

THE INVESTIGATION OF COMMONALITIES IN HUMAN BRAIN SEMANTIC REPRESENTATIONS ACROSS PEOPLE AND ACROSS LANGUAGES

Augusto Buchweitz

Universidade Federal do Rio Grande do Sul

Abstract

Recent studies on the organization of conceptual knowledge in the human brain have reported the remarkable ability to predict the word a person is thinking about, or the picture a person is seeing, from their brain activity. By using Machine Learning techniques to analyze neuroimaging data, researchers have been able to find stable patterns of brain activity across different people. These patterns allow computer algorithms to identify the brain activity associated with a specific word or picture. The studies have also reported striking commonalities across different people's neural signature for the conceptual knowledge associated with thinking about words. The communal characteristic of the organization of meaning allows for the prediction of what one person is thinking about based on another person's brain activity. The results of some of these studies and the implications for research on cognitive processes and second language learning/acquisition are discussed. Preliminary results from a brain imaging study on cross-language thought identification are also presented. These

Ilha do Desterro	Florianópolis	nº 60	p. 105- 120	jan/jun 2011
------------------	---------------	-------	-------------	--------------

recent findings in neuroimaging of human semantics suggest the presence of a common semantic neural representation across people and across languages.

Keywords: fMRI; bilingualism; machine-learning.

Resumo

Estudos recentes sobre a organização da ativação cerebral associada com diferentes conceitos demonstram a extraordinária capacidade de prever, a partir da atividade cerebral, qual palavra ou imagem uma pessoa estaria lendo ou imaginando. Com o uso de técnicas de Aprendizado de Máquina para analisar dados coletados com neuroimagem, é possível identificar padrões estáveis de ativação cerebral em grupos diversos de pessoas. Estes padrões permitem que algoritmos de computação de dados identifiquem a atividade cerebral associada com uma palavra ou imagem. Estes estudos também demonstram que há características comuns entre pessoas diferentes e a atividade cerebral associada com os conceitos de palavras. Esta característica comunal da organização de significados no cérebro humano permite que os algoritmos sejam utilizados para prever o que uma pessoa está pensando com base na atividade cerebral de outra pessoa. Neste trabalho, os resultados de alguns desses estudos e suas consequências para a pesquisa em cognição e aprendizado/aquisição de uma segunda língua são discutidos. Resultados parciais de um estudo de neuroimagem e identificação de atividade cerebral com duas línguas também serão discutidos. Os estudos sobre a semântica e o cérebro humano sugerem a presença de uma representação neural que é comum entre pessoas e entre diferentes línguas.

Palavras-chave: ressonância magnética funcional; bilinguismo; aprendizagem de máquina

Introduction

When we read a word, we activate its meaning, thereby bringing various associated concepts into an activated state. With the advent of brain imaging techniques, it has become possible to identify the

areas of the brain that are biologically activated when the meaning of a word is (conceptually) activated. One of the issues that studies of brain imaging and word reading try to address is how meaning is organized in the human brain. For second language researchers, it has been a lasting challenge to understand how bilinguals (re)organize conceptual knowledge when a new language is learned.

The process of learning a second language inevitably leads to some level of merging and association of meaning across languages. In the light of some of the most recent cognitive neuroscience studies on bilingualism, the present paper addresses the use of artificial intelligence techniques (Machine Learning) and neuroimaging data as a promising combination for the identification and prediction of the neural underpinnings of the representation of word meaning in the brain. These studies have been construed as the first steps towards “mind reading,” “brain reading” (Cox and Savoy, 2003), or as the Carnegie Mellon researchers like to call them, “thought identification” experiments (see interview with Marcel Just and Tom Mitchell in Frank, 2009). Regardless of the label, the combination of Machine Learning and brain imaging data is beginning to shed light into the black box of human neural representations for the meaning of objects that surround us and how we perceive them, and into whether these representations have commonalities across people. The objective of the present paper is to briefly introduce machine learning and its use in cognitive neuroscience, and to address some of the implications of these new findings for research on cognition and second language learning/acquisition. Partial results from a new study on cross-language thought identification are presented and discussed (Buchweitz et al., 2009).

