

Technical University of Denmark



Cognitive semantic networks: emotional verbs throw a tantrum but don't bite

Petersen, Michael Kai; Hansen, Lars Kai

Published in:

2012 3rd International Workshop on Cognitive Information Processing (CIP)

Link to article, DOI:

[10.1109/CIP.2012.6232911](https://doi.org/10.1109/CIP.2012.6232911)

Publication date:

2012

[Link back to DTU Orbit](#)

Citation (APA):

Petersen, M. K., & Hansen, L. K. (2012). Cognitive semantic networks: emotional verbs throw a tantrum but don't bite. In 2012 3rd International Workshop on Cognitive Information Processing (CIP) IEEE. DOI: 10.1109/CIP.2012.6232911

DTU Library

Technical Information Center of Denmark

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.

Cognitive semantic networks: emotional verbs throw a tantrum but don't bite

Michael Kai Petersen
Cognitive Systems, DTU Informatics
Technical University of Denmark
Building 321, DK-2800 Kgs.Lyngby, Denmark
Email: mkp@imm.dtu.dk

Lars Kai Hansen
Cognitive Systems, DTU Informatics
Technical University of Denmark
Building 321, DK-2800 Kgs.Lyngby, Denmark
Email: lkh@imm.dtu.dk

Abstract—Neuroimaging studies have over the past decades established that language is grounded in sensorimotor areas of the brain. The same neuronal circuits seem involved whether we literally pick up a ball or in a phrase refer to grasping an idea. However recent findings have demonstrated that not only leg, hand and face related but also emotional action verbs activate premotor systems in the brain. Hypothesizing that the force and spatial parameters which define action based language might also be reflected in the latent semantics of words, we select motor and emotion related verbs and apply latent semantic analysis, multidimensional scaling, hierarchical clustering and network graph analysis to quantify their interaction and identify parameters of force and spatial differentiation which we propose cognitively relate emotions to sensorimotor action schemas.

I. INTRODUCTION

Whether we move our hands, only imagine doing so, or recognize a similar gesture when looking at other people, mirror neurons in the brain are activated in Broca's area, a part of the frontal lobe involved in speech articulation [1]. Mounting evidence for a multimodal mapping of actions to words has provided a foundation for the recently proposed theory of action-based language [2], where the literal or metaphorical meaning of words is mentally simulated by mapping how we physically interact with objects into abstract concepts that relate verbs constrained by parameters of force and motion. In turn reflecting how mirror neurons in the brain translate spatial positions of entities into sequences of motor programs, that enable us to audiovisually locate, reach out and grasp an object [3]. And thereby impose a structure on action concepts based on neural parameters which might e.g. define the levels of duration or force to differentiate an action like 'push' from 'shove' [4]. Lately neuroimaging studies have demonstrated that the sensorimotor system provides a binding not only for word categories related to leg, arm or face motion but also for verbs with an abstract emotional meaning [5]. Showing a large overlap in activation elicited by emotion, face and hand related words, the findings indicate that the neuronal circuits of both action concepts and affective expressions are closely linked to their verbal representations.

It has recently been proposed that neurosemantics might benefit from simulating the correlated activation of word forms in language areas of the brain using combinatorial statistics of word co-occurrences in large scale text corpora [6], whether

based on neural networks [7], independent components analysis ICA [8] or latent semantic analysis LSA [9]. Cognitive component analysis COCA, defined as an unsupervised learning of features resembling how we perceive the world, suggests that sensory structures can be modeled by reducing dimensionality and treating objects in space and time as linear mixtures that enable machine learning ranging from classification of musical genres [10] to phonemes in speech processing [11]. We have earlier applied COCA to song lyrics by combining LSA with tensor decomposition [12], in order to extract parameters defining emotional dimensions.

