# eNeuro

*Research Article: New Research | Cognition and Behavior*

## Language in context: MEG evidence for modality general and specific responses to reference resolution

Language in context

Christian Brodbeck[1,3], Laura Gwilliams[1,3] and Liina Pylkkänen[1,2,3]

[1]*Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA*
[2]*Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003, USA*
[3]*NYU Abu Dhabi Institute, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates*

**Correspondence should be addressed to** CB (christianbrodbeck@nyu.edu)

**Alerts:** Sign up at eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

1. Manuscript Title: Language in context: MEG evidence for modality general and specific responses to reference resolution

2. Abbreviated Title: Language in context

3. List all Author Names and Affiliations in order as they would appear in the published article:

    i.   Christian Brodbeck[1,3]
    ii.  Laura Gwilliams[1,3]
    iii. Liina Pylkkänen[1,2,3]
    iv.  [1]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
    v.   [2]Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003, USA
    vi.  [3]NYU Abu Dhabi Institute, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates

4. Author Contributions: CB and LP designed research, CB and LG performed research, CB and LG analyzed data, CB wrote the paper.

5. Correspondence should be addressed to CB (christianbrodbeck@nyu.edu)

6. Number of Figures: 6

7. Number of Tables: 0

8. Number of Multimedia: 0

10. Conflict of Interest: Authors report no conflict of interest

12. Number of words for Abstract: 213

13. Number of words for Significance Statement: 109

14. Number of words for Introduction: 747

15. Number of words for Discussion: 3000

# Language in context: MEG evidence for modality general and specific responses to reference resolution

Christian Brodbeck[1,3]

Laura Gwilliams[1,3]

Liina Pylkkänen[1,2,3]

[1]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA
[2]Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003, USA
[3]NYU Abu Dhabi Institute, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates

## Abstract

Successful language comprehension critically depends on our ability to link linguistic expressions to the entities they refer to. Without reference resolution, newly encountered language cannot be related to previously acquired knowledge. The human experience includes many different types of referents, some visual, some auditory, some very abstract. Does the neural basis of reference resolution depend on the nature of the referents or do our brains utilize a modality general mechanism for linking meanings to referents? Here we report evidence for both. Using MEG, we varied both the modality of referents, which consisted either of visual or auditory objects, and the point at which reference resolution was possible within sentences. Source localized MEG responses revealed brain activity associated with reference resolution that was independent of the modality of the referents, localized to the medial parietal lobe and starting around 415 ms after onset of reference resolving words. A modality-specific response to reference resolution in auditory domains was also found in the vicinity of auditory cortex. Our results suggest that referential language processing cannot be reduced to processing in classical language regions and representations of the referential domain in modality-specific neural systems. Instead, our results suggest that reference resolution engages medial parietal cortex which supports a mechanism for referential processing regardless of the content modality.

## Significance statement

Reference resolution is an elementary mechanism for language comprehension, connecting language meaning to pre-existing knowledge. It is unknown whether reference resolution depends on brain mechanisms specific to the modality of the referents, for example whether they are visual or auditory objects, or whether our brains utilize a modality general mechanism for linking meanings to referents. Here we show using source localized MEG that reference resolution is associated with a response in the medial parietal lobe, independent of referent modality, supporting a modality-general mechanism for reference resolution. An additional response associated with resolving reference to auditory objects in auditory cortex suggests that modality-specific representations of the referents are also involved.

## Introduction

A crucial precondition for understanding a sentence in context is identifying the entities that the sentence is about. This was demonstrated in a classic study, which showed how a text passage that appears incomprehensible when presented in isolation becomes perfectly natural when presented after a picture that provides meaningful referents for the text (Bransford & Johnson, 1972). However, much of the research on the neural basis of language comprehension ignores this referential dimension, studying generic sentences presented without a specific context. Here we report a study in which we deliberately manipulated the relationship between background knowledge and linguistic expressions to uncover the neural basis of successful reference resolution.

Unlike in most laboratory experiments, language comprehension in the real world takes place in a rich context. Language comprehenders must not only decode the literal meaning of a message, but also connect it to mental models representing what the message is about (Graesser, Millis, & Zwaan, 1997). A body of evidence suggests that such models are connected to modality-specific cognitive systems. For example, even when participants are looking at a blank screen while listening to stories, their eye movement patterns reflect spatial configurations described in the language input (Spivey & Geng, 2001; Altmann, 2004; Altmann & Kamide, 2009). Furthermore, EEG data suggests that readers resolving reference to an item on a previously seen visual display access a retinotopic representation (Brodbeck, Gwilliams, & Pylkkänen, 2015).

These observations resonate with theories of embodied meaning, according to which meaning is represented in the same cognitive systems that also process sensory information (Barsalou, 1999; Hauk, Davis, Kherif, & Pulvermüller, 2008). Theories of strong embodiment go as far as proposing that lexical meanings are represented in sensory regions (Pulvermuller, 2013), implying that the task of mapping meanings to referents could be performed by modality-specific systems exclusively (red path in Figure 1). But even if lexical semantics involves amodal regions, it is still possible that situation models, in which specific referents are represented, are constrained to sensory-specific cortices (blue path in Figure 1). In such a model, amodal regions should be sensitive to lexical, but not referential properties of language input.

3

110    Alternatively, reference resolution could involve an amodal or modality-general mecha-
111    nism that mediates between lexical and referential meanings. Such a mechanism could be
112    related to amodal discourse representations (Graesser et al., 1997), but it might also be re-
113    quired by embodied theories of meaning, for example to coordinate referents of different
114    modalities.