Brain Imaging and Machine Learning

In the past few years, the field of cognitive neuroscience has seen an increase in the use of Machine Learning methods for the analysis of brain activation data (Pereira, Mitchell, Botvinick, 2009). The use of Machine Learning in cognitive neuroscience involves the application of computer algorithms to decode (or artificially “learn” from) the brain activation patterns found in massive imaging data sets. The computer algorithm then uses patterns found in the data set to generate predictions. These patterns help scientists create directly testable models on different types of data (for more in-depth information on Machine Learning, see Mitchell, 1997).

Through the use of machine learning, brain imaging data can be investigated as a pattern recognition problem. A pattern of brain activity is initially discerned from a set of data (called the training set of data, on which the computer algorithm is literally trained) and then the pattern learned from the data can be used to identify that same cognitive state in a new set of data (called the test set, on which the predictions are tested). Brain activity from reading instantiations of a word can be used to identify brain activity for that same word read at a different time, or by a different person.

The machine learning approach to brain imaging data has allowed for the identification of cognitive states associated with thinking about semantic exemplars from different categories, such as tools and dwellings (Haxby et al., 2001; Carlson, Schrater, & He, 2003; Cox & Savoy, 2003; Hanson, Matsuka, & Haxby, 2004; O’Toole et al., 2005; Shinkareva et al., 2008). Cox and Savoy (2003) successfully identified 10 categories of object exemplars presented as images to the participants. In a recent study, Shinkareva et al. (2008) were able to identify a pattern of brain activation for 10 different tools or dwellings participants were thinking about. The authors were able

to test their predictions on the data across different participants, demonstrating a remarkable power of thought identification based on a communal neural signature for concrete nouns. Figure 1 shows a schematic representation of the combination of neuroimaging data and machine learning analysis techniques.

Mitchell et al. (2008) made one of the most impressive contributions from the combination of machine learning techniques and neuroimaging. The researchers were able to predict brain activation for words not shown in the scanner based on word co-occurrences. Extrapolating from the brain imaging data collected, the researchers used machine learning to predict the brain activation for words the participants had *not* seen in the scanner. The predictions were generated based on the semantic features of words as drawn from an online trillion-token text corpus (Google).

In studies of image identification, Kay et al. (2008) were able to identify which complex image a participant was viewing in the scanner by simply using their brain activation. The authors also showed the first steps towards image reconstruction from brain activity; that is, recreating a photograph from one's brain activation associated with it. The authors hypothesized that in the near future it may be possible to reconstruct mental phenomena, such as a picture of a person's visual experience or a dream, from their brain activity alone. The studies that combine brain imaging and machine-learning classification provide new means for the generation of computational models and predictive theories of the organization of conceptual knowledge and psychological phenomena in the human brain.

Distributed organization of brain activation for the meaning of words

Further contribution to the understanding of the organization of conceptual knowledge in the brain comes from the location of the areas that reliably predict the brain activation for words. The locations in the brain that predicted the words that were being viewed suggest that the neural activity common to the participants encodes information about the properties of the word (specific to the categories of the words, that is, tools and dwellings). Shinkareva et al. (2008) generated predictions based on activation associated with motor activity, such as handling a hammer. This indicates that commonalities in brain activation were not simply based on activation associated with the fact that participants were shown the same visual stimulus. The commonality in activation was not due to lower-level cognitive processing of a string of letters, which was the same for all participants (e.g., the word form, or orthographic appearance, for “hammer” or the picture of a hammer).

