

## Latent semantics of action verbs reflect phonetic parameters of intensity and emotional content

**Petersen, Michael Kai**

*Published in:*  
P L o S One

*DOI:*  
[10.1371/journal.pone.0121575](https://doi.org/10.1371/journal.pone.0121575)

*Publication date:*  
2015

*Document Version*  
Early version, also known as pre-print

[Link to publication](#)

### *Citation (APA):*

Petersen, M. K. (2015). Latent semantics of action verbs reflect phonetic parameters of intensity and emotional content. P L o S One, 10(4), [ e0121575]. DOI: 10.1371/journal.pone.0121575

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# When push comes to shove verbs literally shake due to latent semantic parameters of size and intensity

Michael Kai Petersen\*

Cognitive Systems  
 DTU Compute, Building 324  
 Technical University of Denmark  
 DK-2800 Kgs.Lyngby, Denmark  
 \* mkai@dtu.dk

## Abstract

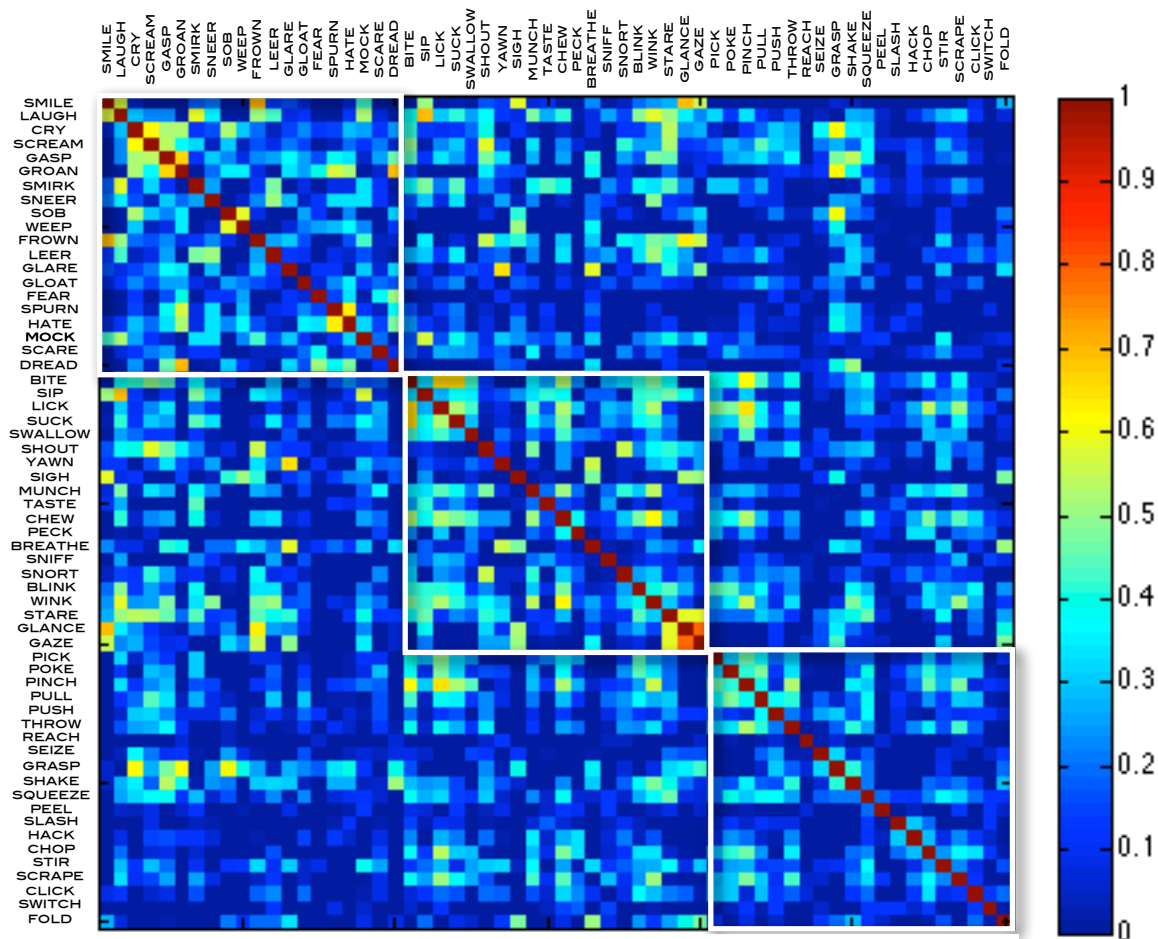
The ability to predict which patterns are formed in brain scans when imagining a celery or an airplane, based on how these concepts as words co-occur in texts, suggests that it is possible to model mental representations based on word statistics. Whether counting how frequently nouns and verbs combine in Google search queries, or extracting eigenvectors from matrices made up of Wikipedia lines and Shakespeare plots, these latent semantics approximate the associative links that form concepts. However, cognition is fundamentally intertwined with action; even passively reading verbs has been shown to activate the same motor circuits as when we tap a finger or observe actual movements. If languages evolved by adapting to the brain, sensorimotor constraints linking articulatory gestures with aspects of motion might also be reflected in the statistics of word co-occurrences. To probe this hypothesis  $3 \times 20$  emotion, face, and hand related verbs known to activate premotor areas in the brain were selected, and latent semantic analysis LSA was applied to create a weighted adjacency matrix. Hierarchically clustering the verbs and modeling their connectivity within a force directed graph, they divide into modules of mouth and hand motion, facial expressions and negative emotions. Transforming the verbs into their constituent phonemes, the corresponding consonant vowel transitions can be represented in an articulatory space defined by tongue height and formant frequencies. Here the vowels appear positioned along a front to back continuum reflecting aspects of size and intensity related to the actions described by the verbs. More forceful verbs combine plosives and sonorants with fricatives characterized by sustained turbulent airflows, while positive and negative emotional expressions tend to incorporate up- or downwards shifts in formant frequencies. Suggesting, that articulatory gestures reflect parameters of size and intensity which might be retrieved from the latent semantics of action verbs.

