Brain & Language 134 (2014) 1–10

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

The role of visual representations during the lexical access of spoken words

Gwyneth Lewis *, David Poeppel

Department of Psychology, New York University, 6 Washington Place, 2nd floor, New York, NY 10003, United States

ARTICLE INFO

Article history:
Accepted 22 March 2014
Available online 9 May 2014

Keywords:
MEG
pMTG
Lexical access
Speech recognition
Imageability

ABSTRACT

Do visual representations contribute to spoken word recognition? We examine, using MEG, the effects of sublexical and lexical variables at superior temporal (ST) areas and the posterior middle temporal gyrus (pMTG) compared with that of word imageability at visual cortices. Embodied accounts predict early modulation of visual areas by imageability – concurrently with or prior to modulation of pMTG by lexical variables. Participants responded to speech stimuli varying continuously in imageability during lexical decision with simultaneous MEG recording. We employed the linguistic variables in a new type of correlational time course analysis to assess trial-by-trial activation in occipital, ST, and pMTG regions of interest (ROIs). The linguistic variables modulated the ROIs during different time windows. Critically, visual regions reflected an imageability effect prior to effects of lexicality on pMTG. This surprising effect supports a view on which sensory aspects of a lexical item are not a consequence of lexical activation.

1. Introduction

Speech perception can be intuitively described as a sequential process involving the piecemeal mapping of continuous acoustic signals onto phonetic units of some form. Less straightforward are the transitional processes and representations leading to lexical retrieval (Poeppel, Idsardi, & Wassenhove, 2008). One particularly thorny problem in the context of lexical processing concerns the hypothesized role of perceptual representations, an issue emphasized by embodiment models in lexical semantic access (e.g., Pulvermüller, 1999). How and when do the acoustic signals and/or phonetic units of speech activate visual representations of a word's real world referent? Does the word *strawberry*, for example, automatically activate a mental image of the color red (for example in the sense of feature spreading), and if so, is such activation a requirement of or merely incidental to lexical access? One especially important issue concerns the temporal dynamics of lexical access. To answer such questions, we examine these issues in the context of the most notable models of lexical processing, which differently emphasize the temporal dynamics and access stages of speech recognition.

2. Models of lexical processing

Neurocognitive models of lexical access suggest the participation of distinct neural regions in the activation of, competition between, and selection of basic sound-meaning representations. In the visual domain, studies of lexical access have determined that certain MEG components are sensitive to the orthographic, morphological, and semantic features of words during different time windows. For example, the frequency of a word's adjacent letter strings (bigram frequency) modulates occipital activation at ~100 ms post word onset (Solomyak & Marantz, 2010). Around 50 ms later, the morphological transition probability of words (the probability of the whole word form, given the stem) modulates responses in the fusiform gyrus (Lewis, Solomyak, & Marantz, 2011; Solomyak & Marantz, 2010). At ~300 ms, properties of the whole word form modulate a superior temporal response (Lewis et al., 2011; Pylkkänen & Marantz, 2003; Simon, Lewis, & Marantz, 2012; Solomyak & Marantz, 2009, 2010). How do the stages of spoken word recognition compare with those involved in visual word recognition?

Interaction-competition models of lexical access and comprehension have attempted to describe the mappings between phonetic units and lexical representations in terms of activation and competition between lexical competitors. The original Cohort model, for example, describes lexical retrieval as a strictly bottom-up selective process wherein the incoming speech signal activates all words beginning with the initial phoneme and

---

* Corresponding author.
E-mail address: gwyneth@nyu.edu (G. Lewis).

http://dx.doi.org/10.1016/j.bandl.2014.03.008
0093-934X/© 2014 Elsevier Inc. All rights reserved.

gradually winnows the selection as word information accrues, eventually discarding all competitors until only the target word remains (Marslen-Wilson, 1987). Another interaction-competition model, TRACE, depicts speech comprehension as a cumulative process wherein a new representation is generated for previous and incoming elements over time, with bidirectional feedback between entries at featural, phonemic, and lexical levels (McClellend & Elman, 1986). The Neighborhood Activation Model (Luce & Pisoni, 1998) conceptualizes the cohort of competitors in terms of phonological neighbors, which are words that differ by one phone from the target word. Similar to Cohort, incoming acoustic information activates "word decision units" (the neighborhood). As in TRACE, there is bi-directional feedback between higher-level information (e.g., context and frequency) and the lower level sensory input.

Embodied perspectives on language processing focus less on activation and competition and more on the role of perceptual (or sensorimotor) representations in, for example, (lexical) semantic access. Strong theories of embodiment view semantic knowledge as grounded in perceptual experience rather than in the relationships between words (Bickhard, 2008). Semantic access is thought to require perceptual simulation and directly engage areas of the brain that are active while perceiving the referent in the real world (Gallese & Lakoff, 2005). Weak-embodiment theories view lexical-semantic access as only moderately dependent on the participation of sensory and motor systems. On such models, semantics may be grounded in sensory and motor information but may also be accessed from higher-level representations (Meteyard & Vigliocco, 2008). In opposition to embodied-based accounts, abstract, symbolic theories view semantic knowledge as derived from correspondences between internal symbols and their extensions to objects in the real world (Fodor & Pylyshyn, 1988). New data could shed light on these theories and disambiguate among some of these predictions.

