closeness, less than half the relationships remain, leading to a sparse graph where even some words are isolated.

shows an example between two modalities, auditory and haptic, using only four words in each one (*barking*, *buzzing*, *rhythmic* and *soundless* for the auditory modality and *feverish*, *pulsing*, *weightless* and *grainy* for the haptic modality).

### 2.3. Generation of the final crossmodal correspondence network

In order to generate a sparse graph that represents only those relationships between words whose scaled cosine distance is particularly low for each pair of modalities, a threshold was determined by implementing the approach suggested by Tukey (1977) for the selection of extreme points in a univariate distribution. For our particular case, this criterion selected pairs of words from different modalities with distances less than 1.5 times the interquartile range below the mean distance. The finalized sparse graph formed a network of crossmodal associations whose vertices are words from different sensory modalities, and the edges are only those associations between two words from different modalities that have the aforementioned short distances. For instance, Figure 2b shows a reduction in associations after applying the threshold, where the word *pulsing* retains almost all relationships, *grainy* remains connected solely with *soundless*, *barking* gets isolated, and less than half the relationships remain, leading to a sparse graph.

It is important to highlight three aspects: First, the final network is constituted of all the edges (relationships below the selected threshold) found across any pair of modalities. Each edge implies a potential crossmodal correspondence, as it can only occur between words from different modalities. Second, each pair of senses had a different threshold that depends solely on the distribution of distances among these specific pair of senses, avoiding some potential biases due to the hierarchy of senses in English. Third, some vertices (words) may be isolated from any connection with any other modality, and therefore, they do not have any candidate for a crossmodal correspondence (for instance, *barking* in Figure 2b). Such words were removed from the network. We assessed whether there were strong differences between removed and selected words due to dominant modality (using Cramer's V) and modality exclusivity index (using eta square). From this point of analysis, scaled cosine

distances were no longer used, and the only element that remained is the network of vertices and relationships that construct the sparse graph, that is, our network of crossmodal correspondences. We also calculated centrality measures for the vertices, particularly, the eigenvector centrality, with higher values corresponding to more connected vertices, in order to detect words that are highly connected in the network.

#### *2.4. Community detection in the crossmodal correspondence network*

When examining a network, such as the network of sensory associations across modalities, it can be insightful to identify potential communities within it. Community detection is a well-known task performed on networks and involves extracting vertices that have a high density of interconnections among them, thereby constituting a community (Fortunato & Hric, 2016). Community detection can be likened to clustering but is specifically tailored for networks. A common method of community detection revolves around maximizing modularity, a metric that compares the observed number of relationships between vertices to the expected number of relationships that would occur at random. Newman's leading eigenvector method (Newman, 2006) computes principal components on the modularity matrix between vertices and selects the number of components with an eigenvalue greater than one, implying a potential association of vertices higher than expected. Hence, when the eigenvalues are greater than one, it indicates that specific sets of words belong to a community within the network and are even more closely related. Newman's eigenvector method was applied to the resulting crossmodal association network explained in Step 2.3.

#### *2.5. Assessment of robustness of community detection*

To assess the robustness or stability of the communities identified in Step 2.4, we applied the following procedures: (a) implementing a different threshold to select the crossmodal candidate relationships, namely, three standard deviations below the mean for each pair of modalities, and (b) implementing different types of community detection. Specifically, we employed a popular alternative method for maximizing modularity, the Louvain method (Blondel et al., 2008), and a community detection method (Infomap) based on a different strategy, namely, minimum description length of a random walk through the network (Rosvall & Bergstrom, 2008). We then compared the results of the modified communities with the initial communities obtained in Section 3.4 using the purity metric. Such a metric, for each initial community, counts the shared number of words between the highest matching modified community and the initial community and divides it by the total number of words in the initial community.

In addition, it can be expected that sensory domains of experience, in general, overlap with the crossmodal domains of experience, because in contexts that are rich in sensory information people can make use of crossmodality to reduce the complexity of information. To assess this, we compared our communities with a clustering constructed over sensory representations. Particularly, we compared our results with the clusters made by Winter (2019, Ch. 13) over the vector of modality ratings of Lynott and Connell (2009) described in Section 2.1., that yielded 13 clusters showing several large clusters for visual and haptic dimensions, whereas the chemical

senses were constricted to smaller less specialized clusters. In addition, a large cluster of multisensory words was identified.

### 2.6. Identifying and comparing dominant modalities across communities

We analyzed communities with more than five words in order to have enough data to draw meaningful conclusions. We measured the strength of the association between the sensory modality of the words and the community to which they belong using Cramer's  $V$ , and calculated the chi-square statistic to show the most represented modalities in each community.

### 2.7. Identification and comparison of valence, arousal and dominance across communities

We matched each word with values of emotional valence, arousal and dominance extracted from the NRC VAD Lexicon (Mohammad, 2018). Valence is the core emotional dimension of pleasure/displeasure; arousal is the emotional dimension of excitation/calm (Bliss-Moreau et al., 2020; Russell, 2003); and dominance is the emotional dimension of control over the context. The NRC VAD lexicon assigned values of valence, arousal and dominance ranging from zero to one, being zero the lowest and one the highest. Table 1 shows an example of words for each dominant modality with all their correspondent information.

We selected the NRC VAD lexicon over the more widely known affective lexicon by Warriner et al. (2013) for two primary reasons. The first reason pertains to the coverage each lexicon provided for the 423 words from Lynott and Connell (2009). In the case of the NRC VAD lexicon, 111 words (26.2%) from Lynott and Connell's list were not initially found; in the case of Warriner et al. (2013), that number increased to 169 (39.9%). To address this issue, we extracted canonical forms of the words not found in each lexicon (linguistic stems and lemmas). Only 31 (7.3%) words remained without emotional values in the NRC VAD lexicon, whereas 94 words (22.2%) were left without emotional values using the lexicon developed by Warriner et al. (2013). The second reason pertains to a slight improvement in the results of the NRC VAD lexicon compared to that of Warriner et al. (2013). Notably, the NRC VAD lexicon does not rely on obtaining ratings for each word. Instead, it involves the simultaneous comparison of four words, from which the highest and lowest in the quality being measured (for example, valence) are selected. This approach allows for the determination of five out of the six potential relationships among the four words in a single trial (Louviere et al., 2015). This led to a better split-half reliability for the NRC VAD lexicon compared to the lexicon by Warriner et al. (2013). It is worth noting that there

**Table 1.** Final word example information by modality

Word	Dominant modality	Valence	Arousal	Dominance	Centrality	Modality exclusivity
<i>Ugly</i>	Visual	.167	.630	.254	.010	.457
<i>Melodious</i>	Auditory	.932	.530	.587	.997	.644
<i>Bitter</i>	Gustatory	.218	.267	.411	.013	.537
<i>Burnt</i>	Olfactory	.082	.713	.356	.001	.261
<i>Silky</i>	Haptic	.802	.397	.271	.691	.528

is low correlation between the two lexicons regarding arousal and dominance scores (Mohammad, 2018).

We then assessed the emotional differences across communities using eta square as a metric of the difference between communities, and calculated confidence intervals for each estimate of valence, arousal and dominance of each community.