## Introduction

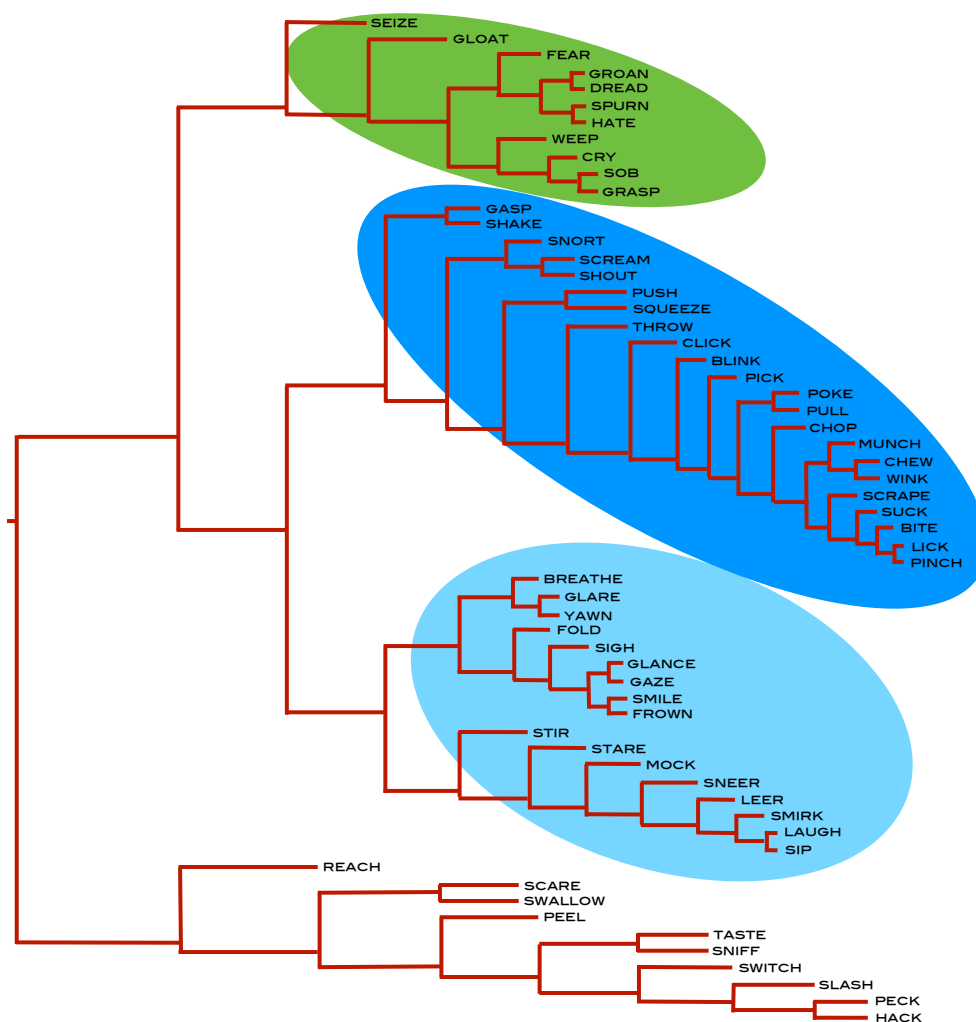
Aspects of motion are fundamental in cognition; spatiotemporal constraints define how we internally represent affordances for potential action, and perceptual states seem to be reenacted from memory traces formed by sensorimotor circuits [1]. Adding to a growing amount of evidence for embodied cognition [2], where not only action verbs like ‘push’ are associated with trajectories, but also terms like ‘argue’ and ‘respect’ appear to be grounded in a conceptual space framed by horizontal and vertical axes [3]. Spatial metaphors are ubiquitous in phrases like ‘hitting the road’, ‘bouncing back’ or ‘thinking out of the box’, where we reinterpret ourselves as colliding objects that are subject to forces of gravity or moving along virtual time lines [4]. Anatomically speaking, parts of Broca’s area (BA 44) are involved in shaping both language and gestures, as motor areas in the brain representing the dominant hand are co-activated in both spontaneous speech and reading [5]. Concrete verbs and nouns seem to be combine within action schemas that semantically link perception with objects, rather than being differentially processed accord-

ing to their respective lexical categories [6]. Passively reading verbs like ‘kick’, ‘pick’ and ‘lick’, has been found to activate premotor areas in the brain associated with the respective movements [7] [8], while Mu brainwave oscillations become desynchronized over the sensorimotor cortex similar to when we imagine tapping a finger [9]. Neural processing within the motor circuits seems to be shared with language, as working memory for action verbs like ‘seize’ or ‘chop’ become impaired if simultaneously moving the hand [10]. In line with neuroimaging studies showing that not only neurons in the premotor cortex but also in the primary motor cortex are firing during the processing of action verbs [11] [12]. Although it might be argued that cognitive modeling based on word statistics is unrelated to embodied semantics [13], sensorimotor constraints associated with words repeatedly encountered in multiple contexts could have been woven into the surface structure of language [14]. Thus providing a semantic bootstrapping that would facilitate language learning through shared distributional and phonological cues [15]. If aspects of action based language have through Hebbian learning been associated with sequences of verbs and nouns [16], the underlying parameters of motion might potentially also be retrieved from the latent semantics of action verbs. Selecting  $3 \times 20$  emotion, face, and hand related verbs known to activate motor circuits in the brain [17], and applying latent semantic analysis LSA [18] their mutual cosine similarities were defined in an adjacency matrix, based on a large-scale text corpus consisting of 22829 words found in 67380 excerpts of Harvard Classics literature, Wikipedia articles and Reuters news [19]. Hierarchically clustering the verbs and modeling their connectivity within a force-directed graph [20], the emerging network components were compared against user rated word norms of valence and arousal [21], defining their emotional polarity and perceived intensity [22]. Subsequently transforming the verb clusters into ARPAbet phonemes using the CMU text to speech pronunciation dictionary [23], as well as acoustic features defined by their average F1 and F2 formant frequencies [24], the primary stress vowels of the action verbs can be represented in an articulatory space defined by tongue height and front-back position related to the international phonetic alphabet (IPA).

