NeuroImage 106 (2015) 72–85

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Beyond the word and image: characteristics of a common meaning system for language and vision revealed by functional and structural imaging

A.L. Jouen [a], T.M. Ellmore [b], C.J. Madden [a], C. Pallier [c], P.F. Dominey [a], J. Ventre-Dominey [a,*]

[a] INSERM Stem Cell and Brain Research Institute U846, Bron, France
[b] Department of Psychology and Program in Cognitive Neuroscience, The City College and Graduate Center of the City, University of New York, New York, NY, USA
[c] Cognitive Neuroimaging Unit, INSERM-CEA, Gif-sur-Yvette, France

ARTICLE INFO

Article history:
Accepted 11 November 2014
Available online 15 November 2014

Keywords:
Semantic
Language
Vision
DTI
fMRI
Human

ABSTRACT

This research tests the hypothesis that comprehension of human events will engage an extended semantic representation system, independent of the input modality (sentence vs. picture). To investigate this, we examined brain activation and connectivity in 19 subjects who read sentences and viewed pictures depicting everyday events, in a combined fMRI and DTI study. Conjunction of activity in understanding sentences and pictures revealed a common fronto-temporo-parietal network that included the middle and inferior frontal gyri, the parahippocampal-retrosplenial complex, the anterior and middle temporal gyri, the inferior parietal lobe in particular the temporo-parietal cortex. DTI tractography seeded from this temporo-parietal cortex hub revealed a multi-component network reaching into the temporal pole, the ventral frontal pole and premotor cortex. A significant correlation was found between the relative pathway density issued from the temporo-parietal cortex and the imageability of sentences for individual subjects, suggesting a potential functional link between comprehension and the temporo-parietal connectivity strength. These data help to define a "meaning" network that includes components of recently characterized systems for semantic memory, embodied simulation, and visuo-spatial scene representation. The network substantially overlaps with the "default mode" network implicated as part of a core network of semantic representation, along with brain systems related to the formation of mental models, and reasoning. These data are consistent with a model of real-world situational understanding that is highly embodied. Crucially, the neural basis of this embodied understanding is not limited to sensorimotor systems, but extends to the highest levels of cognition, including autobiographical memory, scene analysis, mental model formation, reasoning and theory of mind.

Introduction

A significant portion of human mental life is built upon the construction of perceptually and socially rich internal scene representations, or mental models (Johnson-Laird, 2010). These models can be used for reasoning, exploring specific memories of the past, planning for the future, or understanding current situations that may be presented in real-time through immersion in an actual life-activity scene, or more passively, through reading, watching a movie or viewing a photograph. It is possible that these mental models play an intricate role in generating predictions as part of perception (Friston, 2005), and that they are at the core of meaning. In this context, embodied theories of meaning argue that the human conceptual system is implemented in distributed brain networks whose mechanisms are shared by perceptual and action processing (Barsalou, 1999; Barsalou et al., 2003). In a review of the neurophysiology of meaning, and the continuum from symbolic to embodied models, Binder and Desai (2011) argue that the semantic system consists of both modality specific and supra-modal representations. They advocate a semantic system which consists of multiple levels of abstraction, grounded on sensori-motor and affective representations. At the more abstract level, they highlight the important role of the angular gyrus, in the inferior parietal cortex near the temporo-parietal junction, as an abstract supramodal convergence zone (Damasio, 1989) participating in the semantics of event concepts.

Such representations of meaning should be accessible by different perceptual routes, including language and vision. Indeed, one of the most basic problems of cognition is "understanding" or generating a coherent mental model from an image or sentence. This problem was at the core of cognitive science in the 1980s (Jackendoff, 1987; Johnson-Laird, 1981, 1987, 1988; Kintsch, 1988), with the challenge to find the unified representation of meaning common to language and vision. Behavioral research suggests indeed that language and vision

* Corresponding author at: INSERM Stem Cell and Brain Research Institute, U846, 18 Avenue Doyen Lepine, 69500-Bron, France.
E-mail address: Jocelyne.ventre-dominey@inserm.fr (J. Ventre-Dominey).

http://dx.doi.org/10.1016/j.neuroimage.2014.11.024
1053-8119/

interact thus in a common representation, as cross modal priming can occur between pictures and words (Bajo, 1988).

Examining the possible neurophysiological bases of this relation between language and vision, Vandenberghe et al. (1996) observed significant overlap in brain activation when subjects read single words and saw pictures of simple objects in semantic matching tasks. The common activation was seen in areas including the temporal parietal junction, left middle temporal gyrus and inferior frontal gyrus. Recent studies have examined the commonality between cognitive processing of words and pictures in variety of tasks including priming (Kircher et al., 2009), and brief masked presentation at the threshold of consciousness (Van Doren et al., 2010). Because of their relatively restricted nature of word and object stimuli, such tasks may not tap into the more elaborated processes of comprehension that would be elicited by richer stimuli. For example, in the linguistic domain, as the structural complexity of the input stimulus increases, so does the extent of activation. Xu et al. (2005) thus demonstrated that as stimulus complexity increases over words, sentences and narratives, there is an expansion of activity from perisylvian language areas, to frontal operculum and temporal pole for words and sentences respectively. Processing in a narrative sense extended this recruitment to areas including precuneus, medial prefrontal and dorsal temporo-parietal-occipital cortices. One could imagine a similar expansion in the representation of meaning as image stimuli increase in complexity from simple objects, to richer spatial scenes (Harel et al., 2013).