### 2.8. Identification of domains of experience across communities

To identify the domains of experience of each community, we decided to pinpoint the most closely related semantic domains for each of these communities. Semantic domains refer to classifications of concepts, which are reflected in groups of words whose meanings are highly related (Hills et al., 2015; Nerlich & Clarke, 2000). This interconnectedness stems from their association with real-world phenomena, causing these words to revolve around specific subjects or thematic experiences (Brinton & Brinton, 2010). While there are various types of semantic domains, we chose those whose classification is based on areas of interest common to human experience across multiple cultures, specifically, the classification from the Intercontinental Dictionary Series (IDS) (Key & Comrie, 2021) and the Summer Institute of Linguistics (SIL) (Moe, 2003). IDS semantic domains are derived from Buck's (2008) thesaurus, and comprise 22 semantic domains with 1,310 lexical entries across many languages, including English. On the other hand, SIL developed, based on the family of Bantu languages, a list of semantic domains for any language, with nine high level domains in a hierarchy that has up to four levels, comprising around 1,600 total semantic domains. All words from each community were searched on the SIL webpage of semantic domains (Summer Institute of Linguistics, 2021) to identify the most frequently recurring semantic domains. Additionally, for each community, average raw cosine distance between all the words of the community and all the IDS semantic domains was calculated, and the three semantic domains with lowest average distance were extracted for each community. **Sense Perception** is one of the IDS domains, and by its very nature, is closely related to all the communities, without adding any interesting information. That was the reason to discard such a domain of the analyses. Finally, a label was selected for each community assessing the following elements: SIL related domains, IDS related domains, average valence, arousal and dominance and most represented modalities in the community. The previous elements helped to describe each domain of experience (i.e. a semantic domain related to everyday life experiences).

## 3. Results

### 3.1. The crossmodal network

A complete set of 60,265 scaled cosine distances were generated between pairs of words of different senses. A subset of 1,206 pairs of words were selected as crossmodal semantic associations, corresponding to pairs of words from different modalities with particularly small scaled cosine distances between them (distances less than 1.5 times the interquartile range below the mean distance). These 1,206 pairs had 378 distinct words from the original set of 423 words (89.9% of the words from now on the selected words), and comprised 2% of all possible combinations, ranging from 1.1% to 2.6% of all possible combinations for each pair of sensory modalities. The 1,026

**Table 2.** Selection of words by modality

Sense	Not selected	Selected	Total
Visual	29	176	205
Haptic	2	68	70
Auditory	6	62	68
Gustatory	5	49	54
Olfactory	3	23	26

Note: Number of words selected/unselected as crossmodal correspondences candidates divided by modality.

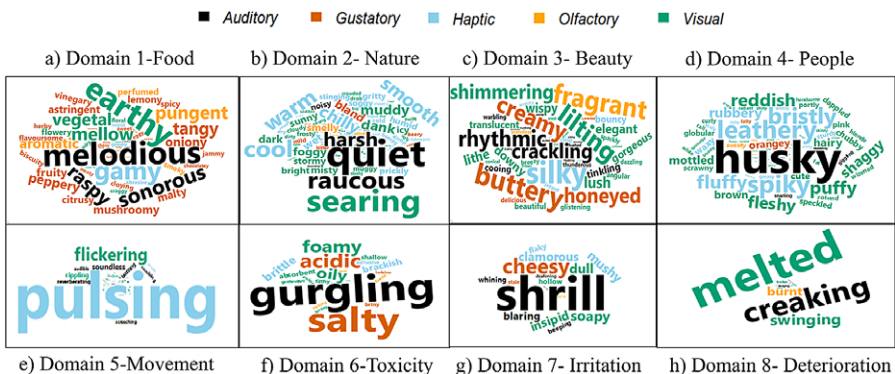
pairs are available as supplementary material in the link at the data availability statement.

Difference between dominant modality of selected words and dominant modality of removed words was little, showing that there is not apparent bias against a particular modality ( $\Phi_{Cramer} = 0.13$ ) (Table 2).

On the other hand, selected words had slightly lower modality exclusivity than removed ones, but the effect size is low ( $\eta^2 = .007$ ), showing a small bias in selection to multisensory words. Overall, there was little evidence of dominant modality or modality exclusivity affecting the formation of the network of crossmodal correspondences, showing that methodology decisions helped to avoid potential biases.

### 3.2. Exploration of communities

Community detection extracted 13 communities or clusters (hereafter, domains of experience). Each domain of experience comprised highly interconnected words, where a connection indicates low cosine scaled distances between the words and, therefore, higher semantic proximity. Five domains of experience with fewer than five words associated, and they were removed from the analysis. Figure 3 shows the words belonging to each of the domains of experience divided by their dominant modality, with sizes proportional to relative centrality within each group. An explanation of the assigned labels is presented in Section 3.3.



**Figure 3.** Words in the domains of experience. Words belonging to each domain of experience. Colors indicate the dominant modality of word, as calculated by Lynott and Conell (2009). Word size is proportional to the centrality of the word in the network.

Figure 4 shows the two largest domains, **Domain 2** and **Domain 4** as a network of relationships, in more detail given their size (for a graph of the remaining domains, see the supplementary material in the data availability statement) Specific areas of the domains of experience are visible.

### 3.3. Domains of experience associations and labelling

Domains of experience differ in the modality of the words contained in them ( $\Phi_{Cramer} = 0.327$ ). In fact, six out of eight domains of experience had a high count of words of a single sense (see Table 3).

Figure 5 shows error bars (Mean and SD) of valence, arousal and dominance for each domain. There are clear differences in average valence ( $\eta^2=.162$ ). Differences in average arousal ( $\eta^2=.041$ ) and average dominance ( $\eta^2=.053$ ) were smaller but important between domains.

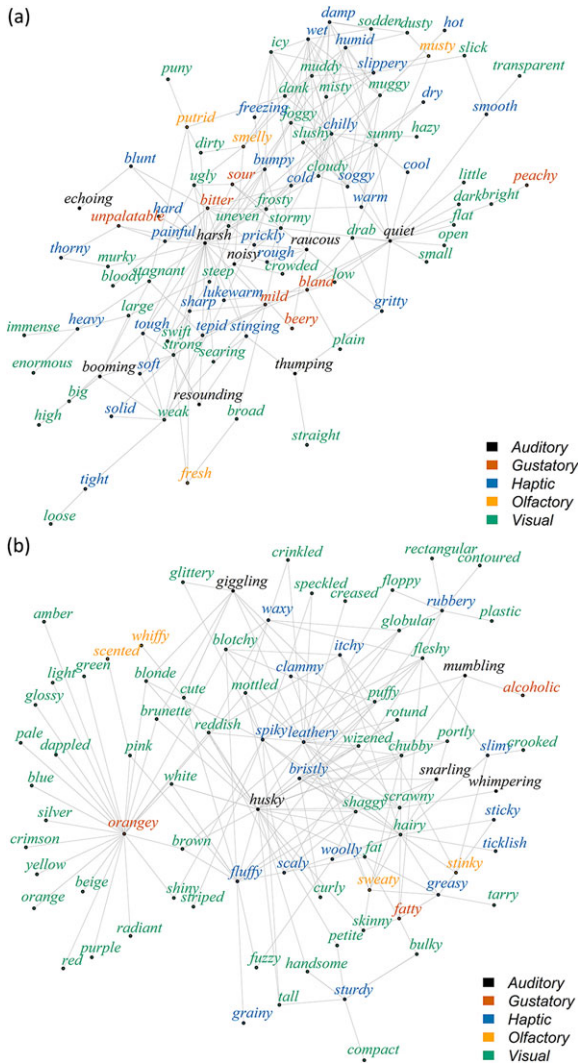
Table 4 shows the results of the labelling process for domains of experience. Across domains of experience, it was noticeable that some IDS semantic domains are highly repetitive: **Emotions and Values** (4 times), **The Body** (4 times), **Food and Drink** (3 times), and **The Physical World** (3 times). Final labels were heuristically selected for better recall and brevity, although their correspondence might not necessarily be one-to-one and the domains might be broader than the labels can capture. This issue is further debated in the discussion section.

Figure 6 shows the labeled domains of experience, binding together their dominant modality, and the three emotional elements assessed, namely, valence, arousal and dominance.