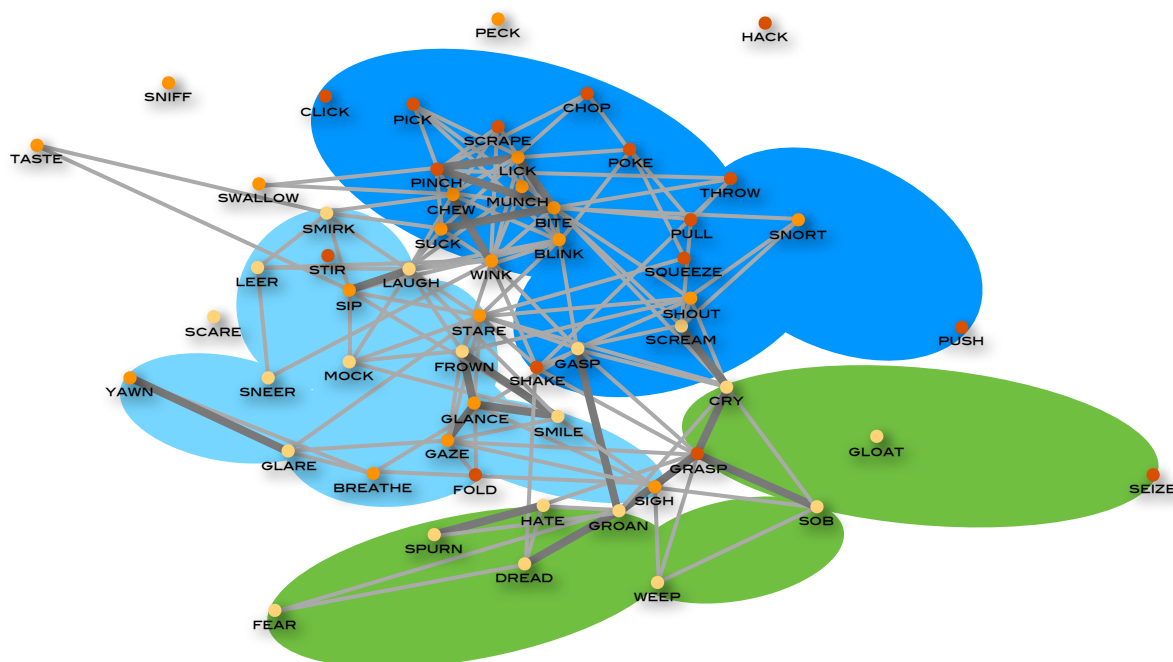
## Results



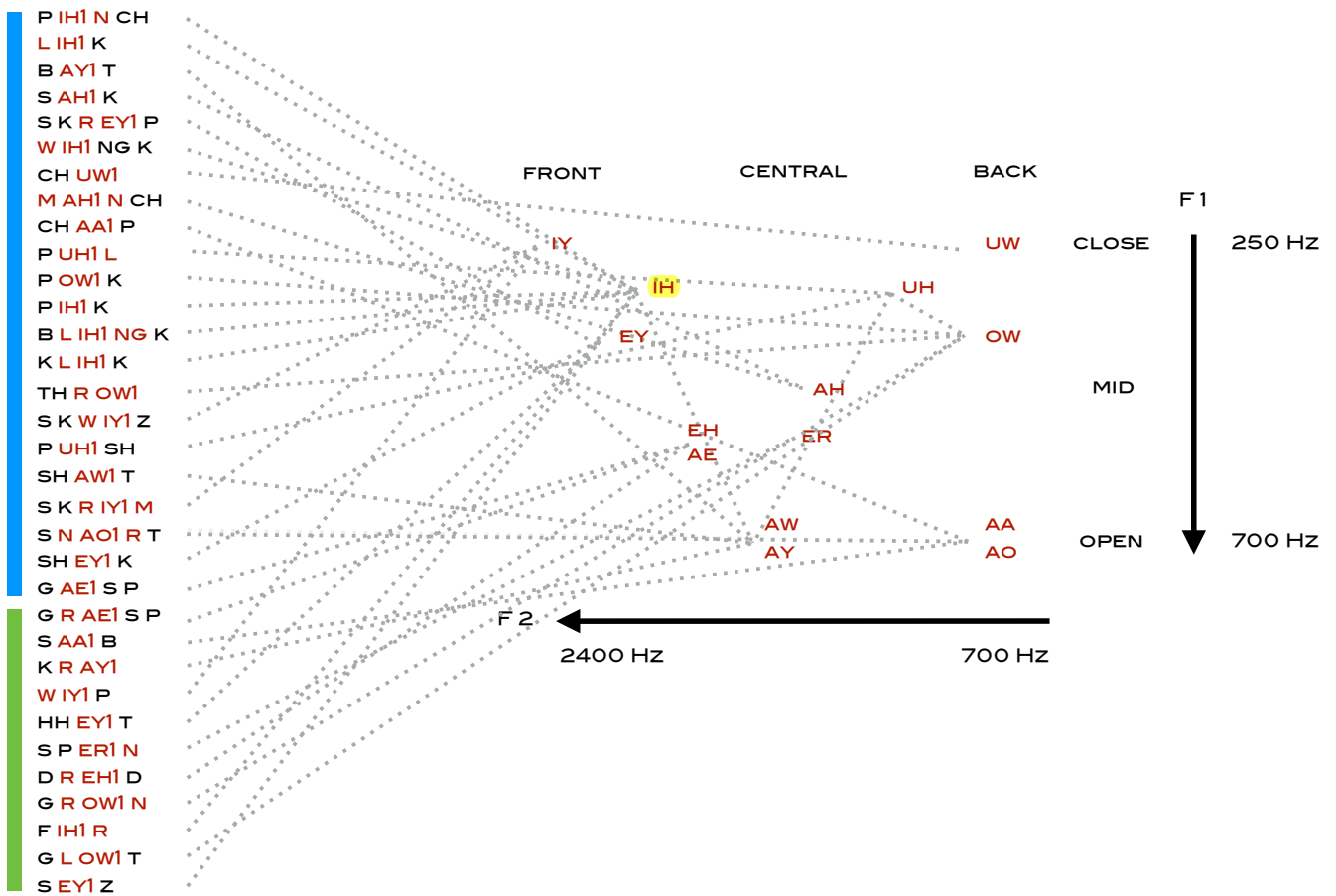
**Figure 1.** Adjacency matrix of  $3 \times 20$  emotion, face and hand action verbs, illustrating their mutual cosine similarities generated by applying LSA latent semantic analysis and reducing the dimensionality to the 125 most significant eigenvalues. The latent semantic relations are based on the HAWIK text corpus, where the original term document matrix consists of 22829 words found in 67380 excerpts of Harvard Classics literature, Wikipedia articles and Reuters news. In the sparsely activated adjacency matrix emotional category verbs such as ‘smile’, ‘laugh’ and ‘frown’ co-occur with facial eye movement verbs like ‘gaze’ ‘stare’ and ‘glance’, but are almost orthogonal to representations of hand related verbs such as ‘pick’, ‘push’ or ‘poke’. Whereas facial verbs related to the jaw and tongue motion such as ‘bite’, ‘lick’ or ‘suck’ are associated with hand movements like ‘pinch’, ‘chop’ and ‘scrape’.



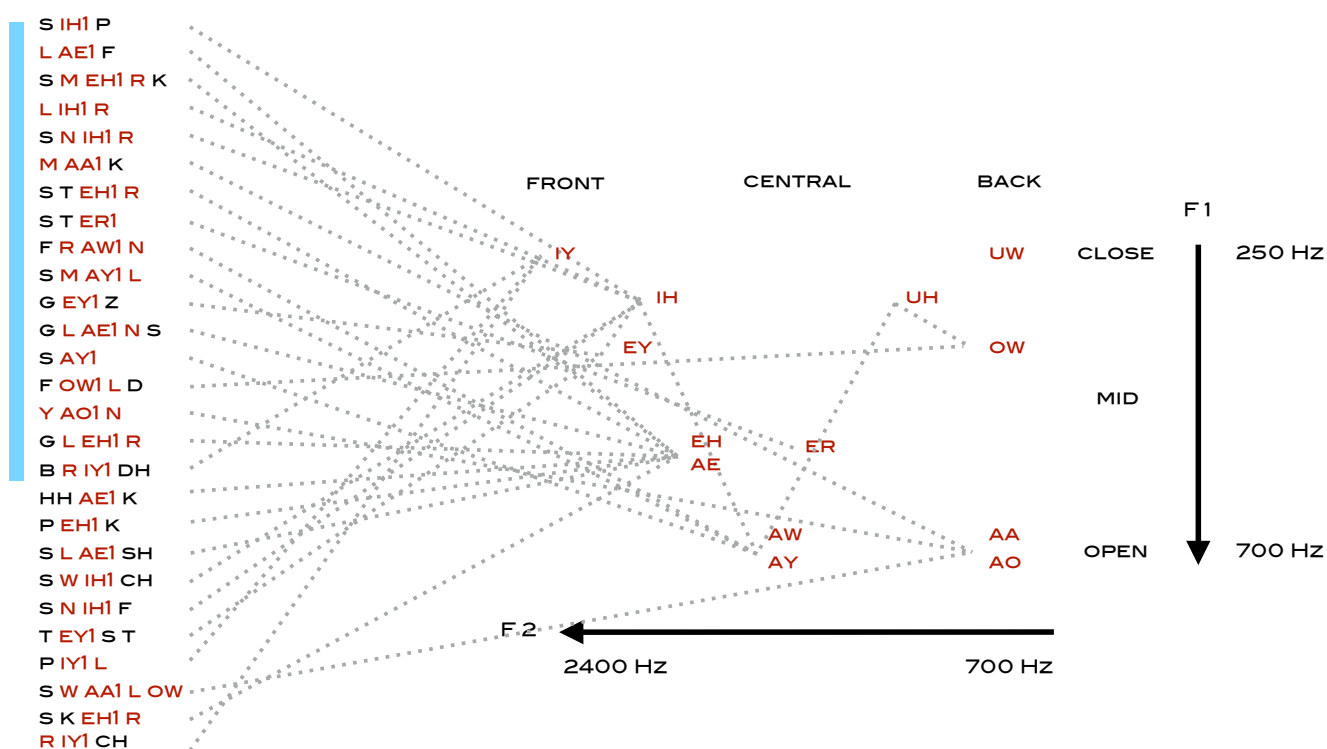
**Figure 2. Hierarchical clustering of  $3 \times 20$  emotion, face and hand action verbs**, based on their mutual LSA cosine similarities generates four clusters: 1. **Negative emotional verbs (green cluster)** characterized by low valence expressions such as ‘hate’ (1.96, SD 1.33), ‘sob’ (2.65, SD 1.81), ‘weep’ (2.88, SD 2.07), ‘cry’ (3.22, SD 2.41), ‘fear’ (2.93, SD 1.79), ‘dread’ (3.00, SD 1.89). 2. Combined mouth and hand motion verbs (blue cluster), characterized by increasing levels of arousal ranging from small size finger precision grip and oscillatory jaw motion as in ‘click’ (2.81, SD 2.20), ‘pick’ (3.62, SD 2.25), ‘munch’ (3.62, SD 1.96), ‘chew’ (3.80, SD 2.24), to whole hand object manipulation like ‘chop’ (4.43, SD 2.29), ‘throw’ (4.52, SD 2.29), ‘poke’ (5.41, SD 2.70), ‘shake’ (5.20, SD 2.71), and forceful expressions such as ‘gasp’ (5.61, SD 2.41), ‘shout’ (6.29, SD 2.05) and ‘scream’ (6.74, SD 1.66). 3. Facial motion verbs (cyan cluster) ranging from low valence expressions like ‘sneer’ (3.30, SD 1.92), ‘frown’ (3.35, SD 1.35), ‘glare’ (3.70, SD 1.59), ‘mock’ (3.81, SD 1.57) ‘stare’ (4.45, SD 1.61), to high valence verbs such as ‘glance’ (5.71, SD 1.65), ‘gaze’ (6.15, SD 1.27), ‘breathe’ (7.17, SD 1.69), ‘laugh’ (7.56, SD 2.64), ‘smile’ (7.89, SD 2.19). 4. **High velocity verbs (transparent cluster)** characterized by increasing levels of arousal ranging from ‘switch’ (3.90, SD 2.10), ‘sniff’ (4.95, SD 2.13), ‘hack’ (5.48, SD 1.91), ‘slash’ (5.65, SD 2.81) to ‘scare’ (7.10, SD 2.13).