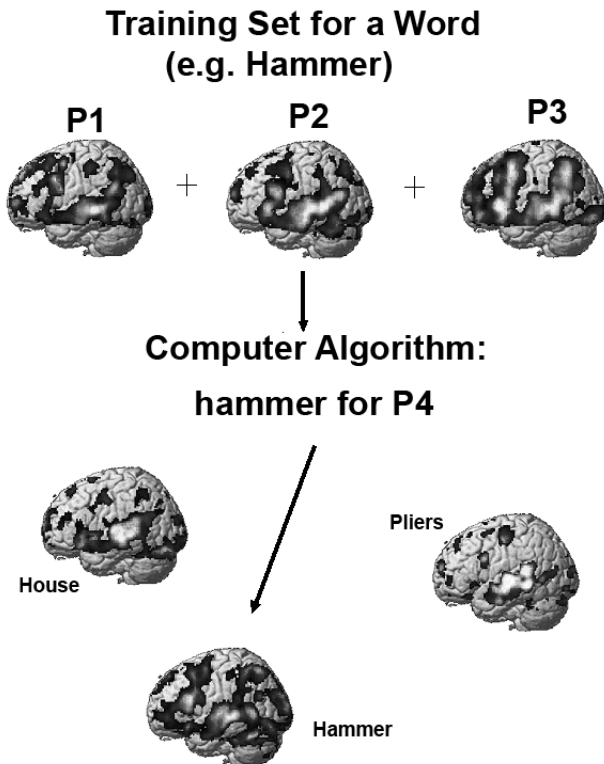
The areas of activation associated with the classification of brain activation for the various individual words indicate that meaning is represented in several locations in the brain. The diffuse representation suggests that word meanings are represented in terms of a large number of features or properties that encode varied types of information. The brain activation for thinking about a word is not just in one place in the brain, it is in a network of places.

The location of the areas of brain activation that predict the semantic properties people are thinking about provides evidence for a perceptual theory of knowledge called “embodied cognition.” Embodied cognition is a theoretical position that holds that meaning representations contain perceptual and motor components corresponding to one’s interaction with objects in the physical

environment (e.g., Barsalou, 1999). Embodied cognition advocates that human perception and cognition are intrinsically bound. They are not two separate modular systems. Cognition is grounded in simulations that combine physical and introspective events (Barsalou, 1999). Meaning and conceptual knowledge are represented in the interaction between the environment and our bodies.

Machine learning studies of thought identification show that despite idiosyncrasies in individual representations of meaning (idiosyncrasies that may be due to different individual experiences) there is an identifiable communal representation of meaning for concrete words in the brain. This representation is distributed and forms networks of activation that may be encoding representations of what it feels like to interact with an object (to hold it, to smell it, to imagine yourself moving about in a dwelling). The alert researcher of second language acquisition or teacher of a second language at this point may be associating the theory of embodied representation of meaning with the facilitation that interaction with actual objects can provide in second language vocabulary acquisition, for example.

Figure 1. Schematic representation of the combination of neuroimaging data and machine learning analysis techniques. Brain activation data for thinking about the properties of a word for a group of participants is averaged and computed to find, among the brain activation for several words, what is the activation data for that same word in another participant.



The Organization of Meaning in the Bilingual Brain

One of the most investigated issues in the field of second language acquisition is the effect of age of acquisition (AoA) on the ultimate attainment of a second language and on brain activation. The variable Age of Acquisition (AoA) divides speakers of two languages into early and late bilinguals. The early bilingual (EBL) is one who acquires two languages, at the same time, from infancy. The late bilingual (LBL), in turn, is one who acquires or learns a second language after the age of seven years (Paradis, 2003). This dichotomy in the AoA of a second

language is warranted by the different cognitive processes (drawing on what is known about procedural and declarative knowledge) that take place during the learning and acquisition processes of LBLs and EBLs.

A question that arises from early or late acquisition of language is whether older age of acquisition hinders the ultimate attainment of second language skill. It is a general consensus in the field of second language acquisition that early (i.e., during childhood), rather than late (i.e., mid teens and older) second language acquisition favors achievement of better L2 proficiency (Hernandez & Li, 2007). In terms of the organization of brain activation for processing a second language, studies have shown that early acquisition of a new language is associated with a striking similarity in the patterns of activation between two languages (Kim et al., 1997) (for more detailed reviews of brain imaging studies of bilingualism, see Abutalebi, Cappa, & Perani, 2001; Buchweitz, 2005; Perani & Abutalebi, 2005; Hernandez & Li, 2007).