### 3.4. Robustness assessment

Table 5 shows purity results of the three methods of robustness assessment. Results were robust (mean purity above 50%) for five domains (**1**, **3**, **4**, **5** and **7**). This indicates that, even when using alternative clustering methods, a high proportion of words were grouped into similar communities. In the other three domains, two scenarios were observed. **Domain 2 (Nature)**, which comprises a significant number of words, tended to be further subdivided into two or three groups. Conversely, the words from the **Domain 6 (Toxicity)** and **Domain 8 (Deterioration)** often remained isolated or tended to integrate into other communities. Overall, results were fairly robust, with some caveats for two of the smallest communities and a potential subdivision within **Domain 2 (Nature)**.

When comparing our results with Winter (2019), purity is clearly lower than that for variations in method/threshold, as shown in Table 5, on almost every domain. Further analysis showed that the 'multisensory' cluster of Winter (2019) is divided among all our communities, indicating that our results aim to more accurately capture this multisensory phenomenon. In fact, the high purity of **Domain 7 (Toxicity)** is due exactly to this 'multisensory' cluster. Other than that, **Domain 1 (Food)** mixes Winter's 'chemical' and 'taste' clusters (plus some 'multisensory' words) and **Domain 2 (Nature)** extracts words from all the clusters with a bias toward touch and sight clusters. In summary, although some overlap is present, there are clear differences among Winter (2019) clusters and our communities due to their different goals (i.e. grouping the senses vs. capturing crossmodal correspondences in different senses).

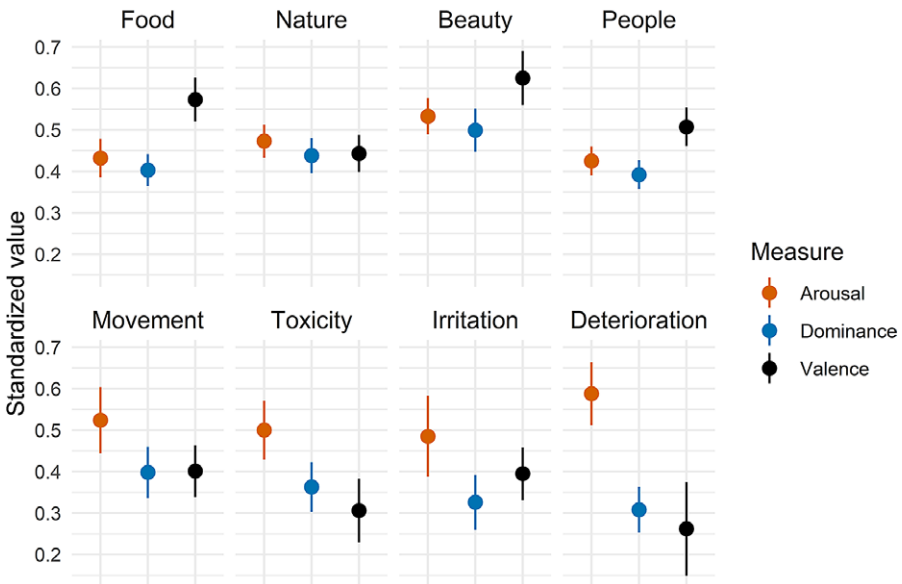


**Figure 4.** Detailed network of the two largest domains of experience. (a) Domain 2. **Domain 2 – Nature** as an example of the matches uncovered in the five sensory modalities (visual, haptic, auditory, gustatory, and olfactory). Links between words indicate semantic closeness. Colors indicate the dominant modality of word, as calculated by Lynott and Conell (2009). In **Domain 2** from left to right, four clear zones are visible. At the extreme lower left, there are mostly haptic and visual words related to mass and size (examples: *immense*, *heavy* and *large*); at the inner lower left, a highly crossmodal zone connected to roughness, and, in general, negative valence words (examples: *painful*, *smelly*, *ugly*, *harsh*, *bitter*) is visible. At the inner upper right, atmospheric patterns related to temperature, humidity and light (examples: *chilly*, *stormy*, *cool*, *warm*, *slippery*) stand out. Finally, at the extreme right, predominantly visual qualities connected with the central words *quiet* and *cool* are found. (b) Domain 4. **Domain 4 – People** as an example of the matches uncovered in the five sensory modalities (visual, haptic, auditory, gustatory and olfactory). Links between words indicate semantic closeness. Colors indicate the dominant modality of word, as calculated by Lynott and Conell (2009). In **Domain 4**, we detected three specific zones. The word *orangey* shows a color-related zone (examples: *blue*, *beige*, *pink*, *blonde*) to the left. On the right, the words *spiky* and *leathery* lead to an upper zone mostly related with textures (examples: *blotchy*, *mottled*) and shape-related words (examples: *rectangular*, *globular*, *speckled*) whereas the words *bristly* and *husky* lead to a lower zone connected with animal and human body properties (examples: *fatty*, *sweaty*, *hairy*, *snarling* and *scrawny*).

**Table 3.** Domains of experience association with modalities (standardized residuals)

Domain of experience (number of words)	Auditory	Gustatory	Haptic	Olfactory	Visual
Domain 1 – Food (53)	-0.82	<b>7.87</b>	<b>-1.86</b>	<b>1.58</b>	<b>-3.14</b>
Domain 2 – Nature (95)	<b>-1.81</b>	<b>-1.60</b>	<b>2.96</b>	-0.71	0.30
Domain 3 – Beauty (52)	<b>2.37</b>	-0.74	-0.85	-0.63	-0.22
Domain 4 – People (83)	<b>-2.24</b>	<b>-2.43</b>	0.16	-0.44	<b>2.67</b>
Domain 5 – Movement (33)	<b>3.41</b>	<b>-2.10</b>	-0.86	-1.41	0.18
Domain 6 – Toxicity (22)	-1.33	0.62	-0.53	<b>2.33</b>	-0.06
Domain 7 – Irritation (16)	<b>2.18</b>	-0.09	0.02	0.04	-1.25
Domain 8 – Deterioration (13)	1.36	-1.32	-0.91	0.25	0.40

Note: The numbers shown represent the standardized residuals. Numbers in bold have an absolute value greater than 1.5 standard deviations. Red cells mean frequency words lower than expected for this sense in this community; green cells mean frequency words higher than expected for this sense in this community



**Figure 5.** Error bars of valence, arousal and dominance for each domain of experience. Middle point depicts the average of each emotional feature in the group; emotional features range from 0 to 1. Upper and lower limits are limits with 95% of confidence, Bonferroni corrected.

#### 4. Discussion

The present study focused on studying the emergence of crossmodal correspondences in language, offering a comprehensive analysis of how crossmodal experiences are interconnected in linguistic constructs. Overall, our data support both the semantic coding and embodied lexicon hypotheses. The discussion revolves around this particular support, which we present in four sections. In section 4.1, the evidence for perceptual crossmodal associations transferred to language is presented for three specific cases (pitch, color and tastes), supporting the embodied lexicon hypothesis. In sections 4.2 and 4.3, evidence of the importance of the statistical and emotional mechanisms of crossmodal correspondences for the semantic organization of



**Table 4.** Domains of experience features

Domain of experience #	IDS closer domains	Related SIL semantic domains	Label
1	Food and drink Agriculture and vegetation Animals	Daily life (food), work and occupation (agriculture), person (medicinal plants).	Food
2	Emotions and values The physical world Spatial relations	Universe (world, sky, nature, weather).	Nature
3	Clothing and grooming The physical world Food and drink	Person/perceive (appearance, beautiful, types of sounds); social behavior (music); physical actions (move gracefully), states of quality (texture, shine).	Beauty
4	Clothing and grooming Animals The body	States (color), animals, person body (particularly hair, skin and appearance), social behavior and relationships (race), daily life (cosmetics, clothing style).	People
5	Speech and language Emotions and values The body	Body (heart, eye, stomach), water (movement), animals (birds, animal sounds), perceive sounds (loud and quiet), physical actions (moves noisily or moves a part of the body).	Movement
6	Food and drink The body The physical world	Daily life (food), universe (substance, matter), work and occupation (artificial).	Toxicity
7	Emotions and values Speech and language Food and drink	Language and thought (voice, complain, uninterested), person/perceive (loud, types of sounds), states of quality (decay).	Irritation
8	Basic actions and technology The body Emotions and values	Physical actions (move noisily and fall, physical impact, divide into pieces, break), person (disease, fall and die), states (decay), social behavior (conflict and trouble).	Deterioration

*Note:* Domains of experience description based on SIL and IDS semantic domains. Domains of experience labels were assigned not only based on SIL and IDS semantic domains, but also in emotional scores and dominant modalities per domain.