**Figure 3. Force directed graph based on mutual cosine similarities of action verbs**, where nodes are thresholded at values above 0.2, while light and bold edges denote LSA cosine similarities above 0.4 and 0.6. Partitioning the network based on hierarchical clustering, the graph is characterized by densely clustered maximum cliques ( $\omega(G) = 12$ , average clique size = 7) grouping action verbs of increasing intensity ranging from small size motion such as ‘pinch’, ‘pick’, ‘lick’, ‘bite’, ‘suck’, ‘chew’, ‘wink’ and ‘blink’, to more forceful gestures like ‘pull’, ‘poke’, ‘throw’, ‘chop’ or ‘scrape’ (arousal 2.81 - 6.74,  $M = 4.47$ , blue component). This component is sparsely connected to the other network modules characterized by low velocity facial expressions (valence 3.30 - 7.89,  $M = 5.34$ , cyan component), negative emotions (valence 1.96 - 5.45,  $M = 3.11$ ), and less densely clustered high velocity gestures of short duration (arousal 3.35 - 7.10,  $M = 4.56$ , transparent background). The nodes with the highest eigenvector centrality values ‘wink’ (0.25), ‘bite’ (0.23) and ‘laugh’ (0.23) function as hubs connecting the combined mouth and hand action verbs (blue component) with the low velocity facial expressions ‘smile’, ‘sigh’, ‘frown’, ‘laugh’, ‘smirk’ and ‘mock’ as well as eye motion like ‘glance’, ‘gaze’, ‘stare’, ‘glare’ and ‘leer’ (cyan component). While the nodes with the highest betweenness centrality ‘scream’ (0.43) and ‘gasp’ (0.56), channel the largest number of shortest paths forming the links between the combined and hand and mouth related gestures (blue component), and the less densely clustered subgraph of negative emotional verbs such as ‘cry’, ‘grasp’, ‘sob’, ‘weep’, ‘groan’, ‘hate’, ‘spurn’, ‘dread’ and ‘fear’ (green component).



**Figure 4. Articulatory projections of primary stress vowels in clusters of action verbs,** where vocal gestures are mapped according to tongue height, front-back position and rounding, while auditory features are defined by the F1 and F2 formant frequencies in the sound spectrum of the voice. Within the combined mouth and hand action verbs (blue cluster) small size gestures like ‘P IH1 K’ (arousal 3.62, SD 2.25) and ‘K L IH1 K’ (arousal 2.81, SD 2.20) are articulated using high frontal ‘IH’ vowels, which acoustically result in higher F2 values that are maximally dispersed from the F1 formants. In contrast to more forceful action verbs like ‘P UH1 L’ (arousal 4.10, SD 2.47) and ‘P UH1 SH’ (arousal 4.40, SD 2.78) articulated by back ‘UH’ vowels as well as the diphthongs ‘OW’ in ‘P OW1 K’ (arousal 5.41, SD 2.70) and ‘TH R OW1’ (arousal 4.52, SD 2.29), which acoustically have a small gap between the F2 and F1 formant frequencies. Open jaw diphthong transitions and vowels characterize aroused actions produced by voiced plosives B and G like ‘B AY1 T’ (arousal 5.10, SD 2.31) and ‘G AE1 S P’ (arousal 5.61, SD 2.41). Sustained tension is emphasized by the turbulent airflow generated by fricatives such as S and SH in ‘S K R EY1 P’ (arousal 4.50, SD 2.28) ‘SH EY1 K’ (arousal 5.20, SD 2.71), ‘S AH1 K’ (arousal 5.6, SD 2.19), ‘S K R IY1 M’ (arousal 6.74, SD 1.66) and ‘SH AW1 T’ (arousal 6.29, SD 2.05). Several of the negative emotion verbs (green cluster) are characterized by back vowels and diphthongs as in ‘S AA1 B’ (valence 2.65, SD 1.81) ‘G L OW1 T’ (valence 3.68, SD 1.11), ‘G R OW1 N’ (valence 3.90, SD 1.59) as well as R liquid consonants like ‘D R EH1 D’ (valence 3.00, SD 1.89) and ‘F IH1 R’ (valence 2.93, SD 1.79).



**Figure 5. Articulatory projections of primary stress vowels in clusters of action verbs,** where vocal gestures are mapped according to tongue height, front-back position and rounding, while auditory features are defined by the F1 and F2 formant frequencies in the sound spectrum of the voice. In the action verbs describing low velocity facial expressions (cyan cluster), dynamic shifts in formant frequencies appear associated with positive and negative emotions. Exemplified by the upward F2 frequency transition in the diphthongs of ‘S M AY1 L’ (valence 7.89, SD 2.19) versus the downwards shift in ‘F R AW1 N’ ( valence 3.35, SD 1.35). Downward frequency shifts due to the lowered F3 third formant characteristic of the liquid consonant R, appear reflected in negative emotions such as ‘S N IH1 R’ (valence 3.30, SD 1.92), ‘G L EH1 R’ (valence 3.70, SD 1.59) and ‘S K EH1 R’ (valence 3.55, SD 2.11). Whereas the action verbs describing high volicy motion of short duration (transparent cluster), are emphasized by the turbulent airflows generated by the fricatives ‘S ’, ‘F ’, ‘SH ’, ‘HH ’ and affricate ‘CH ’ as in ‘S W IH1 CH’ (arousal 3.90, SD 2.10), ‘S N IH1 F’ (arousal 4.95, SD 2.57), ‘HH AE1 K’ (arousal 5.48, SD 1.91), ‘S L AE1 SH’ (arousal 5.65, SD 2.81).