115    A number of fMRI and PET studies have investigated linguistic contrasts that involved ref-
116    erential properties. Coherent language, which involves repeated reference to the same enti-
117    ties, consistently activates medial frontal and medial parietal regions in addition to classic
118    perisylvian language areas (Ferstl, Neumann, Bogler, & von Cramon, 2008). More specifical-
119    ly, medial as well as lateral parietal areas were more active in response to sentence pairs
120    introducing a conjoined subject (e.g., *Jeremy and Roger*) compared to a individually intro-
121    duced referents, suggesting that these regions might be involved in creating and tracking
122    discourse referents (Boiteau, Bowers, Nair, & Almor, 2014). Medial and lateral parietal, as
123    well as frontal areas were also activated by sentences containing unresolved referential
124    ambiguities (Nieuwland, Petersson, & Van Berkum, 2007). While these results point to a
125    number of brain regions that could be involved in referential processing, the low temporal
126    resolution of fMRI and PET makes it difficult to link brain activation to specific linguistic
127    processes, which follow each incoming word in a rapidly cascading fashion. Finally, none of
128    these studies explicitly manipulated the modality of the referents, so their results could be
129    related to imagery associated with language comprehension.

130    Reference resolution itself is inherently difficult to separate from other cognitive processes
131    associated with managing referents, such as retrieving referents form memory and locating
132    them in the environment (see Discussion). Here we tried to create an experimental situa-
133    tion which minimizes memory retrieval and scanning of the environment by presenting
134    simple referential contexts immediately preceding linguistic stimuli that contained referen-
135    tial expressions. We used MEG to directly contrast reference resolving words with non-
136    resolving controls. In order to locate a response involved in reference resolution independ-
137    ent of the modality of the referential domain, we constructed parallel conditions with audi-
138    tory and visual referential domains (see Figures Figure 2 and Figure 3). We predicted two
139    possible response patterns: regions that track the set of possible referents should exhibit
140    more activation for the ambiguous condition, in which the expression is compatible with
141    two referents, while regions that become active when reference is resolved should exhibit
142    more activity in the resolving condition.

## Materials and methods

## Participants

145    We collected data from 26 right-handed native English speakers recruited from the com-
146    munity on the New York University, Abu Dhabi campus. Two participants were excluded
147    because of excessive artifacts leading to less than 70% good trials, leaving 24 participants
148    for the final analysis (18 female, mean age 24, range 19-50). New York University Abu Dha-
149    bi is an English speaking university located just outside the city of Abu Dhabi. Our partici-

4

150 pants were thus immersed in an English speaking environment that involves little contact
151 with local languages unless students and staff actively seek it out. Out of the 24 participants
152 in the final analysis, 16 had grown up speaking only English, whereas 8 had grown up bi-
153 lingually. The second native language varied between participants (Hindi, Japanese (2),
154 Malayalam, Mandarin, Spanish, Urdu, Vietnamese). The protocol was approved by the Insti-
155 tutional Review Board of New York University Abu Dhabi, and all participants provided
156 written consent before beginning the experiment.

## Materials

158 Each trial consisted of presentation of a referential domain, followed by a question about
159 the domain, presented word for word (see Figure 2 and Figure 3). Participants' task was to
160 answer questions such as *Was the grunt in the end?* by pressing one of two response but-
161 tons. Target stimuli were referential expressions which were identical in their linguistic
162 surface properties but differed in whether they resolved reference, for example, *the grunt*
163 in a context that contained one or two grunts.

## Auditory referential domains

165 For auditory referential domains, we selected 10 sounds with monosyllabic names that
166 were easily identifiable and for which short (~100-300 ms) and long (~500-600 ms) vari-
167 ants were clearly distinguishable. Nine sounds (bark, buzz, caw, chirp, cluck, croak, honk,
168 mew, splash) were extracted from sounds available under a creative commons license
169 (http://freesound.org) and one (beep) was created from a 1000 Hz sine wave.

170 Figure 2 shows a sample trial for the auditory referential domains. Each referential domain
171 consisted of 3 sounds, played sequentially. All domains were constructed such that 2
172 sounds shared the name, and 2 other sounds had the same length. Presentation of the do-
173 main started with a fixation cross for 600 ms, followed by the 3 sounds. For long sounds,
174 the stimulus onset asynchrony was 1000 ms, for short sounds it was 700 ms. The domain
175 was followed by the question in serial visual presentation, with function words presented
176 for 400 ms and content words presented for 600 ms. The last word of the question was
177 ended with a question mark and stayed on the screen until the participant made a yes/no
178 response via button press.

179 The target for analysis was the first noun in the question, marked with an arrow in Figure
180 2. In trials of the reference resolving condition, the noun named a unique sound, i.e., the
181 sound that did not share its name with another sound in the domain. In control trials, the
182 noun did not resolve reference because it was compatible with two sounds in the domain
183 that shared the same name. In this case, the noun was followed by a prepositional phrase
184 that used the referent's location for disambiguation, for example *the grunt in the middle*.
185 When the noun resolved reference, the question continued asking for the temporal location
186 of the sound, for example *Was the grunt in the middle?*. In control trials the question asked
187 for the length of the sound, as in *Was the grunt in the middle long?*. The correct answer to
188 the question was "yes" on exactly half of the trials, counterbalanced between conditions.
189 This design assured that conditions were indistinguishable up to the critical noun.

5

190  When hearing one of the sounds by itself, it might have been difficult to judge its length, i.e.,
191  whether it was the long or the short token. However, because the question only asked for
192  the length of the sound in the control condition, this question always followed domains that
193  included both the long and the short version of the same sound, allowing for a direct com-
194  parison of the two versions.

195  A list of all possible referents was generated by permuting referent name (10 nouns), ref-
196  erent length (long or short) and referent location (first, second or third sound). One such
197  list was assigned to the resolving noun condition, another one to the ambiguous noun con-
198  dition, and a third was split between the two condition to produce 30 trials in each cell of
199  the Reference × Location design. The remaining elements of each trial were filled in using a
200  balanced randomization procedure. The same trials were presented to each participant, but
201  the order was randomized for each session.