Going beyond the original COCA approach, we explore in the current paper the connectivity and clustering of 4×10 leg, hand, face and emotion related action verbs by modeling them as nodes within weighted graphs constituting cognitive semantic networks. Hypothesizing that the force and spatial parameters which define action-based language might also be reflected in the co-variances of words underlying their latent semantic structure, we apply LSA to the selected verbs based on two large scale text corpora HAWIK and TASA available online¹². Subsequently we use multidimensional scaling MDS [13] to analyze the parameters linking action verbs, hierarchical clustering and network graph analysis to quantify their connectivity [14], in order to compare the cognitive semantic network structures against current findings in action-based language and neuroimaging. In the following sections we outline methods, results and discuss the parameters which we propose cognitively relate the meaning of emotion and action verbs to sensorimotor schemas.

II. METHODS

Initially selecting 10 leg motion verbs plus a subset of 3×10 hand, face and emotion related action verbs used in a recent neuroimaging study [5], which applied fMRI functional magnetic resonance imaging to measure hemodynamic responses in a passive reading task, we submit this subset of action verbs to an LSA analysis [9], in order to define similarities between each pair of the 4×10 verbs based on the HAWIK and TASA

¹Technical University of Denmark: LSA software & HAWIK corpus matrices <http://dl.dropbox.com/u/5442905/LSA.zip> DTU Informatics, 2010

²University of Colorado: LSA latent semantic analysis based on TASA general reading corpus <http://lsa.colorado.edu> CU Boulder, 1998

text corpora. After removing the most common words using a standard stop words list, the text corpora are indexed to generate matrices consisting of rows of words and columns of documents which in turn makes it possible to model the verbs as linear combinations of the multiple sentences in which they occur. Applying singular value decomposition SVD new matrices of lower dimensionality are generated 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 matrix. Dimensionality is reduced from millions to a few hundred by keeping only the eigenvectors which correspond to the highest eigenvalues in the original matrix. In order to explore to what degree the latent semantics of action verbs generalize we define cosine similarities based on two different large scale collections of texts: the HAWIK corpus generated at The Technical University of Denmark (consisting of 22829 words found in 67380 excerpts of Harvard Classics literature, Wikipedia articles and Reuters news - both LSA software & HAWIK matrices are available for download), and for reference the online LSA made available by the University of Colorado based on the TASA corpus (consisting of 92409 words in 37651 texts taken from the general reading material US students have been exposed to when entering college).

To determine the optimal number of eigenvectors for the HAWIK matrix we adjusted the LSA setup using a synonymy test based on questions from the TOEFL ‘test of english as a foreign language’ i.e. comparing the distances between synonyms while varying the number of dimensions until an optimal percentage of correct TOEFL answers are returned [9]. For the HAWIK corpus we found a best fit of 71,2% correctly identified synonyms for 125 dimensions, above the 64.5% TOEFL average test results achieved by non-native speaking college applicants, and in line with previous results obtained using either LSA or latent dirichlet allocation LDA [15]. Subsequently we project the verbs into this latent semantic space and define the similarity of words as the cosine to the angle between their vector representations. And after setting all negative values to zero generate a 40×40 adjacency matrix defining the pairwise interaction between verbs based on their cosine similarity. As well as performing a hierarchical clustering to analyze how the verbs are merged into groups.

Visualizing the interaction of verbs, we take the Pearson correlation of their mutual cosine similarities and project them into a two-dimensional space using MDS [13]. Repeatedly adjusting the pairwise distances between nodes until a stress function is minimized, action verbs with high similarity values will spatially be positioned close to each other. Next to quantify the connectivity we model the 4×10 action verbs as nodes using a force directed graph algorithm [16], 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 . And to quantify the connectivity we calculate the *betweenness centrality*