## Discussion

Applying statistical approaches like LSA generates matrices of lower dimensionality, in which words that have similar meanings in different contexts are squeezed into a reduced number of rows and columns, corresponding to eigenvectors which capture orthogonal directions in the original high dimensional semantic space. In the adjacency matrix which captures the cosine similarities between action verbs (Figure 1), emotional category verbs such as the low velocity sustained expressions ‘smile’, ‘laugh’ and ‘frown’ are coupled with facial eye movement verbs like ‘gaze’ ‘stare’ and ‘glance’, but are almost orthogonal to more forceful hand motion related verbs such as ‘pick’, ‘push’ or ‘poke’. Whereas facial verbs related to cyclical jaw and tongue motion such as ‘bite’, ‘lick’ or ‘suck’ trigger hand movements like ‘pinch’, ‘chop’ and ‘scrape’. These latent semantic links between hand, mouth as well as eyelid opening and closing action verbs resemble the co-activations of gestures found in motor maps in the brains of monkeys. Rather than representing individual movements they store prototypical sequences of connected hand to mouth gestures involved when eating, or body postures related to more forceful manipulation of objects [25]. It has been proposed that such hierarchical coordination of movements, might through Hebbian learning have been associated with the sequences of verbs and nouns that make up action based language [16], being constrained by the physical parameters of distance and gravity which define how we interact with objects [26]. Single neuron recordings in monkeys indicate that sequences of gestures form a vocabulary of motor schemas, which are recursively combined into object oriented motion patterns. Thus reducing the large number of dimensions involved when manipulating objects to a few parameters of orientation and size related to stored representations of motion patterns [27]. If such sensorimotor parameters are encoded in language they might potentially be retrieved using LSA. Earlier studies have documented that horizontal and vertical dimensions are encoded in language to the degree that it is feasible to reconstruct the geographical layout of cities from how they contextually co-occur in news articles [28] or are described in fiction like “Lord of the Rings” based on LSA [29]. Likewise perceptuomotor aspects encoded in language have been retrieved using LSA [14], ranging from vertical movement and scaling of objects [30] to aspects of motor resonance in manual rotation [31].

When analyzing complex networks like the brain, hierarchical clustering is frequently applied to find heavily interconnected subgraphs. These are typically only sparsely linked to other network components, which are in turn connected through intermediary nodes functioning as hubs [32]. Applying hierarchical clustering to the adjacency matrix (Figure 1), the most similar pairs of action verbs were iteratively merged into a tree structure. In the resulting dendrogram shorter horizontal lines between leaf nodes indicate higher degree of similarity, and the length of branches signify the tightness of the cluster (Figure 2). Subsequently the action verbs were annotated using the “Norms of valence, arousal and dominance for 13915 English lemmas” available online [21], which define how pleasant, intense and controlled the words are perceived as being based on user ratings on a scale from 1 to 9 [22]. Taking these psychological dimensions into consideration, the hierarchical clustering appears to capture the increasing intensity in the concrete motion action verbs (blue cluster), as reflected in the perceived values of arousal, ranging from small size finger precision grip and jaw motion to large scale gestures incorporating the arms and upper body (arousal 2.81 - 6.74,  $M = 4.47$ ). Yet other aspects of motion define the less densely grouped rapid movements (transparent cluster) characterized by high velocity motion in gestures of short duration (arousal 3.35 - 7.10,  $M = 4.56$ ). In contrast to the more abstract action verbs, where emotional polarity defined along the parameter of valence separates the cyan cluster of low velocity facial expressions (valence 3.30 - 7.89,  $M = 5.34$ ) from the green cluster of mostly negative emotions (valence 1.96 - 5.45,  $M = 3.11$ ). In line with cognitive psychology studies indicating that sensorimotor elements remain statistically more significant for the representation of concrete actions, whereas abstract concepts rely increasingly on affective associations the more abstract they are perceived as being [33]. The hierarchical clustering of negative action verbs (green cluster) versus the more positive facial expressions (cyan) highlights their contrasting polarities in relation to valence. Polarity can be interpreted as not only defining positive or negative features of a concept, but also providing a fundamental foundation for adaptive behavior and

reward mechanisms. Possibly explaining why these positive / negative contrasts become so consolidated that the recall of antonyms might still be preserved in aphasiac stroke patients [34].

Neuroimaging studies have demonstrated that passively reading emotional verbs activate not only motor circuits controlling facial muscles but also hand and arm gestures which might contextually facilitate comprehension of affective expressions [35] [17]. In order to further explore whether such couplings might also be reflected in the connections between action verbs, a force directed algorithm was applied to construct a graph based on their mutual cosine similarity, where the nodes are repositioned until reaching a mechanical equilibrium (Figure 3) [20]. Partitioning the network based on hierarchical clustering, the graph is characterized by densely clustered maximum cliques ( $\omega(G) = 12$ , average clique size = 7) grouping action verbs of increasing intensity ranging from small size motion such as ‘pinch’, ‘pick’, ‘lick’, ‘bite’, ‘suck’, ‘chew’, ‘wink’ and ‘blink’, to more forceful gestures like ‘pull’, ‘poke’, ‘throw’, ‘chop’ or ‘scrape’. Neuroimaging studies indicate that canonical neurons which respond differentially to picking up something with two fingers or grasping it using the whole hand, also fire when viewing correspondingly small or large objects. Suggesting that aspects of gravity and size appear to be combined into affordances for potential motor actions [36]. Eigenvector centrality is often used to assess which nodes function as hubs linking the modules within a graph, when modeling functional brain connectivity [37] and semantic word associations [38]. Similar to Google’s PageRank algorithm [39], it considers not only the number of links, but also whether these connections between nodes are themselves significant within the network. The nodes with the highest eigenvector centrality values ‘wink’ (0.25), ‘bite’ (0.23) and ‘laugh’ (0.23) function as hubs connecting the combined mouth and hand action verbs (blue component) with the low velocity facial expressions ‘smile’, ‘sigh’, ‘frown’, ‘laugh’, ‘smirk’ and ‘mock’ as well as eye motion like ‘glance’, ‘gaze’, ‘stare’, ‘glare’ and ‘leer’ (cyan component). While the nodes with the highest betweenness centrality ‘scream’ and ‘gasp’ channel the largest number of shortest paths forming the links to the less densely clustered subgraph of negative emotional verbs (green component). Prototypical emotions are in affective computing experiments often assessed by measuring the amount of muscle activity. Either related to zygomaticus major AU12, which is activated when raising the corners of the lips upwards in a ‘smile’, or risorius AU20 when laterally pulling them apart in a ‘cry’. In both cases the mouth opening is coupled with the muscle activity around the eyes; the so-called Duchenne constriction, which was earlier interpreted as a unique marker of spontaneous positive emotion. However, the degree of eye constriction and mouth opening has been found to be correlated with the perceived intensity of the facial expressions in infants, regardless of whether they are associated with a ‘smile’ or a ‘cry’ [40]. Understood in that context, the degree of mouth opening and eye constriction might also be discernible in the word norm arousal ratings of the negative emotions (green component), going from from a tight lipped ‘weep’ (arousal 4.00 SD 2.47) to an increasingly wider open ‘sob’ (arousal 4.89 SD 2.76) and ‘cry’ (arousal 5.45 SD 2.82). As well as within the low velocity facial expressions (cyan component), where the word norm arousal ratings might reflect the increasing amount of mouth opening described by the action verbs ‘frown’ (arousal 3.61 SD 2.17) ‘smile’ (arousal 4.62 SD 3.09) ‘smirk’ (arousal 4.70 SD 2.47) and ‘laugh’ (arousal 6.62 SD 1.91).