### Visual referential domains

203  For the visual referential domains, we selected ten objects with monosyllabic names that
204  could be used to generate easily identifiable outline images (truck, star, house, car, fish, pig,
205  tree, bird, boat, cat). Outlines were created based on photos using the Inkscape vector
206  drawing application (http://www.inkscape.org). To match visual referential domains in
207  complexity to the auditory domains, where sounds were either long or short, we chose just
208  two colors, black and white.

209  Trials were constructed analogously to the trials with auditory referential domains. Figure
210  3 shows a sample trial for each condition. Each referential domain consisted of 3 objects
211  arranged vertically. All domains were constructed such that 2 objects shared the shape, and
212  2 other objects shared the color. Each trial started with presentation of a fixation cross for
213  600 ms, followed by the referential domain presented for 300 ms, and the question after an
214  interstimulus interval of 300 ms. Presentation of the questions followed the same protocol
215  as for auditory referential domains.

216  In visual domain trials, the form of the referential expression was varied as an additional
217  factor. In half of the trials referential expressions were constructed in parallel to the audi-
218  tory domain condition, with resolving noun contrasted with ambiguous nouns followed by
219  a prepositional phrase specifying the location (e.g., *the bird* vs *the bird in the middle*). In the
220  other half of the trials, the referential expression was an adjective-noun phrase, such as *the*
221  *black bird*. In those expressions, the adjective was the target. In a domain with only one
222  black item, the adjective resolved reference, while in trials with two black items the noun
223  did. Because the location was never included in the referential expression these questions
224  always asked for the location of the referent. Referential expressions with adjective targets
225  were not included with auditory referential domains because sounds would have been
226  harder to distinguish based on length alone.

227  Including reference resolving adjectives gave us an additional distinction between refer-
228  ence resolution at complete and incomplete linguistic phrases. There is evidence that the
229  status of an expression as a complete linguistic phrase interacts with referential pro-
230  cessing. Eye tracking studies suggest that listeners predict whether a potentially complete

6

231  noun phrase will be followed by a prepositional phrase or not, based on whether they could
232  resolve reference, i.e., if a noun phrase is compatible with several potential referents they
233  expect additional information to disambiguate between competing referents (Spivey,
234  Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy,
235  1995). A response to reference resolving nouns (*the bird*) could thus also indicate the com-
236  pletion of a linguistic phrase, since the comprehender would expect the phrase to be elabo-
237  rated in the control condition (*the bird in the middle*). Adjectives did not entail this contrast
238  since they were always followed by a noun (*the black bird*). In contrast, referential pro-
239  cessing should happen at incomplete phrases too. Eye tracking as well as EEG evidence
240  suggest that language comprehenders use the information in reference resolving adjec-
241  tives. For example, when participants are instructed to *Touch the starred yellow square* in a
242  context with only one starred item, they look at the target item shortly after the word
243  *starred* (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Sedivy, K. Tanenhaus,
244  Chambers, & Carlson, 1999). Similarly, readers presented with a referential expression in
245  which the adjectives allows resolving reference to an object on the left or the right side of a
246  visual display exhibit an event-related potential that is sensitive to the location of the refer-
247  ent starting around 333 ms after adjective onset (Brodbeck et al., 2015). Including adjec-
248  tives as targets thus allowed us to distinguish between phrasal and referential processing.

249  The same procedure as for auditory domains was employed to create 30 trials per condi-
250  tion in the Reference (target resolving or ambiguous) by location (top, middle, bottom) by
251  Target (noun, adjective) design (cf. Figure 3).

## Lexical variables

253  The present study was specifically designed to assess reference resolution in different con-
254  texts: in auditory and visual referential domains, and with information conveyed by a noun
255  or an adjective. Stimuli were developed with a focus on creating natural referential situa-
256  tions. Within each context, target words were identical between the resolving and ambigu-
257  ous target conditions, and since our hypotheses did not pertain to main effects of context,
258  matching target words between different contexts was not a priority. Between contexts,
259  target items differed both in variability (two adjective tokens were used whereas 10 noun
260  tokens were used per modality) and in word frequency. Lexical frequency was assessed us-
261  ing the contextual diversity variable in the SUBTL corpus (Brysbaert & New, 2009), a sim-
262  ple count variable that reflects in how many out of 8388 films and television episodes this
263  word occurred at least once. Nouns describing auditory objects were less frequent than
264  nouns describing visual objects (range 6-383 for auditory nouns and 1056-6040 for visual
265  nouns); visual adjectives fell in the range of the visual nouns (*white*: 3355; *black*: 3190).

## Procedure

267  Before the recording session, participants' head shape was digitized using an optical Fast-
268  SCAN scanner (Polhemus, VT, USA, http://www.polhemus.com). The scan included the po-
269  sitions of five marker coils that were later attached to the participant's head. At the begin-
270  ning and end of each experimental session, the position of those marker coils was recorded
271  relative to the MEG sensors, and this record was later used to coregister the head shape
272  relative to the MEG sensor positions for the source localization procedure.

7

273  Participants were familiarized with the task and introduced to the MEG recording proce-
274  dure. MEG acquisition took place in a magnetically shielded chamber, in which participants
275  were lying in a supine position with their head resting in the helmet-shaped dewar contain-
276  ing the SQUIDs. Stimuli were presented with PsychoPy (http://www.psychopy.org,
277  RRID:SCR_006571), projected onto a screen above the participants' head. Participants were
278  allowed to complete as many practice trials as they needed to feel comfortable with the
279  task with both visual and auditory referential domains. They were asked to move as little
280  as possible and to try not to blink while reading the questions. They were given the option
281  to pause the experiment after every trial by pressing the response button twice when giv-
282  ing their answer.