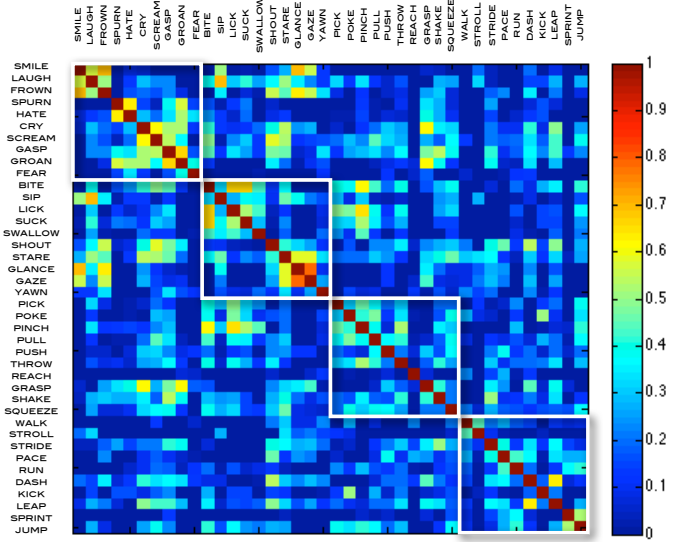


Fig. 1. Adjacency matrix defining the interaction between 4×10 emotion, face, hand and leg related verbs, weighted by their LSA cosine similarity using the HAWIK corpus consisting of 22829 words found in 67380 excerpts of Harvard Classics literature, Wikipedia articles and Reuters news.

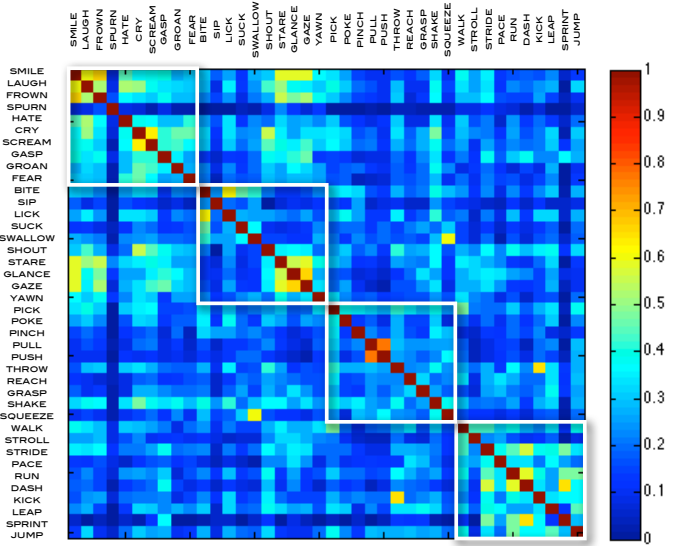


Fig. 2. Adjacency matrix defining the interaction between 4×10 emotion, face, hand and leg related verbs, weighted by their LSA cosine similarity using the TASA corpus consisting of 92409 words found in 37651 texts reflecting the reading material students have been exposed to when entering college.

$$x_i = \sum_{i \neq j, i \neq k, j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

where σ_{jk} are the total number of shortest paths from j to k and $\sigma_{jk}(i)$ the fraction of those which go through node i , assuming that these connections would reflect the most efficient global structure for linking subgraphs within the network.

III. RESULTS

A. Multidimensional scaling

Taking the adjacency matrices defining the LSA cosine similarities of action verbs (Fig. 1 and 2) and applying MDS, the resulting low dimensional representations can be interpreted as reflecting neural parameters similar to those providing the structure in action-based language. Based on the HAWIK corpus (Fig. 3), parameters of distance and size differentiate verbs such as ‘reach’ characterizing the outer bounds when extending the arm (top), from close-up small gestures related to eating like ‘swallow’ (bottom). Likewise on a vertical axis contrasting parameter values of velocity as well as energy separate ‘yawn’ (upper left) from ‘jump’ (lower left). While horizontally different parameter values of applied force and temporal duration juxtapose ‘stroll’ (left) against ‘shake’ (right). Similar linear contrasts can be found in the layout of action verbs based on the TASA corpus (Fig.4), where parameters of distance and size divide small close range gestures like ‘swallow’ (top right) from actions of extended motion like ‘reach’ (bottom right). Horizontally parameters of velocity and energy differentiate drawn out expressions such as ‘yawn’ (left) from powerful sudden actions like ‘jump’ (right). Whereas different levels of applied force and temporal duration characterize aspects of relaxed leisurely motion like ‘stroll’ (lower left) versus controlled repetitive patterns as in ‘shake’ (center).