Sensorimotor connections in the brain linking perception of shapes and motion, may similarly have constrained how aspects of size and intensity are mapped onto the consonants and vowels of words [41]. Underlying structural dimensions of size seem hardwired into speech articulation, as grasping objects of increasing size has been shown to simultaneously enlarge both the lip kinematics and mouth aperture when pronouncing vowels [42]. Behavioral studies have shown that high front vowels are perceived as lighter and associated with smaller organisms than words involving back vowels [43]. Even preverbal 3-4 months old infants seem to link contrasts between high and low pitch to vertically moving balls [44]. Likewise correspondences between articulatory gestures and the shapes of objects have been found already in toddlers, who associate back produced vowels as in ‘bouba’ with rounded forms and link bright front vowels such as ‘kiki’ to edgy outlines [45]. Exploring whether such couplings between phonemes and physical parameters are reflected in the hierarchical clusters, the action verbs were transformed into ARPAbet phonemes [23]. Next, the phonemes constituting their primary stress vowels were projected

into an articulatory space related to the international phonetic alphabet (IPA) and the corresponding acoustical F1 and F2 formant frequencies [46]. Behavioral studies have shown that high front vowels are perceived as lighter and associated with smaller organisms than words involving back vowels [43]. Essentially, the varying positions of the jaw and the vocal tract articulators of lips, velum, larynx and tongue provide a framework for transforming articulatory gestures into phonetic structures [47] [48]. Although phonemes demand complex coordination of articulatory gestures, neuroimaging studies have established that most of the variance during pronunciation of plosives such as ‘P’, ‘T’ and ‘K’ are explained by tongue height and front versus back position. Likewise for consonant vowel transitions the main contrasts within an articulatory space are between frontal unrounded ‘IH’ versus the back rounded ‘UH’ sonorants [49]. Such contrasts also come out within the clustered mouth and hand action verbs (Figure 4, blue cluster), as small size gestures like ‘P IH1 K’ (arousal 3.62, SD 2.25) and ‘K L IH1 K’ (arousal 2.81, SD 2.20) are articulated using high frontal ‘IH’ vowels, which acoustically result in higher F2 values that are maximally dispersed from the F1 formants. While more forceful action verbs like ‘P UH1 L’ (arousal 4.10, SD 2.47) and ‘P UH1 SH’ (arousal 4.40, SD 2.78) are articulated by back ‘UH’ vowels as well as the diphthongs ‘OW’ as in ‘P OW1 K’ (arousal 5.41, SD 2.70) and ‘TH R OW1’ (arousal 4.52, SD 2.29), which acoustically have a small gap between the F2 and F1 formant frequencies. Meaning, that the dispersion between the primary formant frequencies might acoustically be perceived as a cue for contrasts such as light contra heavy, and soft versus hard [50].

It has been proposed that the articulatory features of plosives, sonorants and fricatives, may themselves have evolved by mimicking the sounds that occur in nature when solid objects collide, resonate, or slide against a surface [51]. Suggesting, that phonemes as the building blocks of speech reuse neural circuits for making sense of auditory events. Similar to how written languages appear to have adapted to the brain, by recombining frequently occurring low level visual features into alphabetic characters [52] [53]. Understood in that context, small size hand gestures initiated by unvoiced plosives that extend the gap before the sonorant as in ‘P IH1 N CH’, acoustically resemble the impact of soft objects with a flexible texture. While the harder attack of voiced plosives that reduce the onset before the sonorant as in ‘B AY1 T’, creates a resonance similar to collisions of larger more rigid structures [51]. In contrast to the brief attacks of plosives, fricatives create a feeling of sustained tension caused by the turbulence generated when the flow of air is directed towards the teeth, like ‘SH’ in ‘SH EY1 K’ and ‘SH AW1 T’. Or when forcing the air over the edge of the teeth as in the fricative ‘S’ in verbs like ‘S K R EY1 P’ and ‘S K R IY1 M’. A simplified representation of articulatory features could thus be understood as a continuum going from open jaw resonant diphthongs like ‘AY’ in ‘cry’ to near close frontal vowels such as ‘IY’ in ‘weep’ or a back rounded ‘UW’ in ‘chew’. These sonorants undergo a phase shift as the airflow turns turbulent in fricatives such as ‘S’ and ‘Z’ in ‘squeeze’, or when the airstream is abruptly cut off by plosives like ‘B’ and ‘K’ in ‘blink’. These parameters of intensity appear also reflected in the phonemes of the action verbs, as more open jaw diphthong transitions and vowels characterize aroused actions produced by voiced plosives B and G like ‘B AY1 T’ (arousal 5.10, SD 2.31) and ‘G AE1 S P’ (arousal 5.61, SD 2.41). Sustained tension is emphasized by the turbulent airflow generated by fricatives such as S and SH in ‘S K R EY1 P’ (arousal 4.50, SD 2.28) ‘SH EY1 K’ (arousal 5.20, SD 2.71), ‘S AH1 K’ (arousal 5.6, SD 2.19), ‘S K R IY1 M’ (arousal 6.74, SD 1.66) and ‘SH AW1 T’ (arousal 6.29, SD 2.05). Several of the negative emotion verbs (Figure 4, green cluster) are characterized by back vowels and diphthongs as in ‘S AA1 B’ (valence 2.65, SD 1.81) ‘G L OW1 T’ (valence 3.68, SD 1.11), ‘G R OW1 N’ (valence 3.90, SD 1.59) as well as R liquid consonants like ‘D R EH1 D’ (valence 3.00, SD 1.89) and ‘F IH1 R’ (valence 2.93, SD 1.79). From an acoustic perspective, the corresponding auditory cues which are related to the amount of dispersion between the F2 and F1 formant frequencies, might also contribute to defining the emotional polarity. Dynamically lowering the pitch in vowels is perceived as threatening in human speech sounds, while upwards moving formant transitions are to a larger degree associated with positive emotions [43]. In the action verbs such up- or downward shifts in pitch of the F2 formants are evident in the diphthongs of facial action verbs (Figure 5, cyan cluster) as in ‘S M AY1 L’ (valence 7.89, SD 2.19)

versus ‘F R AW1 N’ (valence 3.35, SD 1.35). Downward frequency shifts due to the lowered F3 third formant characteristic of the liquid consonant ‘R’, appear reflected in negative emotions such as ‘S N IH1 R’ (valence 3.30, SD 1.92), ‘G L EH1 R’ (valence 3.70, SD 1.59) and ‘S K EH1 R’ (valence 3.55, SD 2.11). Whereas the action verbs describing high velocity motion of short duration (Figure 5, transparent cluster), implement sustained tension generated by the fricatives ‘S’, ‘F’, ‘SH’, ‘HH’ and affricate ‘CH’ as in ‘S W IH1 CH’ (arousal 3.90, SD 2.10), ‘S N IH1 F’ (arousal 4.95, SD 2.57), ‘HH AE1 K’ (arousal 5.48, SD 1.91), ‘S L AE1 SH’ (arousal 5.65, SD 2.81).