283  Stimuli were presented in blocks of 45 trials. Auditory and visual referential domains were
284  presented in separate interleaved blocks, and the beginning of each block announced the
285  modality of the domains, so that participants knew which modality to attend to. Within au-
286  ditory and visual domain trials, the order of trials was randomized. Since there were twice
287  as many visual as auditory trials, blocks were presented in V-A-V-V-A-V-V-… sequence. Af-
288  ter each block, participants had the opportunity to take a brief self-terminated break.

289  For 3 participants the MEG session was interrupted. One session was interrupted by a fire
290  alarm, one due to excessive tiredness and one for a bathroom break. All affected partici-
291  pants finished the experiment on the same (two participants) or a different day (one partic-
292  ipant). Marker positions from the first session were used online to match the head position
293  in the second session closely to the first session.

## Data acquisition and analysis

295  Continuous MEG was recorded with a 208 channel axial gradiometer system (Kanazawa
296  Institute of Technology, Kanazawa, Japan) at a sampling rate of 1000 Hz. Data were band
297  pass filtered between 0.1 and 200 Hz online.

### Preprocessing

299  Nonperiodic environment noise was removed from the raw data by regressing the signal
300  against 16 orthogonal reference sensors using the continuously adjusted least squares
301  method (Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001). Data were then converted
302  to the FIFF format and processed with mne-python (V 0.11, http://martinos.org/mne,
303  RRID:SCR_005972) (Gramfort et al., 2013, 2014) and additional tools available in Eelbrain
304  (V 0.22.1, http://pythonhosted.org/eelbrain, RRID:SCR_014661)

305  Bad channels were excluded from analysis based on visual inspection. Data were low-pass
306  filtered at 40 Hz. Epochs from -100 to 600 ms relative to onset of the target words were ex-
307  tracted and screened for artifacts. Epochs exceeding a +/- 2000 femto-tesla absolute
308  threshold were removed automatically. In addition, sub-threshold epochs were manually
309  removed if channels close to the eyes significantly diverged for longer than 300 ms, indicat-
310  ing presence of an ocular artifact. If an individual channel within an epoch significantly de-
311  viated from the group average, it was interpolated just for that epoch. Good epochs were

8

312 downsampled to 200 Hz, averaged per condition and baseline corrected with the 100 ms
313 pre-stimulus interval.

## Source estimation

315 Because structural MRI datasets were not available for our participants, we used the "fsav-
316 erage" average brain model included with FreeSurfer (http://surfer.nmr.mgh.harvard.edu,
317 RRID:SCR_001847) for source localization. To provide better localization accuracy, the
318 fsaverage model was scaled to match each individual's head shape, acquired before the ex-
319 periment (see Procedure). The complete co-registration procedure proceeded as follows:
320 First, the fsaverage head model was aligned with the participant's head shape by matching
321 the nasion position. The fsaverage head model was then modified using rotation and uni-
322 form scaling, with the nasion as center, to minimize the distance of the pre-auricular points
323 on the two head models using an iterative least squares procedure. Finally, minor adjust-
324 ments to the translation were made as necessary, in order to fit the head scan to the fsav-
325 erage head shape while taking skull-external properties (e.g., amount of hair) into consid-
326 eration.

327 The source space was defined on the white matter surface with the topology of a recursive-
328 ly subdivided icosahedron ("ico-4" option). Sources lying in the corpus callosum and sub-
329 cortical structures were excluded based on to the PALS-B12 atlas (Van Essen, 2005). Ana-
330 tomical areas were labelled based on the Desikan-Killiany Atlas (Desikan et al., 2006). Pre-
331 and post-rolandic gyri and the insula, where we did not expect any meaningful effects, were
332 excluded from the analysis, resulting in approximately 1950 source locations in each hemi-
333 sphere.

334 For each subject a separate inverse solution was computed based on the covariance matrix
335 from the 100 ms baseline period (good trials only). Brain activity was estimated across
336 space and time using distributed noise normalized minimum norm current source esti-
337 mates (Dale et al., 2000). Noise-normalization provides advantageous localization accuracy
338 over raw minimum norm estimates for non-superficial sources (Hauk, Wakeman, &
339 Henson, 2011), which was relevant because of the midline regions which were of interest
340 (see Introduction). For each source location, current was estimated at 3 orthogonal dipoles
341 to form a 3-dimensional current vector, of which only the length was retained to provide a
342 non-directional measure of activation. These orientation-free source estimates take into
343 account that the fsaverage brain, while providing approximate locations of anatomical fea-
344 tures, might not accurately reflect precise individual cortical folding patterns.

345 Because source activity was estimated on the same brain model, though scaled to different
346 sizes, estimates for different subjects were directly comparable without morphing data
347 from one brain model to another.

## Statistical analysis

349 The primary statistical analysis was based on a mass-univariate analysis with spatio-
350 temporal permutation cluster tests (Holmes, Blair, Watson, & Ford, 1996; Maris &
351 Oostenveld, 2007). Source estimates of condition averages for each subject entered the
352 analysis. A repeated measures ANOVA $F$-value was computed for each source at each time

353  point in a pre-specified anatomical area and time window. This *F*-map was thresholded at
354  an *F*-value corresponding to an uncorrected *p*-value of .05. Clusters were formed based on
355  direct adjacency in space and time. For each cluster, the exceedance mass was calculated
356  (the sum of all *F*-values in the cluster). The same procedure was repeated in 10000 random
357  permutations of the original data, shuffling condition labels within subject to take into ac-
358  count the within-subject nature of the design. For each permutation, the largest cluster
359  mass value was retained to form a non-parametric estimate of the distribution of the larg-
360  est cluster mass value under the null hypothesis that condition labels are exchangeable. Fi-
361  nally, a *p*-value was computed for each cluster in the original *F*-map as the proportion of
362  permutations that yielded a cluster with a larger mass than the cluster under question. We
363  report all clusters that reached a *p*-value of .05 or smaller.