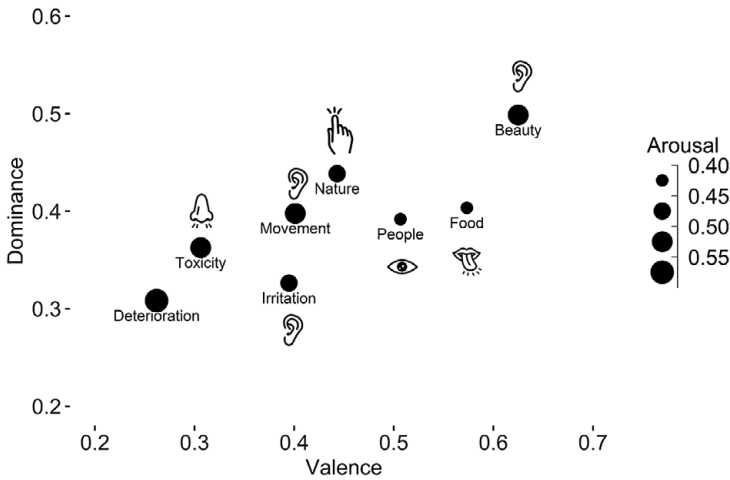
associations between the senses is discussed. Section 4.4 presents the domains of experience and their overall connection with the semantic coding hypothesis.

#### 4.1. Perceptual crossmodal associations in the semantic network

We focused on three particular cases: pitch, color and shapes.

##### 4.1.1. Pitch

There are no words that solely represent high or low pitch in our dataset. However, several auditory words have pitch as a part of their meaning. Surprisingly, most words



**Figure 6.** Characterization of domains of experience. X-axis depicts valence. The Y-axis depicts dominance. Point size shows arousal. Values of the variables are in the range [0, 1]. Predominant significant sense (i.e. the sense with higher positive residual as shown in Table 2) is depicted aside the domain of experience. Icons of senses by Takao Umehara from NounProject.com.

**Table 5.** Purity of communities after robustness assessment

Community	Louvain	Infomap	3σ threshold	Mean purity	Winter (2019)
Food	62.3% (53)	58.5% (53)	77.3% (44)	<b>66.0%</b>	33% (53)
Nature	43.2% (95)	25.3% (95)	35.5% (76)	<b>34.6%</b>	23% (95)
Beauty	67.3% (52)	40.4% (52)	51.1% (47)	<b>52.9%</b>	18% (52)
People	78.3% (83)	47.0% (83)	42.9% (56)	<b>56.1%</b>	20% (83)
Movement	75.8% (33)	39.4% (33)	40.9% (22)	<b>52.0%</b>	27% (33)
Toxicity	36.4% (22)	36.4% (22)	21.1% (19)	<b>31.3%</b>	48% (22)
Irritation	75.0% (16)	68.8% (16)	69.2% (13)	<b>71.0%</b>	31% (16)
Deterioration	61.5% (13)	46.2% (13)	33.3% (9)	<b>47.0%</b>	23% (13)

Note: Values of purity (number of words of original community) for each robustness assessment and for Winter (2019) clusters. Notice that 3σ threshold has a lower number of words per community, due to a tighter threshold for candidate crossmodal correspondence selection. Mean purity for the three robustness assessments – except Winter (2019) – is also depicted.

related to high pitch are concentrated in **domain 3 (Beauty)**: *cooing, tinkling, wailing, whistling, jingling, crackling*. Notice that other domains do not have clear pitched words, such as **Domain 1, 5, 6 or 8** (see Figure 3, particularly panel (c)). There is also a group of brightness-related word in the same domain: *shimmering, glistening, brilliant*. The relationship of brightness and high pitch is a well-known crossmodal correspondence, appearing with both perceptual and semantic stimuli (Martino & Marks, 1999; Spence, 2011).

Moreover, **Domain 3 (beauty)** is a high valence domain that includes very specific gustatory words. Two of them, *honeyed* and *caramelised*, are semantically related to sweetness. Concurrently, there is a tested association between sweet taste and high pitch, whether stimuli are gustatory words (Crisinel & Spence, 2009, 2010a) or real tastants (Crisinel & Spence, 2010b; Wang et al., 2016). Importantly, such relationships seem to be emotionally mediated (Wang et al., 2016).

Studies of the pitch–flavor relationship have found that sour is associated with high pitch too (Crisinel & Spence, 2009, 2010b; Knoeferle et al., 2015). It is important to notice that the word *sour* has no close relations with high-pitched words and it is located in **Domain 2 (Nature)** close to words of negative valence (*ugly, harsh, painful*). In addition, four important high pitch words are not in **Domain 3 (Beauty)**: *giggling/whimpering* in **Domain 4 (People)** (Figure 4b) and *shrill/whining* in **Domain 7 (Irritation)** (Figure 3f). All four words have negative valence. Overall, there is no clear relationship between high-pitched negative words and tastes in our evidence.

Low-pitched words are not that clear in our stimuli. They seem to be more located in **Domain 2 (Nature)** (*resounding, raucous*), and be confounded with loudness, hindering the possibility of assessing pitch–taste relationships, since loudness has been connected to the concentration of tastants, rather than its taste (Wang et al., 2016).

Is pitch associated with high and low spatial locations in our results? Words *high* and *low* are both in **Domain 2 (Nature)** (Figure 4a). The only crossmodal connection of *high* is *booming* which is arguably related with low pitch but also with loudness. In turn, *booming* is also connected with *big* and *large*, suggesting the well-known association between pitch and size, i.e. lower pitches are associated with bigger objects (Korzeniowska et al., 2022). On the other hand, *low* is connected with *quiet* and *mild*, suggesting an association based upon intensity. Thus, the presence of crossmodal correspondences between pitch and vertical position in our evidence is unclear, likely due to the interference from other perceptual dimensions affecting the meaning of words.

In summary, there is fair evidence of known pitch crossmodal associations in the semantic crossmodal network, although not all relationships are clearly present in our results.

#### 4.1.2. Color

As showed in Figure 4b, the most part of color words related to hue are located in **Domain 4 (People)** and share a single link with the gustatory word *orangey*. At a first glance, apparently crossmodal associations with color are mostly absent in the evidence. A closer look reveals a different picture.

**Domain 4 (People)** deals with animals, human bodies and clothing. The presence of the colors in such domain is particularly relevant, given the importance of colors in visual recognition (Bramão et al., 2011). In fact, source identification is key to the relationships found between colors, odors and tastes; food color is diagnostic (Saluja & Stevenson, 2018; Spence, 2020b). Without proper source identification, associations with colors display great variability among cultures, individuals (Goubet et al., 2018; Wan et al., 2014) and even food categories (Velasco et al., 2023). Thus, they might not clearly emerge in a semantic network specifically crafted to avoid source objects. However, the visual and central word ‘earthy’ in **Domain 1 (Food)**, referring to a wider range of colors identifiable with multiple sources, show connections with gustatory words (*chocolatey, cloying, flavoursome*) and olfactory words (*aromatic, fragrant, perfumed*). Close crossmodal relationships of colors, odors and tastes in the food context have been long studied (see Spence, 2020b, for a review).