The language skills that are more deeply affected by later learning of an L2 are the ability to parse sentences for grammaticality (syntax), and the ability to segment words into its different written parts or sounds (morphology and phonology) (Hernandez & Li, 2007). Roughly speaking, these are skills associated with learning the workings and taxonomy of how to combine man-made, arbitrary symbols for language into writing systems. But not all L2 skills are completely dependent on AoA. Behavioral studies show that proficient bilinguals achieve comparable performance in lexical and semantic processing tasks regardless of their age (Hernandez & Kohnert, 1999; Hernandez & Reyes, 2002; Kohnert, Bates, & Hernandez 1999). Bilinguals seem to be able to become highly skilled in the processing of meaning in a second language early and late in life. Learning a new writing system and orthography, and being able to verbalize and apply their rules is intrinsically less natural than

making meaning associations, which is something we as humans naturally do everyday.

The general notion is that as bilinguals become more proficient there is a developmental shift from lexical to conceptual mapping of meaning in the second language (Kroll & de Groot, 1997). As bilinguals become more proficient, they evolve from processing words in the second language (L2) in association with words in first language (L1), to processing words in the L2 independently of the word in L1 (Chen & Ho, 1986; Tzelgov, Henik, & Leiser, 1990). In a study of semantic priming, Perea et al. (2008) found significant between-language semantic priming for proficient bilinguals, regardless of age of acquisition. The authors replicated the results in two language directions. They concluded that semantic overlap between two languages is a process that solidifies with increased proficiency in the second language.

Picture naming studies have shown that bilinguals learn to map individual words to their concepts in the second language independently of a word association with the first language. Participants in these studies were able to name pictures at the same speed in L1 and L2 (Potter & Faulconer, 1975; Kroll & de Groot, 1997). Also, others have shown that proficient bilinguals can categorize words into semantic groups (e.g., a hammer is a tool) with the same speed in different languages (Caramazza & Brones, 1980; Potter et al., 1984; Shanon, 1982). Studies with bilingual Stroop tasks have shown that more proficient bilinguals suffer a Stroop effect from naming an ink color in their native language when the word names a different color in the second language, in language pairs with similar and with dissimilar orthographies (Kroll & de Groot, 1997; Chen & Ho, 1986; Tzelgov et al., 1990).

Brain imaging studies with bilingual participants corroborate the findings of behavioral studies. As bilinguals become proficient

(more skilled) in a second language, the cortical organization of first and second language cognitive processes also becomes increasingly similar in lexical, semantic, and sentence processing tasks (Perani et al., 1998; Illes et al., 1999; Wartenburger et al., 2003; Green, Crinion, & Price, 2006; Hernandez & Li, 2007). The areas that process language in L1 tend to continue to do so in L2. In a study of bilingual sentence comprehension, Buchweitz (2006) found considerable similarities in the brain activation for processing sentences in English and Portuguese. The collective findings from brain imaging of bilingualism suggest that the activation of the proficient bilingual brain in word and sentence processing in two languages occurs in similar ways. One question that remains is how similar is the actual organization of meaning in the bilingual brain for a concept elicited by words (translation equivalents) in two languages, and are there enough similarities to allow for cross-language prediction.