364  To describe the pattern of activation in each significant cluster we then extracted and plot-
365  ted the time course of activation, as well as the average activation in the cluster. For the
366  time course, we created a region of interest (ROI) encompassing all sources that were part
367  of the cluster at any point in time. For the average activation we used the cluster as a spa-
368  tio-temporal mask to extract a single average activation value for each subject and condi-
369  tion. All plots indicate within-subject standard errors as a measure of variability (Loftus &
370  Masson, 1994).

### Analysis design and parameters

372  The goal of the primary analysis was to test whether reference resolution was associated
373  with a certain brain response across different contexts, or whether there was a response to
374  reference resolution that depended on the context, i.e., the modality of the referential do-
375  main or the target word type. We thus performed an initial 2 × 3 × 3 ANOVA with factors
376  reference (reference resolving vs ambiguous target), condition (auditory domain with noun
377  targets, visual domain with noun targets, visual domain with adjective targets) and location
378  of the referent in the domain (first/top, middle, last/bottom).

379  Prior fMRI work on referential language processing (Boiteau et al., 2014; Nieuwland et al.,
380  2007; Almor, Smith, Bonilha, Fridriksson, & Rorden, 2007) and coherent language (Ferstl et
381  al., 2008) suggest a large number of brain regions with possible involvement in reference
382  resolution. To account for this, our initial analysis included the cortex of both hemispheres
383  except for the pre- and postrolandic gyri and insula as defined in the Desikan-Killiany Atlas
384  (Desikan et al., 2006). This region is illustrated at the top right of Figure 4.

385  We used a time window of 200-600 ms relative to target word onset for this analysis. Tra-
386  ditional models from electrophysiology have suggested that lexical access occurs shortly
387  after 300 ms relative to a written word presentation (Grainger & Holcomb, 2009; Py-
388  lkkänen & Marantz, 2003). Since a word has to be recognized before it can be related to the
389  referential domain, this estimate would provide a lower bound for expected effects of ref-
390  erential processes. However, more recent evidence suggests that access to lexical and se-
391  mantic information can happen by around 200 ms (Hauk, Coutout, Holden, & Chen, 2012),
392  and shortly after that effects of semantic composition can be detected (Bemis & Pylkkänen,
393  2011). We thus used a large time window to include potential early effects.

10

394  Based on the result of a main effect of reference resolution in the initial analysis, we then
395  performed a conjunction analysis to confirm the presence of a response regardless of the
396  modality of the referential domain (Nichols, Brett, Andersson, Wager, & Poline, 2005). First,
397  two *t*-maps were created, one comparing reference resolving to ambiguous targets in audi-
398  tory domains and another one in visual domains. The conjunction map was defined as the
399  element-wise minimum of those two maps ($t_{conjunction} = \min(t_{reference\ auditory}, t_{reference\ visual})$).
400  Clusters were formed using a threshold equivalent to the one-tailed *p*-value of .05. The sig-
401  nificance of clusters was assessed with a permutation test, as described above, by repeating
402  the analysis in 10000 permutations with shuffled condition labels.

403  After the result of a main effect of reference resolution in the medial posterior region of the
404  left hemisphere, we performed a post-hoc test in a more constrained region to ascertain the
405  absence of an interaction effects which would qualify the result for the main effect. For this
406  purpose we defined a spatial region, slightly larger than the main effect cluster, using labels
407  from the Desikan-Killiany Atlas (Desikan et al., 2006). The labels included were: the para-
408  central lobule, precuneus, cuneus, pericalcarine cortex, the lingual gyrus and the posterior
409  and isthmus divisions of the cingulate gyrus. These roughly correspond to Brodmann areas
410  23, 26, 29-31, the medial aspects of 1-5 and 7 and the medial superior parts of 18 and 19.

411  Finally, we also assessed modality-specific effects of reference resolution for auditory and
412  visual domains in more targeted tests. Targeted anatomical search regions were defined
413  based on labels in the Desikan-Killiany Atlas (Desikan et al., 2006). For auditory domains,
414  we defined a search region broadly construed to auditory cortex. This region included the
415  transverse temporal, superior temporal and supramarginal gyri of both hemispheres
416  (roughly, Brodmann areas 22 and 40-42). Several lines of research suggest that auditory
417  objects are represented in the superior temporal gyrus adjacent to primary auditory areas
418  (Bizley & Cohen, 2013; Ding & Simon, 2012; Giordano, McAdams, Zatorre, Kriegeskorte, &
419  Belin, 2013). We included supramarginal gyrus because activation associated with (non-
420  linguistic) auditory imagery tends to extend into this region (McNorgan, 2012). We also
421  performed an analogous analysis for visual referential domains in the occipital lobe as de-
422  fined by the PALS-B12 atlas (Brodmann areas 17-19), although this last test was more ten-
423  tative, because brain activity related to concurrent visual presentation of the sentence
424  stimuli could interfere with our ability to record activity associated with visual referents.

425  With one exception we are not reporting any analyses of brain responses to events subse-
426  quent to the targets, because our stimuli were not designed to analyze these. As illustrated
427  in Figures 2 and 3, after the target word there were systematic differences between stimuli
428  with reference resolving targets and stimuli with ambiguous targets. The possible excep-
429  tion is nouns in the visual, adjective-noun phrase condition (see Figure 3, bottom). Here,
430  nouns following ambiguous adjectives resolve reference, whereas nouns following refer-
431  ence resolving adjectives are in a referential sense redundant. However, this prediction is
432  somewhat weakened because EEG data suggest the possibility of a reference resolution-
433  like response to nouns in adjective-noun phrases even after adjectives resolve reference
434  (Brodbeck et al., 2015). This could, for example, reflect a process of double checking that
435  the noun fits with the referent. We analyzed the response to these nouns analogously to the
436  other analyses described above, except that we used -100 to 1200 ms epochs relative to

11

437  onset of the adjective and baseline corrected before onset of the adjective to avoid a base-
438  line in a region that already differed between condition.