Referring to the relationship between colors and music, it is worth noting that emotional mediation is a likely frequent mechanism of correspondence, and that the

relationship encompasses hue, lightness and saturation (Palmer et al., 2013). As shown in the previous section, high-valenced **Domain 3 (Beauty)** comprises many bright-related words (*shimmering, glowing*) along with auditory pleasant words (*cooing, tinkling*). Moreover, the word *rhythmic* is central in such domain. Finally, the word *melodious* in the high-valenced **Domain 1 (Food)** is directly linked to color words such as *earthy* and *flowery*.

Overall, we found evidence of color crossmodal associations mediated by emotion within the semantic crossmodal network.

#### 4.1.3. Tastes

When discussing pitch, we have already evaluated its correspondence with tastes. In addition, crossmodal correspondences literature shows a strong relationship between tastes and visual-haptic features, where sweetness is related to curved shapes (Velasco et al., 2016b), creaminess (Carvalho et al., 2017) and softness (Pistolas & Wagemans, 2023), whereas sourness/saltiness/bitterness are related to sharpness/angularity (Velasco et al., 2016a). In such relationships, some studies have demonstrated emotional mediation (Chuquichambi et al., 2024; Salgado Montejo et al., 2015; Velasco et al., 2015). Sweet-related words are in **Domain 3 (Beauty)** (e.g. *honeyed, caramelised*, see Figure 3c) and in **Domain 1 (Food)** (e.g. *sweet, biscuit*, see Figure 3a). Both domains are high in valence. In **Domain 3 (Beauty)**, there are many haptic central words (*silky, buttery, creamy*) and the word *curved* is linked to *creamy*. On the other hand, *sour* and *bitter* are in **Domain 2 (Nature)**, close to *sharp, bumpy, steep, rough* and *prickly*. It is noticeable that *angular* is in **Domain 3 (Beauty)** and *soft* in **Domain 4 (Nature)**. Such closeness of antonyms is well known in Word Embeddings, and arise from the fact that they probably share similar contexts (Ono et al., 2015). Nevertheless, they are not central in their respective assigned domains.

In summary, we found compelling evidence for taste-related crossmodal correspondences, likely mediated by emotion, in the semantic crossmodal network.

Broadly speaking, the three cases presented earlier show that perceptual associations in the crossmodal phenomena are encoded in language structures, supporting the embodied lexicon hypothesis.

## 4.2. Emotion in the crossmodal semantic network

Our results show that emotion is a key dimension of the semantic organization of concepts between sensory modalities (see Figures 5 and 6). Indeed, valence, arousal and dominance emerged as salient dimensions within the domains of experience. This observation resonates with Spence and Deroy's (2013) work, emphasizing the role of affective factors in mediating crossmodal correspondences (see also Spence, 2020a).

In our research, valence showed the largest effect size of the three emotional elements, suggesting its role as a core component of the crossmodal correspondence architecture in language. The previous section depicted valence-mediated crossmodal relationships about the perceptual features of pitch, color and tastes.

Dominance had the second effect in size across communities. Low dominance highlights domains of experience where there might be out-of-control threat elements. In **Domain 7 (Irritation)**, several elements of powerless perception are bound together: the auditory complaining of *whining* and *clamorous*, with the loose texture

of *flaky* and *mushy* and the visual lacking vigor of *insipid* and *dull*. These elements are not easily linked to a single simultaneous experience and are more easily explained due to their connection to a similar emotional state. The role of dominance in crossmodal correspondences continues to be a topic of exploration, potentially involving high-order cognitive based influences in specific crossmodal relationships.

Regarding arousal, the higher arousal intensity of **Domain 8 – Deterioration**, reflects an increase in the urgency of acting to address potential risks. Conversely, the lower arousal intensity of **Domain 1 – Food** refers to potentially enjoyable experiences, where the word *melodious* is linked to color words such as *earthy* and *flowery*. Several research studies have shown arousal as a mediator in the correspondences of music and color (Lindborg & Friberg, 2015; Whiteford et al., 2018).

The centrality of the positively valenced words *melodious* and *sonorous* within the **Food** domain (Figure 3a) supports and adds to the long-established relationship between music and food, that seems to be mediated by emotion (Reinoso-Carvalho et al., 2020; Spence, 2019). In fact, pleasures, such as food, music and sex appear to activate similar brain regions (Blood & Zatorre, 2001).

One standing issue regarding the role of emotion in the semantic crossmodal network concerns the relative contribution of the emotional mediation versus other potential explanations, such as the statistical co-occurrence of emotional experiences or the semantic use of shared emotional qualities. That is, beyond emotional mediation, certain multisensory experiences might share affect (i.e. *rancid* and *rotten* in **Domain 6 – Toxicity**) or we might frequently use words with shared affect to convey meaning (i.e. *sweet* and *melodious* in **Domain 1 – Food**). Although it is not possible to fully separate the contribution of these factors in the present research, the examples presented earlier (i.e. pitch and taste, music and color) demonstrate the emotional mediation that is present in our evidence. Therefore, emotional mediation likely contributes, to a certain degree and alongside other factors, to the role of emotion in the semantic network of association between senses in language.

#### 4.3. Statistical associations in the crossmodal semantic network

Domains not clearly characterized by emotional features may be better characterized by statistical associations, (i.e. similar arrangement of stimuli in time and space across senses). **Domain 5 – Movement** (Figure 3e) hints toward relationships determined by patterns in space and time. The triad *flickering* (visual)–*pulsing* (haptic)–*reverberating* (auditory) is a good example, suggesting that the temporal crossmodal nature of the sound-induced flash illusion (Hirst et al., 2020) appears in language and might extend to other senses.

**Domain 2 – Nature** (Figure 4a) also shows strong connections between temperature (*hot*, *warm*, *cold*), humidity (*slippery*, *dry*, *wet*) and light words (*sunny*, *foggy*, *hazy*), all related to weather. Crossmodal literature has mostly explored the relationship of hue and pitch with temperature, but has focused less on the relationship with brightness, based upon intensity (Spence, 2020c). Some previous linguistic analyses on weather have found both a component representing weather conditions and an evaluative component, showing that emotional and statistical correlation factors are present in the weather lexicon (Stewart, 2007). However, weather is usually evaluated in valence terms (i.e. positive and negative) (Stewart, 2020), likely affecting the not-polarized valence found as an average in the **Nature** domain of experience.

In **Domain 4 – People** (Figure 4b), textural and visual patterns (*fluffy, leathery, waxy*) connect with people descriptors (*blonde, brunette*) and then with colors. This shows that these patterns might become a mechanism for object, animals and people recognition, and that such recognition might be facilitated by statistical associations between senses. For instance, *fatty* (gustatory), *bulky* (visual), *greasy* (haptic) and *stinky* (olfactory) might arguably stem from a repeated exposure to the same multi-sensory experiences in the environment not necessarily related to a single source.

Intensity is a well-known crossmodal physical underlying phenomenon, frequently associated with the senses of smell and taste (Hanson-Vaux et al., 2013; Velasco et al., 2016a). Our evidence is plentiful, showing intensity-related crossmodal correspondences (examples: *heavy/large/booming* in the **Nature** domain, *crackling/tingly/glowing* in the **Beauty** domain, *soundless/tiny/weightless* in the **Movement** domain). Sometimes the variation in the direction of the intensity (for instance, the *chilly/drab/raucous* connection in the **Nature** domain) is concurrent with the specific context (a cold, dark, windy day).

Crossmodal statistical associations related to changes in space and time are also present in the evidence. There are two striking patterns related to transformation and change: process of disintegration and fire-related changes. Disintegration changes relate to the physical change actions of break and fall (*burnt, melted, creaking* – Figure 3h). This is the reason for naming the domain 'Deterioration': Processes within this domain share a decrease in the organization of systems that can be translated between senses.