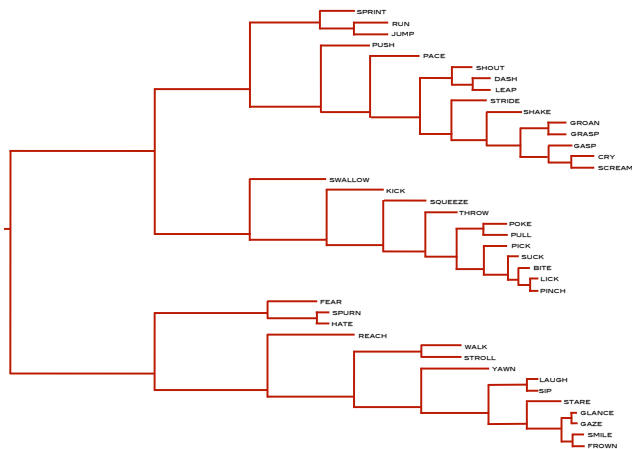


Fig. 7. Hierarchical clustering of action verbs in HAWIK based on their LSA cosine similarity largely divides them into: aroused gestures reflecting high force and velocity parameter values (top), close-up tongue and hand actions coupling low parameter values of distance and size (middle), extended motion, eye and emotional expressions combining low energy with high duration parameter values (bottom).

B. Network structure

Applying a force-directed algorithm to generate weighted graphs of action verbs thresholded at a LSA similarity of 0.20, results in a HAWIK graph with 88 cliques of connected nodes with an average size of 6, among which 3 are fully interconnected subgraphs i.e. maximum cliques of size 10 shaded in grey, while strong links with a LSA cosine similarity above 0.50 are visualized as solid lines (Fig.5). Analyzing the global structure within the HAWIK graph by calculating pairwise distances between all edges to identify the nodes which channel the shortest paths, the highest betweenness centrality values are found for ‘gasp’: 0.066 and ‘stare’: 0.064 ($M = 0.017$). Here the latter node ‘stare’ functions as a bridge connecting strongly linked eye motion and emotional expressions such as ‘glance’ and ‘laugh’, with another subgraph of nodes characterized by forceful aroused gestures such as ‘cry’ and ‘shake’. Overall forming a cluster of strongly linked emotional expressions which are channeled through the central node ‘gasp’, providing the most efficient connection to another part of the graph, consisting of weakly but mutually fully interlinked maximum cliques of hand and mouth gestures such as ‘suck’ and ‘pinch’.

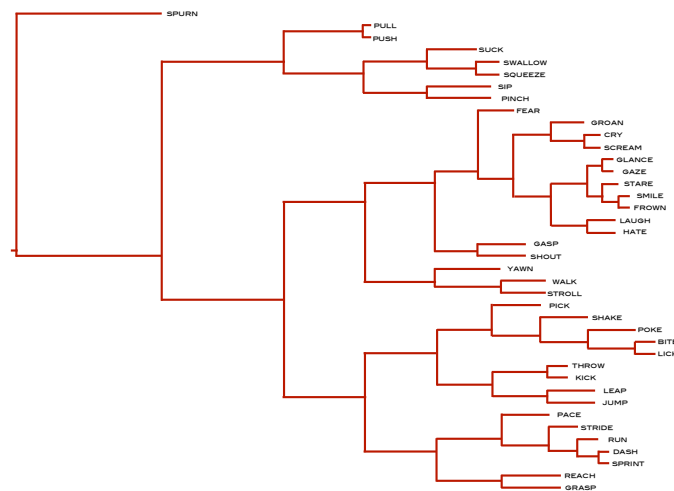


Fig. 8. Hierarchical clustering of action verbs in TASA based on their LSA cosine similarity largely divides them into: close-up tongue and hand actions coupling low parameter values of distance and size (top), extended motion, eye and emotional expressions combining low energy with high duration parameter values (middle), aroused gestures reflecting high force and velocity parameter values (bottom).