Whether language is seen as rooted in symbolic associations constituted by statistical word representations, or being grounded in simulation literally dependent on sensorimotor circuits, there is an emerging consensus on the need to adapt a pluralist view about embodiment and semantics [13] [54]. Over the past decade a growing number of studies indicate that projecting semantics onto matching articulatory gestures facilitate language learning [55]. Such audiovisual synaesthetic mappings appear cross-culturally, as also 3 year old Japanese speaking infants learn verbs faster when associating video clips of forceful contra light motion with syllables like ‘batobato’ versus ‘chokachoka’ [56]. Even native English speaking learn pairs of antonyms in Japanese faster when the speech sounds match the meaning of the actual terms [57]. Even though a limited number of parameters might suffice to model the underlying latent semantics it remains a daunting task. Depending on the context, comprehension involves dynamically fluctuating layers of interaction between perceptuomotor processes and the mental imagery conjured up by abstract representations grounded in long term memory [58]. On the other hand, capturing these complex patterns of features occurring within multiple contexts, might actually now be feasible not only based on existing large scale text corpora, but also by taking advantage of the massive amounts of data continuously being generated within web search and social media. Combining latent semantics and articulatory gestures might thus longer term enable us to model not only how actions relate to objects but also how our inner states are linked to perception [59], constrained by sensorimotor parameters in a space encompassing the extremes of emotional contrasts.

## Methods

Initially selecting  $3 \times 20$  hand, face and emotion related action verbs previously used in an EEG electroencephalography experiment [60], constituting half of the action verbs similarly used in a fMRI functional magnetic resonance neuroimaging study, demonstrating that the selected action verbs activated premotor cortices in the brain during a passive reading task [17], we apply latent semantic analysis LSA [61] [18] in order to retrieve an adjacency matrix based on the HAWIK text corpus consisting of 22829 words found in 67380 excerpts of Harvard Classics literature, Wikipedia articles and Reuters news [19]. Using singular value decomposition SVD to reduce dimensionality [62], the original  $m \times n$  term-document matrix  $\mathbf{X}$  is decomposed into a product of three other matrices:

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$$

where the  $\mathbf{U}$  matrix, similar to the original matrix has  $m$  rows of words, while the columns now consist of  $r$  eigenvectors representing the principal components in the data. Likewise the transpose of the orthonormal matrix  $\mathbf{V}^T$  has as before  $n$  columns of documents but now related to  $r$  rows of eigenvectors or principal components. The very purpose of the decomposition is to scale down the number of parameters based on a  $\mathbf{\Lambda}$  square matrix containing  $r$  singular values  $\lambda$  arranged along the diagonal in decreasing order, which as eigenvalues scale the eigenvectors of the rectangular matrices to each other and thereby derive a matrix of reduced dimensionality:

$$\mathbf{Z}_k = \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{V}_k^T$$

where only the  $k$  largest singular values of the  $\mathbf{\Lambda}$  diagonal matrix are retained. As a result the number of parameters in the rectangular  $\mathbf{U}_k$  and  $\mathbf{V}_k^T$  matrices are reduced to what would correspond to the principal components containing the highest amount of variance in the matrix. Thus allowing us to reconstruct the original input based on a  $\mathbf{Z}_k$  matrix of lower dimensionality which is embedding the underlying structure of the data. Geometrically speaking, the terms and documents in the condensed  $\mathbf{Z}_k$  matrix can be interpreted as points in a  $k$  dimensional subspace, which enables us to calculate the degree of similarity between matrices based on the dot or inner product of their corresponding vectors. Interpreting the matrix multiplication geometrically the cosine similarity between two words represented by their vectors can be expressed as

$$\cos \theta = \frac{x \cdot y}{\|x\| \|y\|}$$

where  $x \cdot y$  signifies the dot product of the vectors, and  $\|x\| \|y\|$  the Euclidean norm corresponding to the square root of the dot product of each vector with itself.

To determine the optimal number of dimensions for the HAWIK corpus a synonymy test was used, which based on questions from the TOEFL ‘test of english as a foreign language’ compared the LSA cosine similarity of the multiple choice test synonyms, while varying the number of eigenvectors until an optimal percentage of correct answers were returned [18]. For the HAWIK matrix, we found a best fit of 71.2% correctly identified synonyms for 125 dimensions, which is above the 64.5% TOEFL average test score achieved by non-native speaking US college applicants, in line with previous results obtained using either LSA or probabilistic topic models [63]. Subsequently, using the LSA derived cosine similarities of word vectors as a distance matrix, we apply multidimensional scaling MDS, which initially distributes all verbs randomly in two dimensions, compares the difference between their current and target distances, repeatedly repositioning every node until a least squares fit is optimized [64].

To quantify the connectivity we model the  $3 \times 20$  action verbs as nodes using a force directed graph algorithm [20] whereby the links are weighted in proportion to their LSA cosine similarity thresholded at values above 0.20. Here the strength of node  $x_i$  is given by its degree and weights of links i.e. the adjacency and weight matrices of nodes  $i$  and  $j$ .

We calculate the eigenvector centrality [37] which weights nodes not only based on their degree of connectivity, but similar to the Google PageRank algorithm also takes into consideration whether the links are formed between nodes that are themselves central within the network. That is, for a  $m \times n$  matrix  $A$  containing pairwise similarity measures the eigenvector centrality  $x_i$  of node  $i$  is defined as the  $i$ -th entry in the normalized eigenvector belonging to the largest eigenvalue  $\lambda$  of  $A$  then

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j$$

so that  $x_i$  is proportional to the sum of similarity scores of all nodes connected to it.

## References

1. Engel AK, Maye A, Kurthen M, König P (2013) Where’s the action ? the pragmatic turn in cognitive science. Trends in Cognitive Sciences 17.
2. Barsalou LW (2008) Grounded cognition. Annual Review of Psychology 59: 617-645.
3. Barsalou LW, Santos A, Simmons WK, Wilson CD (2008) Language and simulation in conceptual processing. Symbols, embodiment and meaning : 245-283.