## Analysis of behavioral data

440  Behavioral performance was evaluated using mixed effects logistic regression models with
441  correctness of the response as binary outcome measure (see Fitzmaurice, Laird, & Ware,
442  2011). Models were fitted with the glmer command of the lme4 package (Bates, Mächler,
443  Bolker, & Walker, 2015, p. 4) in R (R Core Team, 2016). All models included random inter-
444  cepts for participant as well as item, considering each unique trial to be a separate item.
445  Significance was evaluated using type-II Wald $X^2$ tests implemented in the car package (Fox
446  & Weisberg, 2011).

447  It should be noted that the post-target regions of the stimulus sentences differ markedly
448  between sentences with reference-resolving and ambiguous targets. The analysis of behav-
449  ioral data thus does not directly speak to the difficulty of resolving reference early or late in
450  the sentence, but rather reflects a compound measure of difficulty of the different condi-
451  tions. The one effect for which conditions were adequately matched concerns the location
452  of the referent. Thus, a significant effect of referent location would indicate that the loca-
453  tion of the referent affected the difficulty of the task.

454  Finally, to test whether the results of the MEG analysis are related to behavioral perfor-
455  mance we computed the correlation between estimated brain activity and the proportion
456  of correct responses. For each subject and each condition, we extracted the mean activation
457  in the spatio-temporal region identified by the significant clusters from the MEG analysis.
458  We then calculated the difference between the response to reference resolving and ambig-
459  uous targets. To test for a correlation across subjects, we correlated this measure with the
460  total proportion of correct responses for each subject. To test for a correlation within sub-
461  ject, we computed for each subject the correlation between the difference values and pro-
462  portion correct responses in each condition (3 conditions × 3 referent locations) and sub-
463  mitted the resulting $r$ values to a one-sample $t$-test. For determining the proportion of cor-
464  rect responses we only considered those trials which also entered the MEG analysis.

## Results

## Behavioral performance

467  Among the participants included in the final analysis, behavioral performance ranged from
468  71.1% to 98.7% correct answers. A mixed effects logistic regression model with fixed ef-
469  fects Reference (2 levels: reference resolved by the target or in the post-target region), Mo-
470  dality/target (3 levels: auditory domain with noun target, visual domain with noun target
471  or visual domain with adjective target) and Location of the referent (3 levels) indicated that
472  the effect of condition interacted significantly with reference ($X^2(2) = 6.54$, $p = .038$) as well
473  as location ($X^2(4) = 9.58$, $p = .048$). To resolve these interactions we proceeded with a sepa-
474  rate analysis for auditory and visual domains.

475 In auditory domains, there were significant main effects of both reference ($X^2$(1) = 9.78, $p$ =
476 .002) and location ($X^2$(2) = 13.82. $p$ = .001). The significant main effect of reference indicat-
477 ed that the percentage of correct response was higher when reference was resolved early
478 by the noun ($M$ = 87.0% correct) than when reference was resolved late, by the preposi-
479 tional phrase ($M$ = 83.8%). This is not unexpected, since questions in the latter condition
480 were slightly more complicated. The main effect of location indicated that responses were
481 more accurate when the referent was the last of the three sounds ($M$ = 88.1%) than when it
482 was to the second ($M$ = 83.5%, $t$(23) = 3.37, $p$ = .003) or the first sound ($M$ = 84.7%, $t$(23) =
483 3.02, $p$ = .006). Behavioral performance thus indicated a recency effect with improved per-
484 formance on trials in which the referent was the most recently heard sound.

485 No effects were significant for the visual domains. Overall, responses were more accurate
486 in blocks with visual domains ($M$ = 91.1%) than in blocks with auditory domains ($M$ =
487 85.4%, $t$(23) = 6.46, $p$ < .001).

## Modality-general response

489 Our primary analysis used a spatio-temporal cluster based permutation test to find effects
490 associated with reference resolution in the neural response to target words (see Figure 2
491 and Figure 3). The test was based on a repeated measures ANOVA with a 2 (reference re-
492 solving or ambiguous target) × 3 (condition: auditory domain with noun target, visual do-
493 main with noun target or visual domain with adjective target) × 3 (location of the referent)
494 design. The test was performed in the time window from 200-600 ms, and including all
495 parts of the cerebral cortex except the pre- and postrolandic gyri and insula. Results indi-
496 cated a significant cluster for the main effect of reference in the medial parietal lobe (415-
497 600 ms, $p$ = .012, see Figure 4). In this region, activation increased whenever the target
498 word resolved reference. Time course plots for the cluster region suggested that this was
499 the case in visual as well as auditory referential domains.

500 The same test also indicated a main effect of condition at two largely symmetric sites with
501 peaks in the temporal lobes of the two hemispheres (left hemisphere 320-600 ms, $p$ < .001;
502 right hemisphere 350-600 ms, $p$ = .009). Pairwise comparison of average activation in the
503 cluster area suggests that after visual domains, the response was higher for nouns than for
504 adjectives, with an even stronger response for nouns after auditory domains. The spatial
505 distribution of the effect resembles previous reports of N400 effects localized with distrib-
506 uted minimum norm estimates of MEG data (Halgren et al., 2002; Lau, Gramfort,
507 Hamalainen, & Kuperberg, 2013; Lau, Weber, Gramfort, Hamalainen, & Kuperberg, 2014).
508 However, this result is difficult to interpret, because a number of variables were not bal-
509 anced between conditions, leading to differences in lexical frequency and predictability on
510 various levels in addition to possible baseline differences caused by differences between
511 auditory and visual referential domains. In particular, previous research suggests that lexi-
512 cal frequency and predictability interact in influencing N400 amplitudes (Dambacher,
513 Kliegl, Hofmann, & Jacobs, 2006; Van Petten & Kutas, 1990, 1991). Because of these compli-
514 cations and because this effect of condition does not influence the interpretation of our
515 primary result related to reference resolution we will not discuss it further here.