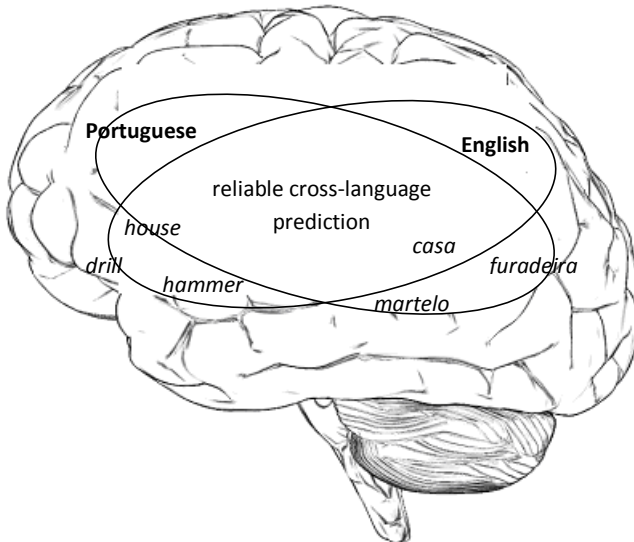
Cross-language identification of brain activation

Buchweitz et al. (2009) carried out a study with the aim to establish a correspondence between the activation patterns associated with thinking about words in two different languages. The study applied classification of brain activity for reading words in English and Portuguese in bilinguals. The initial hypothesis of the study was that there would be common neural representations across languages (the participants were proficient, but late, bilinguals) that would allow for cross-language classification of brain activity. Some of the 14 words for tools and dwellings (and their translation) presented were: hammer (martelo); saw (serrote); house (casa); and hut (cabana).

The results showed a remarkable ability to predict the brain activity associated with thinking about the properties of tools and dwellings across languages (Figure 2). The brain activation for the

properties of words that participants were thinking about in one language reliably predicted the brain activation for words in another language. The ability to identify an individual word and word categories in one language based on data from another language suggests the presence of shared semantic content for concrete nouns across languages in proficient bilinguals. In the highly-proficient bilinguals included in the study, word meaning representations in the second language and in the first language appear to merge significantly to the point that they are identifiable across languages. The merging of concepts did not affect identification of thought in the first language, which was also successful.

Figure 2: The study of shared semantic representation allowed for cross-language prediction of the brain activity associated with words in English and Portuguese based on brain activation for concrete nouns (Buchweitz et al., 2009).



Concluding remarks

The combination of machine learning analysis techniques and brain imaging data provides a glimpse into the future of cognitive neuroscience studies of bilingual brain activation. Recent findings of neuroimaging studies suggest a common semantic neural representation across people and across languages. The organization of the representation of meaning in the brain is not only identifiable, but it has shared characteristics across people and across different languages. The features shared across people for tools and dwellings are associated with their application. Our brains have a similar signature for, for example, the manipulability of objects, and the brain activation for these features remains identifiable in a second language.

Further studies may be carried out to answer whether reliable classification accuracies across languages would be found in less proficient bilinguals, who may have to rely on L2-to-L1 word associations to retrieve the properties of a word. It also remains to be seen whether the classification accuracies for early bilinguals would be different than those obtained for late bilinguals. Optimistically, the use of machine learning techniques in studies of second language acquisition may allow for future development of generative models of the brain representation for different stages of language skill, and for what may be a dynamic overlap of meaning representations.

References

- Abutalebi, J., Cappa, S.F., Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: language and cognition* 4 (2), 179-190.
- Barsalou, L.W. (1999) Perceptual symbol systems. *Behavioral and brain sciences* 22, 577-660.