Fire-related changes proved highly crossmodal in nature in our evidence: The words *crackling* (auditory), *crisp* (haptic) and *charred* (visual) are connected between them and initially located in the **Beauty** domain (Figure 3c). What is more, they are connected with several gustatory-olfactory words (i.e. the connection of *crackling* with *earthy, lemony, oniony, smoky* in the **Food** domain, and the connection of *charred* with *burnt* in the **Deterioration** domain). Furthermore, they are connected to several light-related words (i.e. the connection of *crisp* with *vivid, brilliant* and *dazzling*). In fact, *crackling* is one of the more central words in our network.

#### 4.4. Binding all together: Domains of experience

Here, it is important to mention that the emotional and statistical accounts of crossmodal correspondences might not be mutually exclusive (Spence, 2011). Instead, as suggested by our study, they appear to be part of a broader architecture of crossmodal associations, as captured in language. Importantly, each crossmodal domain of experience depicts a more complex picture than captured by the assigned label. Indeed, each domain of experience might include a mix of statistical correlations and emotional correspondences. The **Nature** domain (Figure 4a), for instance, exhibits a large subnetwork of valenced words (*ugly, painful, harsh, putrid*). The **Beauty** domain (Figure 3c) also shows many patterned words (*rhythmic, shimmering, tinkling, creamy*).

Overall, the combination of emotion and statistical correlations in the present evidence reveals eight key domains of experience. According to our evidence, such domains are connected to the broader and cross-cultural IDS domains of **Food, The Physical World, The Body and Emotions and Values**. We believe that these findings help to further clarify specific sensory domains where diverse types of crossmodal

correspondences might be important. In fact, most of these domains were well-established domains of experience connected before to crossmodal correspondences (for the physical world, see Parise & Spence, 2009; for emotion, see Spence, 2020a; for food, see Velasco et al., 2014). It is also worth highlighting domains present in IDS that did not appear in our analysis. Although absence of evidence is not evidence of absence, our results did not show crossmodal communities strongly associated with IDS semantic domains such as **The House, Warfare and Hunting, Religion and Beliefs, and Social and Political Relations**, among the most striking. For instance, despite the documented connection between moral judgment and taste (Chapman et al., 2009; Eskine et al., 2011), we did not observe this connection reflected in our work, where it could have appeared in the form of connections with **Religion and Beliefs** or **Social and Political Relations**.

We believe that the evidence presented in this research supports a clear connection between crossmodal correspondences and their encoding in language. More particularly, it supports the idea that the emotional and statistical mechanisms of crossmodal correspondences formation are encoded in language. A potential organization of a semantic network, wherein perceptually related concepts are linked according to their relationships and matches across modalities, as predicted by the semantic coding hypothesis.

## 5. Conclusions, limitations, and future work

Our findings indicate that many perceptual crossmodal correspondences are embedded in language. Beyond simply reflecting these correspondences, language can capture their emotional and perceptual mechanisms (statistical correspondences) in a crossmodal semantic network. These results align with the embodied lexicon hypothesis and the semantic coding hypothesis.

The crossmodal semantic network reveals domains of experience where crossmodal correspondences are present across a wide range of everyday life contexts, including food, weather, nature, the body of people and animals, danger, threat and fire.

Our work underscores the connection between perception and language and the importance of emotions. It demonstrates that our feelings and the statistical regularities that we perceive are key in organizing sensory concepts across different domains (Noel et al., 2018).

Our results open interesting opportunities to discover, or hypothesize, specific crossmodal perceptual correspondences contributing to a new perceptual research agenda. Furthermore, studying sensory associations using a computational model allow us to move beyond one-on-one matches and consider how crossmodal (in) consistency across contexts and objects, influences experience, decision-making, and sensory integration. After all, sensory associations impact how we integrate sensory information. Understanding how language encodes crossmodal correspondences is crucial as it ultimately influences how neurons encode perceptual information. Our analysis can also inform the development of hypotheses and guidelines for fields as diverse as design (Velasco & Spence, 2019) or neurological rehabilitation (Tinga et al., 2016).

Future research applying a similar method across languages can provide insights into the evolutionary drivers of linguistic constructs. If certain crossmodal associations are shared across languages, it might suggest deep-rooted evolutionary reasons for these correspondences. Given that distributional semantic encodings have been built on diverse languages, analyzing the differences and similarities in crossmodal correspondences across these languages can shed light on cultural nuances in sensory perceptions.

It is important to mention here, that our methodological decisions might have influenced the results obtained. For instance, expanding the list of words used from Lynott and Connell (2009) could provide a clearer picture of crossmodal correspondences in language. Additionally, our results may not generalize to other languages as we focused solely on English. Word Embeddings have limitations in providing a complete encoding of perception and fully represent semantic memory but can also capture other language features. Moreover, some similarities between words be the result of relationships could be due to relationships like hyponymy or morphology. For instance, *lemony* and *tangy* might be considered hyponyms of *citrusy*, leading to their grouping in **Domain 1**; or the suffix *-ing* in *creaking*, *swinging* and *banging* might tie these words in **Domain 8**. It is important to notice that hyponymy and morphology are also intertwined with perceptual semantics, making it difficult to disentangle them with the available methods.

Measuring emotion-related words presents important caveats. Scoring the words in terms of valence, arousal, and dominance may present some challenges as the field of emotion research is continuously updating and revising its methods and approaches. What is understood by an emotion and the relationship between emotions and language may change as the field presses forward. Finally, the community detection method used, as shown in the robustness analysis, creates larger groups than other methods. We consider this a decision that allows a level of generalization appropriate to an initial exploration of the crossmodal language that can be further explored with other community detection methods.

Another limitation of our study is the potential for imposed labels on the identified domains of experience. Our findings suggest that the communities/domains may convey information beyond the labels we applied, focusing instead on specific features that could be more strongly associated. Future research can delve deeper into this to enhance our understanding of the domains of experience.

Given the complex nature of the crossmodal correspondences and their relation to language, it is reasonable to assume that other factors may influence the architecture of semantic crossmodal correspondences. Nevertheless, our results support the importance of emotion and perceptual statistical associations in the semantic organization of sensory concepts between sensory modalities in language, around domains of experience.

Language cannot fully represent crossmodal correspondences, because many experiences and sensorial features lack corresponding words. Additionally, the extent to which Word Embeddings capture specific perceptual information remains unclear. Cross-cultural analysis may uncover similarities and differences between crossmodality representation in different languages, potentially revealing common underlying phenomena.

**Data availability statement.** All code and data to reproduce results, and final data results are available at: [https://osf.io/yrmhb/?view\\_only=9151b5859e6b4215998802222bf1c38b](https://osf.io/yrmhb/?view_only=9151b5859e6b4215998802222bf1c38b).



**Acknowledgements.** The authors thank the anonymous reviewers, and especially the action editor Bodo Winter, for their thoughtful and thorough comments on the manuscript.