3. Recent empirical findings

Results from fMRI studies of visual perception and mental imagery suggest that the same occipital regions active while perceiving objects are similarly active while mentally 'simulating' visual images of objects (Ganis, Thompson, & Kosslyn, 2004). Evidence that occipital (visual) regions are involved in simulating perceptual visual features during language comprehension also comes from recent fMRI experiments. One study showed, for example, that occipital regions processed shape information of sounds, wherein the stimulus impact sound of an object (such as a ball bouncing) modulated occipital activation when the hearer's instructions were to name the shape (e.g., *round*) rather than the material (e.g., *rubber*) of the object generating the sound (James, Stevenson, Kim, VanDerKlok, & James, 2011).

Results from Pulvermüller and colleagues' EEG, fMRI, and MEG experiments support the hypothesis that (conceptual or lexical) semantic knowledge activation requires (or at least co-occurs with) perceptual simulation. Pulvermüller (2005) reported that action words involving the feet, hands, and face (e.g., *kick, pick, lick*) elicited cortical activation in motor regions associated with performing those actions with the respective body parts, by argument during early recognition stages. Similar results were reported for other novel sensory modalities during fMRI reading experiments, where scent-words such as *cinnamon* activated olfactory cortices (González, Barrós-Loscertales, Pulvermüller, Meseguer, & Sanjuán, 2006) and taste-words such as *salt* activated gustatory cortices (Barrós-Loscertales et al., 2011). While Pulvermüller, Preissl, Lutzenberger, and Birbaumer (1996) argued that imageable nouns and verbs elicited the visual and motor cortices (respectively) in EEG, the results from a later fMRI reading experiment failed to indicate any effect of shape- and color-words such as *square* and *bronze* on activation in the visual cortex (Pulvermüller & Hauk, 2006).

4. Motivation of the current experiment

Based on such findings, we assume that the visual cortex is at least possibly active during spoken word recognition. Whether and when such activation contributes to meaning-based resolution remains controversial. In previous work, we found that the meaning-based resolution of *visual* words can be verified at around 300 ms post-stimulus onset (Simon et al., 2012). This is reflected in the modulation of a superior temporal response (the MEG M350, comparable to the N400/N400m of Helenius et al. (2002)) by the meaning-entropy (semantic ambiguity) of visually presented words. An absence of earlier semantic effects does *not* mean, however, that lexical resolution (selection of the appropriate representation) does not begin much earlier. But can one diagnose lexical resolution and perceptual simulation at earlier stages, and on which brain regions should one focus?

Previous studies of language processing have employed magnetoencephalography (MEG) combined with structural MRIs to examine the various stages of visual word recognition. The rationale is that MEG provides fine-grained temporal resolution (unlike fMRI) and, when enriched by source modeling that is constrained by structural MRI data, provides good spatial resolution (unlike EEG), which allows for examination on a millisecond level of the neural contributors to word recognition. Such work has determined that in *visual* word recognition, occipital brain regions process orthographic features at ~100 ms, inferior temporal regions decompose morphological properties at ~150 ms, and superior temporal regions contribute to lexical access (of the whole word form) at meaning based-resolution by as early as ~300 ms (Lewis et al., 2011; Simon et al., 2012; Solomyak & Marantz, 2009, 2010).

Of particular interest, in the functional anatomic sense, is the posterior middle temporal gyrus (pMTG), which previous work implicates as an indicator of lexical access in spoken word recognition (Hickok & Poeppel, 2007). While traditional accounts of verbal comprehension emphasize the role of Wernicke's (superior temporal) area in speech processing, there is considerable evidence that the MTG plays a central role in lexical processing (see reviews in, e.g., Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004; Hickok & Poeppel, 2007; Lau, Phillips, & Poeppel, 2008). Evidence that the pMTG is a critical node in the language comprehension network comes from lesion studies that find that, compared with patients with lesions in Wernicke's (superior temporal) and Broca's areas, patients with lesions to pMTG demonstrate poor performance in comprehending and naming single words. The pMTG may therefore link conceptual information to lexical representations (Dronkers et al., 2004). Further evidence comes from a study of connectivity profiles of brain areas within the language comprehension network, which determined that the MTG connectivity pattern is extensively integrated with areas of the network previously found to be critical to sentence comprehension (Turken & Dronkers, 2011). Results from neuroimaging show that MTG activation increases as a function of speech intelligibility (Davis & Johnsrude, 2003) and is also modulated by increasing semantic ambiguity (Rodd, Davis, & Johnsrude, 2005). The MTG therefore provides an ideal testing ground for the study of the processes leading to lexical access of spoken words.

Following spectrotemporal analysis of auditory input in the early auditory cortex and phonological analysis in the superior temporal gyrus (STG) and superior temporal sulcus (STS), the circuitry of the pMTG links phonological information to semantic