- Buchweitz, A. (2005). Brain and Language: an overview of neuroimaging studies of bilingual language processing. *Revista brasileira de lingüística aplicada* 5, 87-99.
- Buchweitz, A. (2006). Two languages, two input modalities, one brain: an fMRI study of Portuguese-English bilinguals and Portuguese listening and reading comprehension effects on brain activation. Unpublished doctoral dissertation. *Universidade Federal de Santa Catarina*, Santa Catarina.
- Buchweitz, A., Shinkareva, S., Just, M.A. (2009) Commonality of noun neural representations across languages. *15th Annual Meeting of the Organization for Human Brain Mapping*.
- Carlson, T.A., Schrater, P., He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience* 15, 704-717.
- Caramazza, A., Brones, I. (1980). Semantic classification by bilinguals. *Canadian Journal of Psychology* 34 (1), 77-81.
- Chen, H.C., Ho, C. (1986). Development of stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12, 397-401.
- Cox D.D., Savoy R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19, 261-270.
- Frank, M. (2009). Reading my mind. *60 minutes*. Retrieved January 30th, 2009 from <http://www.cbsnews.com/stories/2009/01/06/60minutes/main4702505.shtml>
- Green, D.W., Crinion, J., Price, C.J. (2006). Convergence, degeneracy, and control. *Language learning* 56 (s1), 99-125.
- Hanson S.J., Matsuka T., Haxby J.V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *NeuroImage* 23, 156-166.
- Haxby J.V., Gobbini M.I., Furey M.L., Ishai A., Schouten J.L., et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425-2430.

- Hernandez A.E., Kohnert K. (1999). Aging and language switching in bilinguals. *Aging Neuropsychology and Cognition*. 6, 69-83
- Hernandez, A.E., Li, P. (2007). Age of Acquisition: Its neural and computational mechanisms. *Psychological bulletin* 133 (4), 638-650.
- Hernandez, A.E., Reyes, I. (2002). Within- and Between-Language Priming Differ: Evidence from Repetition of Pictures in Spanish-English Bilinguals," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28, 726-34.
- Illes, J., Francis, W. S., Desmond, J.E., Gabrieli, J.D.E., Glover, G.H., Poldrack, R., Lee, C.J., Wagner, A.D. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language* 70, 347-363.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L. (2008). Identifying natural images from human brain activity. *Nature* 452, 352-355.
- Kim, K.H.S., Relkin, N.R., Lee, K., Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature* 388, 171-174.
- Kohnert, K.J., Bates, E., Hernandez, A.E. (1999). Balancing bilinguals: Lexical-semantic production and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, and Hearing Research* 42, 1400-1413.
- Kroll, J.F., de Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual. In A.M.B. de Groot and J.F. Kroll (eds) (1997). *Tutorials in bilingualism: Psycholinguistic perspective* (pp. 203-234) Mahwah, NJ: Erlbaum.
- Mitchell, T.M. (1997). *Machine Learning*. McGraw Hill.
- Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science* 320, 1191-1195.
- O'Toole A., Jiang F., Abdi H., Haxby J.V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience* 17, 580-590.
- Paradis, M. (2003). Differential use of cerebral mechanisms in bilinguals. In M.T. Banich, M. Mack (eds.) (2003). *Mind, Brain, and Language: Multidisciplinary perspectives* (pp. 351-370). London: Lawrence Erlbaum.

- Perani, D., Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology* 15, 202-206.
- Perani, D., Paulesu, E., Galles, N.S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S.F., Fazio, F., Mehler, J. (1998). The bilingual brain: proficiency and age of acquisition of the second language. *Brain*, 121, 1841-1852.
- Perea, M., Duñabeitia, J.A., Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language*, 58, 916-930.
- Pereira, F., Mitchell, T., Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage*, 45, S199-S209.
- Potter, M.C., Faulconer, B.A. (1975). Time to understand pictures and words. *Nature* 253, 437-438.
- Potter, M.C., So, K-F, Von Eckhardt, B., Feldman, L.B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior* 23, 23-38.
- Shanon, B. (1982). Identification and classification of words and drawings into two languages. *The quarterly journal of experimental psychology* 34a, 135-152.
- Shinkareva S.V., Mason R.A., Malave V.L., Wang W., Mitchell T.M., Just M.A. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE* 3(1):e1394.
- Tzelgov, J., Henik, A., Leiser, D. (1990). Controlling Stroop interference: evidence from a bilingual task. *Journal of Experimental Psychology* 16, 760-771.
- Wartenburger, I., Heekeren, H.R., Abutalebi, J., Cappa, S.F., Villringer, A., Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron* 37, 159-170.