**Competing interest.** Author 3 is owner of the company Atrianna ([www.atrianna.com](http://www.atrianna.com)).

## References

- Arshamian, A., Gerkin, R. C., Kruspe, N., Wnuk, E., Floyd, S., O'Meara, C., Rodriguez G. G., Lundström J. N., Mainland, J. D., & Majid, A. (2022). The perception of odor pleasantness is shared across cultures. *Current Biology*, 32(9), 2061–2066, e2063. <https://doi.org/10.1016/j.cub.2022.02.062>
- Bakker, I., Van Der Voordt, T., Vink, P., & De Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology*, 33, 405–421.
- Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, 33(3–4), 130–174.
- Bliss-Moreau, E., Williams, L. A., & Santistevan, A. C. (2020). The immutability of valence and arousal in the foundation of emotion. *Emotion*, 20(6), 993–1004. <https://doi.org/10.1037/emo0000606>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), 11818–11823. <https://doi.org/10.1073/pnas.191355898>
- Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In I. Guyon, U. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (Eds.) *Advances in neural information processing systems*. Red Hook, NY, USA: Curran Associates Inc., 4356–4364.
- Bramão, I., Reis, A., Petersson, K. M., & Fáisca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, 138(1), 244–253.
- Brinton, L., & Brinton, D. M. (2010). *Workbook: The linguistic structure of Modern English*. John Benjamins Publishing Company. Retrieved 16/11/2022 from <https://benjamins.com/sites/z.156/exercise/c6q4>.
- Buck, C. D. (2008). *A dictionary of selected synonyms in the principal Indo-European languages*. University of Chicago Press.
- Carvalho, F. R., Wang, Q. J., Van Ee, R., Persoone, D., & Spence, C. (2017). “Smooth operator”: Music modulates the perceived creaminess, sweetness, and bitterness of chocolate. *Appetite*, 108, 383–390.
- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, 323(5918), 1222–1226.
- Chuquichambi, E. G., Munar, E., Spence, C., & Velasco, C. (2024). Individual differences in sensitivity to taste-shape crossmodal correspondences. *Food Quality and Preference*, 115, 105110.
- Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125(3), 452–465.
- Connell, L., & Lynott, D. (2014). I see/hear what you mean: Semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General*, 143(2), 527.
- Crisinel, A.-S., & Spence, C. (2009). Implicit association between basic tastes and pitch. *Neuroscience Letters*, 464(1), 39–42.
- Crisinel, A.-S., & Spence, C. (2010a). A sweet sound? Food names reveal implicit associations between taste and pitch. *Perception*, 39(3), 417–425.
- Crisinel, A.-S., & Spence, C. (2010b). As bitter as a trombone: Synesthetic correspondences in nonsynesthetes between tastes/flavors and musical notes. *Attention, Perception, & Psychophysics*, 72, 1994–2002.
- Das, R., Zaheer, M., & Dyer, C. (2015). Gaussian LDA for topic models with word embeddings. In Ch. Zong & M. Strube (Eds.) *ACL-IJCNLP 2015 - 53rd Annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing of the Asian Federation of natural language processing, Proceedings of the conference*. Association for Computational Linguistics.
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, 24(5), 613–621.

- Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological science*, 22(3), 295–299.
- Feinberg, T. E., & Mallatt, J. (2016). The evolutionary origins of consciousness. In R. Poznanski, J. Tuszynski & T. Feinberg (Eds.) *Biophysics of consciousness: A foundational approach*. World Scientific Pub Co Inc. (pp. 47–86). <https://doi.org/10.1142/9444>
- Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1–44.
- Fu, R., Guo, J., Qin, B., Che, W., Wang, H., & Liu, T. (2014). Learning semantic hierarchies via word embeddings. In K. Toutanova & H. Wu (Eds.) *52nd annual meeting of the association for computational linguistics, ACL 2014 - Proceedings of the conference*. Association for Computational Linguistics.
- Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science*, 8(5), 573–585. <https://doi.org/10.1177/1745691613498098>
- Goubet, N., Durand, K., Schaal, B., & McCall, D. D. (2018). Seeing odors in color: Cross-modal associations in children and adults from two cultural environments. *Journal of Experimental Child Psychology*, 166, 380–399.
- Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, 6, 975–987. <https://doi.org/10.1038/s41562-022-01316-8>
- Gurnee, W., & Tegmark, M. (2023). Language models represent space and time. Preprint, [arXiv:2310.02207](https://arxiv.org/abs/2310.02207).
- Hanson-Vaux, G., Crisinel, A.-S., & Spence, C. (2013). Smelling shapes: Crossmodal correspondences between odors and shapes. *Chemical Senses*, 38(2), 161–166.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162.
- Hauck, P., von Castell, C., & Hecht, H. (2022). Crossmodal correspondence between music and ambient color is mediated by emotion. *Multisensory Research*, 35(5), 407–446.
- Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science*, 7(3), 513–534. <https://doi.org/10.1111/tops.12151>
- Hirst, R. J., McGovern, D. P., Setti, A., Shams, L., & Newell, F. N. (2020). What you see is what you hear: Twenty years of research using the sound-induced flash illusion. *Neuroscience & Biobehavioral Reviews*, 118, 759–774.
- Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, 4(1), 103–120.
- Kenneth, J. H. (1923). Mental reactions to smell stimuli. *Psychological Review*, 30(1), 77–79. <https://doi.org/10.1037/h0068405>
- Key, Mary Ritchie & Comrie, Bernard (eds.) 2021. *The Intercontinental Dictionary Series*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <https://ids.cld.org>, Accessed on 2021-04-03.)
- Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, 28(47), 12224–12230. <https://doi.org/10.1523/JNEUROSCI.3579-08.2008>
- Knoeferle, K. M., Knoeferle, P., Velasco, C., & Spence, C. (2016). Multisensory brand search: How the meaning of sounds guides consumers' visual attention. *Journal of Experimental Psychology: Applied*, 22(2), 196–210. <https://doi.org/10.1037/xap0000084>
- Knoeferle, K. M., Woods, A., K ppler, F., & Spence, C. (2015). That sounds sweet: Using cross-modal correspondences to communicate gustatory attributes. *Psychology & Marketing*, 32(1), 107–120.
- Konkol, M., Brychc n, T., Nykl, M., & Hercig, T. (2017). Geographical evaluation of word embeddings. In G. Kondrak & T. Watanabe (Eds.) *Proceedings of the eighth international joint conference on natural language processing* (Volume 1: Long Papers). Asian Federation of Natural Language Processing.
- Korzeniowska, A., Simner, J., Root-Gutteridge, H., & Reby, D. (2022). High-pitch sounds small for domestic dogs: Abstract crossmodal correspondences between auditory pitch and visual size. *Royal Society Open Science*, 9(2), 211647.
- Kryklywy, J. H., Ehlers, M. R., Anderson, A. K., & Todd, R. M. (2020). From architecture to evolution: Multisensory evidence of decentralized emotion. *Trends in Cognitive Sciences*, 24(11), 916–929. <https://doi.org/10.1016/j.tics.2020.08.002>
- Kuhnke, P., Kiefer, M., & Hartwigsen, G. (2020). Task-dependent recruitment of modality-specific and multimodal regions during conceptual processing. *Cerebral Cortex*, 30(7), 3938–3959. <https://doi.org/10.1093/cercor/bhaa010>

- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, 158(4), 405–414. <https://doi.org/10.1007/s00221-004-1913-2>
- Levinson, S. C., & Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, 29(4), 407–427.
- Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, 4, 1021–1028. <https://doi.org/10.1038/s41562-020-0918-6>
- Lindborg, P., & Friberg, A. K. (2015). Colour association with music is mediated by emotion: Evidence from an experiment using a CIE Lab interface and interviews. *PLoS One*, 10(12), e0144013.
- Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge University Press.
- Louwerse, M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, 35(2), 381–398.
- Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, 3(2), 273–302.
- Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science*, 33(1), 51–73.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41(2), 558–564. <https://doi.org/10.3758/BRM.41.2.558>
- Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, 45, 516–526.
- 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., Woll, B., LeLan, B., de Sousa H., Cansler B. L., Shayan S., de Vos, C., Senft, G., Enfield, N. J., Razak, R. A., Fedden, S., Tufvesson, S., Dingemans, M., Ozturk, O., Brown, P., Hill, C., Le Guen, O., Hirtzel, V., van Gijn, R., Sicoli, M. A., & Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45), 11369–11376. <https://doi.org/10.1073/pnas.1720419115>
- Marks, L. E. (1996). On perceptual metaphors. *Metaphor and Symbol*, 11(1), 39–66. [https://doi.org/10.1207/s15327868ms1101\\_3](https://doi.org/10.1207/s15327868ms1101_3)
- Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, 28(7), 903–923.
- 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>
- Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.
- Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 15(1), 69.
- Mikolov, T. (2013). *word2vec Google Code Archive*. Google. <https://code.google.com/archive/p/word2vec/>.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In Y. Bengio & Y. LeCun (Eds.) *1st international conference on learning representations, ICLR 2013 - Workshop track proceedings*.
- Moe, R. (2003). Compiling dictionaries using semantic domains. *Lexikos*, 13, 215–223.
- Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 1: Long Papers)*, Melbourne, Australia.
- Motoki, K., Marks, L. E., & Velasco, C. (2023). Reflections on cross-modal correspondences: Current understanding and issues for future research. *Multisensory Research*, 1, 1–23.
- Nerlich, B., & Clarke, D. D. (2000). Semantic fields and frames: Historical explorations of the interface between language, action, and cognition. *Journal of Pragmatics*, 32(2), 125–150. [https://doi.org/10.1016/S0378-2166\(99\)00042-9](https://doi.org/10.1016/S0378-2166(99)00042-9)
- Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3), 036104.