The TASA graph thresholded at LSA cosine similarity 0.20 contains 108 cliques of connected nodes with an average size of 10, where 5 are maximum cliques with 15 mutually interconnected nodes shaded in grey, while strong links with an LSA cosine similarity above 0.50 are visualized as solid lines (Fig.6). Analyzing the global network structure in the

TASA graph, the node ‘stare’ again stands out as well as the node ‘shake’ having the highest betweenness centrality values of 0.061 and 0.056 respectively ($M = 0.011$). Also here ‘stare’ fuses the shortest paths of strongly linked nodes characterized by relaxed eye motion and emotional expressions such as ‘gaze’ and ‘smile’, which are connected with other clusters consisting of agitated highly energetic verbs like ‘shout’ and ‘scream’. Whereas the node ‘shake’ connects the five weakly linked but mutually fully interconnected maximum cliques.

C. Hierarchical clustering

Performing a hierarchical clustering analysis on the derived LSA distances to determine structural patterns by iteratively merging the closest pairs into groups largely divides the action verbs into three main branches; forceful aroused gestures, close-up tongue and hand actions, extended motion and emotional expressions (Fig.7 and 8).

IV. DISCUSSION

Neuroimaging experiments using functional magnetic resonance imaging fMRI to measure what areas in the brain are activated during a passive reading task, have established that action verbs somatotopically activate the same parts of the motor cortices involved when actually moving legs, hands or the tongue. Hypothesizing that the neural co-activation of associated word forms in the brain might also be reflected in the combinatorial semantics of words, we selected a subset of action verbs used in a recent fMRI neuroimaging study showing that emotional verbs also activate areas in the brain overlapping with premotor areas in the brain [5]. Modeling the 4×10 emotion, face, hand and leg related action verbs as nodes within LSA weighted graphs to quantify their interaction, a number of shared characteristics within the HAWIK and TASA seem apparent. Despite different patterns of action verb usage in the two large scale text corpora, their betweenness centrality assuming that the shortest paths provide the most efficient patterns of connectivity, are correlated for emotion $r = 0.77$ and facial expressions $r = 0.75$, and to a lesser degree for hand $r = 0.37$ and leg $r = 0.34$ related actions.

In both networks emotional expressions such as ‘smile’ and ‘laugh’ are intertwined with clusters of eye motion formed by nodes like ‘gaze’ and ‘glance’. These parts of the networks stand out separate from subgraphs consisting of weakly linked but fully interconnected maximum cliques of hand and mouth gestures such as ‘suck’ and ‘pinch’. Thereby suggesting modular structures not unlike those identified in the above fMRI neuroimaging study of action verbs, which demonstrated distinct but overlapping patterns characterized by higher activation for emotional verbs in the left prefrontal regions of the brain. In terms of left and right lateralization in the brain, the hand action verbs were found to dominate the left lateral premotor and motor regions, whereas face related verbs additionally triggered the supramarginal gyrus in the right hemisphere. Emotional verbs showed a stronger activation in the orbitofrontal and dorsolateral prefrontal regions compared to the mouth and hand related verbs [5].

Graph theory emphasizes analysis of connectivity patterns based on the topological layout of nodes, which is likewise relevant for analysis of brain networks characterized by a high degree of clustering, where closely positioned regions have a higher likelihood of being connected, while remote modules are typically linked through high centrality nodes functioning as hubs. The inherent modular structure in brain networks can be estimated by hierarchical clustering, which makes the heavily interlinked subgraphs stand out as they are often only sparsely connected with nodes in other modules [14].

Related to the hierarchical clustering and graph analyses of action verbs, the separation of emotional expressions from the hand and mouth related verbs, is in line with neuroanatomical findings of lateralization within the two hemispheres of the brain. In particular the fully interconnected maximum cliques of verbs such as ‘bite’ and ‘pinch’ resemble in functionality the posterior part of Broca’s area, which has been shown to partake in both speech processing and hand movement [2]. A network structure similar to that of the hand and mouth action verbs characterized by clusters of closely connected nodes, would assure a robust resilience against errors if any of the links should fail. While high centrality nodes functioning as bridges channeling the shortest paths would provide the most efficient transfer of information within a larger functional network structure [14].