516 No other effect in the ANOVA revealed significant clusters.

517 Since our study was specifically directed at finding a response to reference resolution that
518 is present for auditory as well as visual domains, we followed up on this initial finding with
519 a test for conjunction of activity related to reference resolution after auditory and visual
520 domains. Results are shown in Figure 5. The conjunction analysis revealed a single cluster
521 in the medial posterior left hemisphere (500-600 ms, $p$ = .002) with very similar localiza-
522 tion as the main effect in the ANOVA. Figure 5 also displays plots of the activation in the
523 ROI defined by the cluster depending on condition. These plots show that reference resolu-
524 tion was associated with an increase in medial parietal activation across the different con-
525 ditions.

526 The ANOVA and the conjunction test results both suggest a shared neural response for ref-
527 erence resolution in auditory and visual domains. A difference in localization and/or timing
528 in the responses should manifest itself in a significant Reference × Condition interaction,
529 but no such interaction was found. However, this null result could also be due to the fact
530 that the primary analysis was very conservative. The spatio-temporal cluster based test,
531 correcting for multiple comparison across a large spatio-temporal region, is maximally sen-
532 sitive to spatially and temporally extended effects. However, given a largely shared re-
533 sponse, an interaction effect reflecting a temporal or spatial difference might be temporally
534 short-lived or spatially constrained. The conservative primary test might thus have missed
535 subtle interaction effects. In order to test for this possibility, we repeated the ANOVA test,
536 but constrained it to a smaller spatio-temporal region. Spatially, the test was restricted to a
537 medial posterior region in the left hemisphere, only slightly larger than the significant clus-
538 ter associated with reference resolution. Temporally, the test was restricted to the time
539 window of the ANOVA cluster, i.e., 415-600 ms. No significant Reference × Condition (or
540 higher level) interaction was found (all $p \geq$ .230). The finding that the conjunction effect
541 starts at 500 ms, while the ANOVA main effect starts at 415 ms might specifically indicate
542 that the effect differs in onset latency between auditory and visual domains. For an even
543 more liberal test of this hypothesis we repeated the same ANOVA restricted temporally to
544 the window from 415-500 ms, but also this test revealed no significant Reference × Condi-
545 tion (or higher level) interaction (all $p \geq$ .170). The lack of a significant interaction effect
546 suggests that apparent differences in timing were not reliable. In sum, we found evidence
547 for a shared response to reference resolution in auditory and visual domains in the left me-
548 dial parietal lobe, and no evidence that this response differs in its spatial or temporal dis-
549 tribution depending on the modality of the referential domain.

550 MEG activation increase in the medial parietal lobe in association with reference resolution
551 was not significantly correlated with behavioral performance. This was the case for the
552 cluster found in the ANOVA (across subjects: $r(22)$ = .30, $p$ = .159, within subjects: mean $r$ =
553 -.09, $t(23)$ = -1.25, $p$ = .22) as well as the cluster from the intersection analysis (across sub-
554 jects: $r(22)$ = .32, $p$ = .129, within subjects: mean $r$ = -.132, $t(23)$ = -1.68, $p$ = .106).

555 In the analysis of nouns in adjective-noun phrases (see Figure 3, bottom) no significant dif-
556 ference was found between reference resolving and redundant nouns.

14

### Effects in modality-specific areas

We also performed directed tests for activation related to reference resolution in brain re-gions known to be involved in modality-specific representations. For auditory domains, this included Heschl's gyrus and the superior temporal and supramarginal gyri of both hemispheres. The 2 (reference) × 3 (location) ANOVA resulted in a cluster with main effect of reference with a maximum in the vicinity of posterior auditory cortex (420-600 ms, $p <$ .001, see Figure 1). Pairwise comparison suggests that this effect was due to an increase in activation when reference was resolved.

Visual inspection of the time course of activation in Figure 6 suggests that the time course differed depending on the position of the referent. While this difference did not result in a significant Reference × Location interaction at the cluster level, the divergence might simp-ly be too short-lived for the spatio-temporal cluster method, which is maximally sensitive to effects that are extended in space and time. We thus performed a post-hoc analysis to test for an influence of referent position on the time course of activation. We extracted the average time course of activation for each of the reference resolving conditions (reference to first, second and last sound) in the spatial ROI identified by the cluster with main effect of reference. We then performed a temporal cluster-based permutation test with a one-way ANOVA (reference resolution to the first, second or last sound) in the time window around the onset of main effect of reference resolution, 350-450 ms. The test procedure was analo-gous to the spatio-temporal cluster test, except that the data was lacking a spatial dimen-sion, and clusters were formed over contiguous time points only. This test revealed a signif-icant effect of referent position (395-420 ms, $p$ = .037). The average activation in this time window was higher for reference to the last sound compared to the first ($t(23)$ = 2.49) as well as the second sound ($t(23)$ = 2.42). The first time point at which the activation to ref-erence resolving nouns differed from the average activation to non-referential nouns at an uncorrected $p ≤$ .05 was 380 ms for reference to the last sound, 490 ms for the second and 440 ms for the first. This analysis suggests a recency effect, with reference to the most re-cent sound leading to relatively earlier activation. While this finding is based on a post hoc test after a more conservative test did not result in a significant interaction effect, and thus requires empirical verification, the finding is consistent with the recency effect found in the behavioral performance data.

The corresponding analysis for activity associated with reference resolution in visual do-mains in the occipital lobe did not yield any significant clusters.