- Noel, J. P., Blanke, O., & Serino, A. (2018). From multisensory integration in peripersonal space to bodily self-consciousness: From statistical regularities to statistical inference. *Annals of the New York Academy of Sciences*, 1426(1), 146–165.
- Ono, M., Miwa, M., & Sasaki, Y. (2015). Word embedding-based antonym detection using thesauri and distributional information. In R. Mihalcea, J. Chai & A. Sarkar (Eds.) *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies*. Association for Computational Linguistics.
- Osgood, C. E. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, 66(3), 171–200.
- Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, 110(22), 8836–8841.
- Parise, C. V. (2016). Crossmodal correspondences: Standing issues and experimental guidelines. *Multisensory Research*, 29(1–3), 7–28. <https://doi.org/10.1163/22134808-00002502>
- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 6104–6108. <https://doi.org/10.1073/pnas.1322705111>
- Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes [Article]. *PLoS One*, 4(5), e5664. <https://doi.org/10.1371/journal.pone.0005664>
- Passaro, L. C., Bondielli, A., & Lenci, A. (2017). Learning affect with distributional semantic models. *IJCoL. Italian Journal of Computational Linguistics*, 3(3–2), 23–36.
- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, 14(2), 119–124.
- Pistolas, E., & Wagemans, J. (2023). Crossmodal correspondences and interactions between texture and taste perception. *i-Perception*, 14(2), 20416695231163473.
- 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. *Journal of Neurolinguistics*, 55, 100895. <https://doi.org/10.1016/j.jneuroling.2020.100895>
- Reinoso-Carvalho, F., Gunn, L., Molina, G., Narumi, T., Spence, C., Suzuki, Y., ter Horst, E., & Wagemans, J. (2020). A sprinkle of emotions vs a pinch of crossmodality: Towards globally meaningful sonic seasoning strategies for enhanced multisensory tasting experiences. *Journal of Business Research*, 117, 389–399. <https://doi.org/10.1016/j.jbusres.2020.04.055>
- 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.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3), 273–294.
- Salgado Montejo, A., Alvarado, J. A., Velasco, C., Salgado, C. J., Hasse, K., & Spence, C. (2015). The sweetest thing: The influence of angularity, symmetry, and the number of elements on shape-valence and shape-taste matches. *Frontiers in Psychology*, 6, 1382.
- Salgado-Montejo, A., Salgado, C. J., Alvarado, J., & Spence, C. (2017). Simple lines and shapes are associated with, and communicate, distinct emotions. *Cognition and Emotion*, 31(3), 511–525. <https://doi.org/10.1080/02699931.2015.1133401>
- Saluja, S., & Stevenson, R. J. (2018). Cross-modal associations between real tastes and colors. *Chemical Senses*, 43(7), 475–480.
- Speed, L. J., & Majid, A. (2020). Grounding language in the neglected senses of touch, taste, and smell. *Cognitive Neuropsychology*, 37(5–6), 363–392. <https://doi.org/10.1080/02643294.2019.1623188>
- Speed, L. J., Vinson, D. P., & Vigliocco, G. (2015). Representing meaning. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 190–211). De Gruyter Mouton. <https://doi.org/10.1515/9783110292022-010>
- Spence, C. (2011). Crossmodal correspondences: A tutorial 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. *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. *Multisensory Research*, 33(1), 1–29. <https://doi.org/10.1163/22134808-20191469>
- Spence, C. (2020b). Olfactory-colour crossmodal correspondences in art, science, and design. *Cognitive Research: Principles and Implications*, 5(1), 52. <https://doi.org/10.1186/s41235-020-00246-1>
- Spence, C. (2020c). Temperature-based crossmodal correspondences: Causes and consequences. *Multisensory Research*, 33(6), 645–682. <https://doi.org/10.1163/22134808-20191494>
- Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences?. *Consciousness and Cognition*, 22(1), 245–260. <https://doi.org/10.1016/j.concog.2012.12.006>
- 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>
- Stewart, A. E. (2007). Linguistic dimensions of weather and climate perception. *International Journal of Biometeorology*, 52, 57–67.
- Stewart, A. E. (2020). Affective normative data for English weather words. *Atmosphere*, 11(8), 860.
- Summer Institute of Linguistics. (2021). *Semantic domains*. Retrieved 09/04/2021 from <https://semdom.org/>.
- Tinga, A. M., Visser-Meily, J. M. A., van der Smagt, M. J., Van der Stigchel, S., van Ee, R., & Nijboer, T. C. W. (2016). Multisensory stimulation to improve low- and higher-level sensory deficits after stroke: A systematic review. *Neuropsychology Review*, 26, 73–91.
- 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., Escobar, F. B., Spence, C., & Olier, J. S. (2023). The taste of colours. *Food Quality and Preference*, 112, 105009.
- Velasco, C., & Obrist, M. (2020). *Multisensory experiences: Where the senses meet technology*. Oxford University Press.
- Velasco, C., Salgado-Montejo, A., Marmolejo-Ramos, F., & Spence, C. (2014). Predictive packaging design: Tasting shapes, typefaces, names, and sounds. *Food Quality and Preference*, 34, 88–95. <https://doi.org/10.1016/j.foodqual.2013.12.005>
- Velasco, C., & Spence, C. (2019). *Multisensory packaging: Designing new product experiences*. Springer.
- 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. (2016a). The semantic basis of taste-shape associations. *PeerJ*, 2016(2), 1644. <https://doi.org/10.7717/peerj.1644>
- Velasco, C., Woods, A. T., Petit, O., Cheok, A. D., & Spence, C. (2016b). Crossmodal correspondences between taste and shape, and their implications for product packaging: A review. *Food Quality and Preference*, 52, 17–26. <https://doi.org/10.1016/j.foodqual.2016.03.005>
- 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.
- Walker, P. (2016). Cross-sensory correspondences: A theoretical framework and their relevance to music. *Psychomusicology: Music, Mind, and Brain*, 26(2), 103.
- Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception*, 13(1), 75–81.
- 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. *Frontiers in Psychology*, 5, 1365. <https://doi.org/10.3389/fpsyg.2014.01365>
- Wang, Q. J., Wang, S., & Spence, C. (2016). “Turn up the taste”: Assessing the role of taste intensity and emotion in mediating crossmodal correspondences between basic tastes and pitch. *Chemical Senses*, 41(4), 345–356.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191–1207.
- Whiteford, K. L., Schloss, K. B., Helwig, N. E., & Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues. *i-Perception*, 9(6), 2041669518808535.

Winter, B. (2019). *Sensory linguistics*. John Benjamins B.V. (pp. 1–303).

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179, 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects [Article]. *Attention, Perception, and Psychophysics*, 81(5), 1609–1623. <https://doi.org/10.3758/s13414-019-01668-w>

---

**Cite this article:** Alvarado, J. A., Velasco, C., & Salgado, A. (2024). The organization of semantic associations between senses in language, *Language and Cognition*, 1–30. <https://doi.org/10.1017/langcog.2024.19>