Also the division between the fully interconnected nodes representing close-up mouth and hand gestures from the subgraph of expanding motion as well as eye and emotional expressions, might reflect differences in attentional processing within the two halves of the brain. Vigilant, alert and sustained attention are primarily right part brain processes related to intensity, whereas selectivity understood as focused and divided attention are shifted towards the left hemisphere. Meaning, the left part provides a narrow and close-up attentional focus complementary to the perception of depth and temporally sustained attention supported by the right part of the brain. When it comes to interpreting emotional facial expressions, clinical studies involving patients with lesions in the left part indicate that this part of the brain is literally only capable of reading the lips and features of the mouth. Whereas the right hemisphere which plays a larger role in interpreting reciprocal affective aspects, is responsible for integrating the more complex expressions conveyed by the eyes as well as storing facial features into memory [17].

These divisions also come out in the hierarchical clustering analysis which by iteratively merging the most similar pairs of action verbs generates three modules; aroused gestures reflecting high force and velocity parameter values, close-up tongue and hand actions coupling low parameter values of distance and size, and eye and emotional expressions combining low energy with high duration parameter values. Here the central node ‘stare’ channelling the largest number of shortest paths is clustered within the module juxtaposing positive and negative feelings. While the other high betweenness centrality nodes ‘gasp’ and ‘shake’ are both grouped within clusters of aroused gestures based on the HAWIK and TASA respectively.

The two clusters thus appear to capture a low dimensional representation of expressions and gestures that reflect the valence i.e. how pleasant something is perceived as being along an axis going from ‘smile’ to ‘fear’, whereas arousal captures the levels of excitability in action verbs like ‘jump’ or ‘leap’ [12]. Although the three clusters which are generated based on the HAWIK corpus (Fig.7) appear less distinct when performing the analysis based on the TASA corpus (Fig.8), 11 out of 14 terms in the clusters representing eye and emotional expressions nevertheless remain shared across the HAWIK and TASA corpora. Similarly 6 of the 11 tongue and hand actions, and 6 of the 15 aroused gestures are shared, when comparing the hierarchical clustering based on the TASA against the HAWIK corpus.

Finding these topological structures across two large-scale text corpora, we propose that the neural co-activation of associated word forms in the brain might also be reflected in the combinatorial semantics of words. And suggest that cognitive component analysis COCA, understood as unsupervised learning of features resembling how we perceive the world, might be extended to simulate how we map actions into concepts using combinatorial statistics of word co-occurrences in large scale text corpora. That is, the latent semantics of action verbs modeled by reducing dimensionality and treating objects in space and time as linear mixtures, might allow us to retrieve the parameters of force and motion which impose the structure on action-based language in the sensorimotor networks of the brain [2].

In that sense one could interpret the MDS derived spatial contrasts of action verbs exemplified by close-up gestures such as ‘lick’ and extended motion like ‘reach’, as being not only defined by the action-based language parameter of distance, but grounded in the mirror neuron system of the brain. Studies of parts of the macaque brain corresponding to human premotor areas show that reachable objects within peri-personal space are transformed into sequences of motor programs. In turn triggering a series of action concepts like if we ‘glance’, ‘reach’ out and ‘grasp’ an apple in order to take a ‘bite’ and ‘swallow’ it. Similarly controlling an opening and closing gesture like ‘pinch’ relies not only on parameter values of granularity, but involves canonical neurons in the brain which transform features of magnitude into temporal sequences for how to manipulate an object based on continuous feedback. Likewise the aspects of motion inherent in action verbs should not only be thought of as abstract parameter values of velocity but could literally reflect the firing rate of neurons. As in the brain of a cat, where the rate at which neurons fire control motor circuits defining the gait, and thereby determine whether the cat struts, trots or gallops [3].