From Vandenberghe's early work on words and images (Vandenberghe et al., 1996) to recent studies and meta-analyses of the semantic system (Binder and Desai, 2011; Binder et al., 2009; Graves et al., 2010; Vandenberghe et al., 2013) the cortex in the temporal-parietal junction including the angular gyrus appears to play the role of an integrating convergence zone in a distributed semantic network.

If such a network exists, allowing these brain regions to encode meaning representations that may be extended in space and time, it is likely that dedicated anatomical white matter pathways contribute to the infrastructure for this distributed network. Diffusion tensor imaging (DTI) provides a method for analysis of white matter pathways in the living brain based on diffusion properties of water molecules within these pathways (Makris et al., 1997). DTI has proven to be a potent tool for examination of language-related white matter pathways (Anwander et al., 2007; Frey et al., 2008; Friederici, 2009; Friederici et al., 2006; Glasser and Rilling, 2008; Makris et al., 2005; Makris and Pandya, 2009; Menjot de Champfleur et al., 2013; Sarubbo et al., 2013; Saur et al., 2008, 2010; Turken and Dronkers, 2011). Catani and colleagues (Catani et al., 2005) used DTI to examine the organization of one of the principal language-related pathways, the arcuate fasciculus (AF) of the left hemisphere. They identified the classical pathway connecting Broca's region (in the territory of the inferior frontal gyrus) and Wernicke's region (in the territory of middle temporal gyrus), and an additional previously undescribed indirect pathway passing through inferior parietal cortex. They suggest that the indirect pathway, which passes via parietal cortex, is involved in semantically based language functions, while the direct pathway is involved in phonological language functions. Their dissection of the indirect pathway highlights the importance of this parietal area, characterized as Geschwind's territory in BA39/40, which Lichtheim (Lichtheim, 1885) referred to as a concept center (Catani et al., 2005). Glasser and Rilling (2008) subsequently demonstrated distinct components of the AF with a superior temporal gyrus (STG) pathway for phonological content and middle temporal gyrus (MTG) pathway for lexical semantic content in accord with Hickok and Poeppel (2004). Further investigating this connectivity, Frey et al. (2008) identified the link between BA45 and STG via extreme capsule system, and BA44 with rostral inferior parietal lobe (IPL) via the third branch of the superior longitudinal fasciculus (SLF3). Additional pathways from this parietal territory that contribute to language include the middle longitudinal fasciculus (MdLF) linking STG and inferior parietal cortex/angular gyrus (Makris et al., 2009; Menjot de Champfleur et al., 2013; Saur et al., 2010; Turken and Dronkers, 2011), and the inferior fronto-occipital fasciculus (IFOF) (Sarubbo et al., 2013; Turken and Dronkers, 2011). Indeed, frontal terminations of IFOF reveal a multifunction pathway with a superficial part that distributes information from posterior cortex (superior parietal, TPJ, occipital extrastriate cortex) to pars triangularis and opercularis in inferior frontal gyrus, and the deep component that projects to orbito-frontal, middle frontal cortex and dorso-lateral prefrontal cortex (DLPFC) (Sarubbo et al., 2013). This TPJ/MTG area in the proximity of the semantic convergence zone of Binder (Binder and Desai, 2011) is at the crossroads of multiple language related pathways including the AF/SLF, the IFOF and the MdLF (Sarubbo et al., 2013; Turken and Dronkers, 2011) that could contribute to its semantic integration function.

The current research tests the hypothesis that comprehension of human events will engage an extended semantic representation system, independent of the input modality. In particular, if the temporo-parietal cortex in the region of the angular gyrus is an integrating convergence zone for event representation as suggested by Binder and Desai (2011), then this area should be activated when subjects understand sentences or images depicting events. Likewise, if this region plays a privileged role in comprehension, then one would expect that structural connectivity to this area would relate to comprehension. In order to assess this we first characterize the fMRI response to the semantic processing of visual scenes and sentences and then, by using DTI tractography we investigate specific pathways within the activated semantic network.

Previous functional imaging studies comparing linguistic and visual input stimuli have tended to use relatively constrained stimuli (Van Doren et al., 2010; Vandenberghe et al., 1996), rather than richer visual scenes and sentences that would potentially tap more heavily into the semantic system. In the current study, we make a compromise, by using relatively complex stimuli. Most importantly, we use stimuli that engage the semantic system by two separate input modalities – scene vision vs. language. We combine functional and anatomical imaging to identify a neural system for the representation of meaning at the sentence and scene level, independent of the input modality.

The resulting event related fMRI experiment allows us to examine the representation of meaning in two processing modes, corresponding to reading sentences, and seeing pictures (see Fig. 1). In both cases, the stimuli (sentences or pictures) depicted rich human events or scenes. In the sentence condition, example sentences include "The man climbs the ladder", and "The little girl rakes the leaves." In the picture condition, the pictures were photographs depicting these same kinds of events. For both the sentences and pictures, a protocol was employed which ensured that subjects processed the sentence or picture stimuli. We examined fMRI results for sentence and picture comprehension, and the conjunction of these modality dependent activations, in order to identify the common network. After the fMRI, while subjects were still in the scanner we obtained DTI data that allowed us to characterize the anatomical connectivity within the network, and establish the link between such measures of connectivity and a behavioral measure of comprehension.

Materials and methods

Subjects

Nineteen healthy, right-handed, native-French speakers with normal or corrected to normal vision participated in the study (11 male; mean age: 23 years; SD: 5.36 years). No participant had a history of neurological, major medical, or psychiatric disorder. The protocol was approved by the regional ethic committee (Comité de Protection des Personnes Ile-de-France VII. Protocole de Recherche Biomedicale #2008-A00241-54/1) and all participants gave their written informed consent before the scanning session.