4. Lakoff G, Johnson M (1999) *Philosophy in the flesh; the embodied mind and its challenge to western thought*. Basic Books.
5. Rizzolati G, Craighero L (2004) The mirror-neuron system. *Annual Review of Neuroscience* 27: 169-192.
6. Moseley RL, Pulvermüller F (2014) Nouns, verbs, objects, actions and abstractions: local fmri activity indexes semantics not lexical categories. *Brain & Language* 132: 28-42.
7. Aziz-Zadeh L, Wilson SM, Rizzolati G, Iacoboni M (2006) Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology* 16: 1818-1823.
8. Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. *Nature Neuroscience* 11: 351-360.
9. Moreno I, de Vega M, Inmaculada L (2013) Understanding action language modulates oscillatory mu and beta rhythms in the same way as observing actions. *Brain and Cognition* 82: 236-242.
10. Shebani Z, Pulvermüller F (2013) Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex* 49: 222-231.
11. Repetto C, Colombo B, Cipresso P, Riva G (2013) The effects of rtms over the primary motor cortex: the link between action and language. *Neuropsychologia* 51: 8-13.
12. Klepp A, Weissler H, Nicolai V, Terhalle A, Geisler H, et al. (2014) Neuromagnetic hand and foot motor sources recruited during action verb processing. *Brain & Language* 128: 41-52.
13. Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G (2012) Coming of age: a review embodiment and the neuroscience of semantics. *Cortex* 48: 788-804.
14. Louwse MM (2010) Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science* 3: 273-302.
15. Monaghan P, Christiansen MH, Chater N (2007) The phonological-distributional coherence hypothesis: cross-linguistic evidence in language acquisition. *Cognitive Psychology* 55: 259-305.
16. Glenberg AM, Gallese V (2011) Action-based language: A theory of language acquisition, comprehension, and production. *Cortex* doi:10.1016/j.cortex.2011.04.01: 1-18.
17. Moseley R, Carota F, Hauk O, Mohr B, Pulvermüller F (2011) A role for the motor system in binding abstract emotional meaning. *Cerebral Cortex* doi:10.1093/cercor/bhr238: 1-14.
18. Landauer TK, Dumais ST (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review* 104: 211-240.
19. Petersen MK (2012) LSA software & HAWIK corpus matrices. Technical University of Denmark <https://dl.dropboxusercontent.com/u/5442905/LSA.zip>.
20. Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. *Software - Practice and Experience* 21: 1129-1164.
21. Warriner AB, Kuperman V, Brysbaert M (2013) Norms of valence, arousal and dominance for 13915 english lemmas. *Behavior Research Methods* : 1-17.
22. Russell JA (1980) A circumplex model of affect. *Journal of Personality and Social Psychology* 39: 1161-1178.

23. CMU (1976) The cmu pronouncing dictionary. Technical report, Carnegie Mellon University.
24. Catford JC (1988) A practical introduction to phonetics. Clarendon Press.
25. Graziano MS, Taylor CS, Moore T, Cooke DF (2002) The cortical control of movement revisited. *Neuron* 36: 349-362.
26. Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74: 1114-1124.
27. Jeannerod M, Arbib MA, Rizzolati G, Sakata H (1995) Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neurosciences* 18.
28. Louwrese MM, Zwaan RA (2009) Language encodes geographical information. *Cognitive Science* 33: 51-73.
29. Louwrese MM, Benesh N (2012) Representing spatial structure through maps and language: Lord of the rings encodes the spatial structure of middle earth. *Cognitive Science* 36: 1556-1569.
30. Kaschak MP, Madden CJ, Therriault DJ, Yaxley RH, Aveyard M, et al. (2005) Perception of motion affects language processing. *Cognition* 94: 79-89.
31. Zwaan RA, Taylor LJ (2006) Seeing, acting, understanding: motor resonance in language comprehension. *Journal of Experimental Psychology* 135.
32. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Neuroscience* 10: 186-198.
33. Kousta ST, Vigliocco G, Vinson DP, Andrews M, Campo ED (2011) The representation of action words: why emotion matters. *Journal of Experimental Psychology* 140: 14-34.
34. Crutch SJ, Williams P, Ridgway GR, Borgenicht L (2012) The role of polarity in antonym and synonym conceptual knowledge: evidence from stroke aphasia and multidimensional ratings of abstract words. *Neuropsychologia* 50: 2636-2644.
35. Aviezer H, Hassin RR, Ryan J, Grady C, Josh S, et al. (2008) Angry, disgusted or afraid ? *Psychological Science* 19: 724-732.
36. Rizzolati G, Strick PL (2013) Principles of neural science, McGraw-Hill Medical, chapter Cognitive functions of the premotor systems. 5th edition.
37. Lohmann G, Margulies DS, Horstmann A, Pleger B, Lepsien J, et al. (2010) Eigenvector centrality mapping for analyzing connectivity patterns in fmri data of the human brain. *PLoS ONE* 5: e10232. doi:10.1371/journal.pone.0010232.
38. Abbott JT, Austerweil JL, Griffiths TL (2012) Human memory search as a random walk in a semantic network. In: *Neural Information Processing Systems NIPS 2012*.
39. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bring order to the web. Technical report, Stanford InfoLab.
40. Messinger DS, Mattson WI, Mahoor MH, Cohn JF (2012) The eyes have it: making positive expressions more positive and negative expressions more negative. *Emotion* DOI: 10.1037/a0026498.
41. Ramachandran V, Hubbard E (2001) Synaesthesia - a window into perception thought and language. *Journal of Consciousness Studies* 8: 3-34.

42. Gentilucci M, Corballis MC (2006) From manual gesture to speech: a gradual transition. *Neuroscience and Biobehavioral Reviews* 30: 949-960.
43. Myers-Schulz B, Pujara M, Wolf RC, Koenigs M (2013) Inherent emotional quality of human speech sounds. *Cognition and emotion* 27: 1105-1113.
44. Walker P, Bremmer JG, Mason U, Spring J, Mattock K, et al. (2010) Preverbal infants sensitivity to synaesthetic cross-modality correspondences. *Psychological Science* 21: 21-25.
45. Maurer D, Pathman T, Mondloch CJ (2006) The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science* 3: 316-322.
46. Ladefoged P (1989) Representing phonetic structure. *Working Papers in Phonetics* 73.
47. Liberman DH Alvin Mand Whalen (2000) On the relation of speech to language. *Trends in Cognitive Sciences* 4.
48. Galantucci B, Fowler CA, Turvey MT (2006) The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review* 13: 361-377.
49. Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495: 327-331.
50. Klink RR (2000) Creating brand names with meaning: the use of sound symbolism. *Marketing letters* 11: 5-20.
51. Changizi MA (2011) *Harnessed - how language and music mimicked nature and transformed ape to man*. BenBella Books.
52. Changizi MA, Zhang Q, Ye H, Shimojo S (2006) The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *The American Naturalist* 167: DOI: 10.1086/502806.
53. Dehaene S (2009) *Reading in the brain: the new science of how we read*. Viking.
54. Pulvermüller F (2013) How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences* 17: 458-470.
55. Kovic V, Plunkett K, Westermann G (2010) The shape of words in the brain. *Cognition* 114: 19-28.
56. Imai M, Kita S, Nagumo M, Okada H (2008) Sound symbolism facilitates early verb learning. *Cognition* 109: 54-65.
57. Nygaard LC, Cook AE, Namy LL (2009) Sound to meaning correspondences facilitate word learning. *Cognition* 112: 181-186.
58. Zwaan RA (2014) Embodiment and language comprehension: reframing the discussion. *Trends in Cognitive Sciences* 18.
59. Wittgenstein L (1953) *Philosophical investigations*. Wiley-Blackwell.
60. Stopczynski A, Stahlhut C, Petersen MK, Larsen JE, Falk Jensen C, et al. (2014) Smartphones as pocketable labs: visions for mobile brain imaging and neurofeedback. *International Journal of Psychophysiology* 91: 54-66.



61. Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman RA (1990) Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41: 39-407.
62. Furnas GW, Deerwester S, Dumais ST, Landauer TK, Harshman RA, et al. (1988) Information retrieval using a singular value decomposition model of latent semantic structure. In: 11th Annual International ACM SIGIR Conference. pp. 465-480.
63. Griffiths TL, Steyvers M, Tenenbaum JB (2007) Topics in semantic representation. *Psychological Review* 114: 211-244.
64. Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1-27.