Although a limited number of parameters of force and motion might suffice to constrain action-based language schemas underlying how we physically interact with objects or mentally simulate abstract concepts, modeling these action schemas still remain daunting as they are constantly being modified based on the changing context. On the other hand one could argue that capturing these very 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. We therefore propose that combining latent semantics with action-based language theory may longer term enable us to model not only how actions relate to objects but also how our inner states are linked to perception [18], constrained by parameters of force and direction in a space encompassing the extremes of emotional contrasts.

V. ACKNOWLEDGEMENT

This research is supported in part by The Danish Council for Strategic Research funded project “Responsible Business in the Blogosphere”

REFERENCES

- [1] F. Pulvermüller and L. Fadiga, “Active perception: sensorimotor circuits as a cortical basis for language,” *Nature Neuroscience*, vol. 11, pp. 351–360, May 2010.
- [2] A. M. Glenberg and V. Gallese, “Action-based language: A theory of language acquisition, comprehension, and production,” *Cortex*, vol. doi:10.1016/j.cortex.2011.04.01, pp. 1–18, 2011.
- [3] V. Gallese and G. Lakoff, “The brain’s concepts: the role of the sensory-motor system in conceptual knowledge,” *Cognitive Neuropsychology*, vol. 22, no. 3-4, pp. 455–479, 2005.
- [4] S. Narayanan, “Moving right along: a computational model of metaphoric reasoning about events,” in *AAAI-99*. American Association for Artificial Intelligence, 1999.
- [5] R. Moseley, F. Carota, O. Hauk, B. Mohr, and F. Pulvermüller, “A role for the motor system in binding abstract emotional meaning,” *Cerebral Cortex*, vol. doi:10.1093/cercor/bhr238, pp. 1–14, 2011.
- [6] F. Pulvermüller, “Meaning and the brain: The neurosemantics of referential, interactive, and combinatorial knowledge,” *Journal of Neurolinguistics*, vol. doi:10.1016/j.jneuroling.2011.03.004, pp. 1–37, 2011.
- [7] F. Pulvermüller and A. Knoblauch, “Discrete combinatorial circuits emerging in neural networks: a mechanism for rules of grammar in the human brain?” *Neural Networks*, vol. 22, pp. 161–172, 2009.
- [8] T. Honkela, A. Hyvärinen, and J. J. Väyrynen, “Word ica - emergence of linguistic representations for words by independent component analysis,” *Natural Language Engineering*, vol. 16, no. 3, pp. 277–308, 2010.
- [9] T. K. Landauer and S. T. Dumais, “A solution to plato’s problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge,” *Psychological Review*, vol. 104, no. 2, pp. 211–240, 1997.
- [10] L. K. Hansen, P. Ahrendt, and J. Larsen, “Towards cognitive component analysis,” in *International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning*, 2004.
- [11] L. Feng and L. K. Hansen, “On phonemes as cognitive components of speech,” in *The 1st IAPR Workshop on Cognitive Information Processing*, 2008.
- [12] M. K. Petersen, M. Mørup, and L. K. Hansen, “Latent semantics as cognitive components,” *IEEE Explore*, vol. doi: 10.1109/CIP.2010.5604233, pp. 1–6, 2010.
- [13] J. B. Kruskal, “Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis,” *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.
- [14] E. Bullmore and O. Sporns, “Complex brain networks: graph theoretical analysis of structural and functional systems,” *Nature Neuroscience*, vol. 10, pp. 186–198, 2009.
- [15] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, “Topics in semantic representation,” *Psychological Review*, vol. 114, no. 2, pp. 211–244, 2007.
- [16] T. M. J. Fruchterman and E. M. Reingold, “Graph drawing by force-directed placement,” *Software - Practice and Experience*, vol. 21, no. 11, pp. 1129–1164, 1991.
- [17] I. McGilchrist, *The master and his emissary: the divided brain and the making of the western world*. Yale University Press, 2009.
- [18] L. Wittgenstein, *Philosophical investigations*. Wiley-Blackwell, 1953.