### Discussion

Our results provide evidence for a brain system recruited during referential language pro-cessing that is independent of the modality of the referents and involves the medial parietal lobe. To our knowledge this is the first report of a brain response that is associated with successful reference resolution and not tied to the modality of the referential domain. Cru-cially, the high temporal resolution of MEG allowed us to attribute this response directly to the reference resolving words, starting approximately 415 ms after word onset. Our results thus go beyond prior hemodynamic studies that implicated the same region in coherent

598  language processing, but did not target reference resolution in particular and were unable
599  to attribute the response to a specific stage of sentence comprehension.

600  Reference resolution is inherently difficult to separate from other cognitive processes asso-
601  ciated with managing situation models. At its core, we consider reference resolution to in-
602  volve identifying an entity in a mental model based on a linguistic description. However, an
603  out-of-the blue reference to an entity that is not immediately present also involves retriev-
604  ing the relevant entity from memory. For example, when captain Ahab asks a passing ship
605  "Have ye seen *the White Whale*?" (italics indicate critical expression) he brings to the atten-
606  tion of the addressees an entity that has not been present in the recent discourse or imme-
607  diate environment. In other situations, reference is made to entities that are not necessarily
608  remembered but are part of the immediate context, as when Ishmael says "Landlord! I've
609  changed my mind about that harpooneer. —I shan't sleep with him. I'll try *the bench* here."
610  While interpreting such a referential expression does not require memory retrieval, it di-
611  rects attention to the environment for a referent. In the present study we tried to create an
612  experimental situation which minimizes memory retrieval and scanning of the environ-
613  ment by presenting referential contexts immediately preceding linguistic stimuli that con-
614  tained referential expressions. The following passage constitutes a textual illustration of
615  such a situation: "The four whales slain that evening had died wide apart; one, far to wind-
616  ward; one, less distant, to leeward; one ahead; one astern. *These last three* were brought
617  alongside ere nightfall; but *the windward one* could not be reached till morning". Here,
618  memory demands should be minimal because the referents are active immediately before
619  the referential expression, although there might still be additional cognitive processes trig-
620  gered by accessing the referent, for example processes related to situating it in the referen-
621  tial domain.

## Modality-general response

623  Our main result is a response to reference resolution in the medial parietal lobe. We did not
624  find any evidence that this brain response was modulated by the modality of the referents;
625  however, this null result should be interpreted with care. It is possible that the follow-up
626  tests we performed, despite relaxed statistical criteria, were not sensitive enough to detect
627  a subtle effect. For example, it is possible that medial parietal cortex is characterized by a
628  subtle anatomical subdivision with spatially alternating areas that respond to different
629  modalities. Such an effect would be difficult to detect with MEG which has a source localiza-
630  tion accuracy in the order of tens of millimeters (cf. Hauk et al., 2011). Importantly, howev-
631  er, we showed a significant conjunction effect for visual and auditory referents, indicating
632  that even if there might be undetected differences between modalities, the medial parietal
633  lobe, broadly defined, responded to referents in both modalities.

634  Given the uncertainty inherent in MEG source localization, there are several regions that
635  could be involved in generating the reported reference-sensitive response, including
636  precuneus, posterior cingulate cortex and retrosplenial cortex. Although anatomical and
637  functional connectivity patterns suggest a more fine grained division of these regions
638  (Margulies et al., 2009), they are frequently co-activated in fMRI studies (Ranganath &
639  Ritchey, 2012). With this in mind we will discuss possible connections to other work in-
640  volving these regions as a group.

16

641  A meta-analysis of hemodynamic studies found the left precuneus among the brain regions
642  that were reliably more active for coherent language compared to incoherent language
643  (Ferstl et al., 2008). Most of the included studies compared comprehension of coherent sto-
644  ries (Xu, Kemeny, Park, Frattali, & Braun, 2005) or sentence pairs (Ferstl & von Cramon,
645  2002) to unconnected counterparts. While this is a broad contrast, a crucial component of
646  coherence is repeated reference to the same entities. Our results showing increased activity
647  directly after presentation of a reference resolving word thus go beyond these previous re-
648  sults and suggest a more specific role for medial parietal cortex in invoking known entities
649  as referents.

650  Interestingly, we observed the same response not only for complete noun phrases, like *the*
651  *bird*, but also for incomplete phrases that provided enough information to resolve refer-
652  ence, like *the black* in a context with only one black item. This suggests that this response is
653  associated with reference resolution as a cognitive process that uses information extracted
654  from language, but is not tied by its formal properties, i.e., does not have to wait for a com-
655  plete noun phrase to resolve reference. This result adds neurophysiological evidence for
656  the hypothesis that reference resolution is quick and incremental (Tanenhaus et al., 1995).

657  An involvement of medial parietal cortex in reference resolution adds a new component to
658  our understanding of medial parietal lobe function. The medial parietal lobe has been con-
659  sistently implicated in episodic memory. Brain damage involving retrosplenial cortex is as-
660  sociated with episodic memory deficits (Maguire, 2001). In addition, functional imaging
661  studies frequently associated the medial parietal lobe with episodic retrieval. This includes,
662  for example, tasks like recalling the second word of a learned word pair given the first
663  (Krause, 1999), or recalling some aspect of the context in which an item was previously
664  seen as opposed to merely recognizing that it has been previously presented (e.g.
665  Lundstrom, 2003; Lundstrom, Ingvar, & Petersson, 2005).

666  If the medial parietal lobe is associated with retrieving referents, then why were ambigu-
667  ous expressions, which were compatible with two referents, not associated with more acti-
668  vation than resolving expressions? At least two answers can be conceived of. First, it is pos-
669  sible that our participants did not retrieve referential candidates when faced with tempo-
670  rary referential ambiguity. This would be in contrast to results from visual world studies, in
671  which people listening to referentially ambiguous expressions tend to look at potential ref-
672  erents in a visual display (Tanenhaus et al., 1995). However, this difference could be due to
673  differences between the two paradigms; for example, visual world studies tend to have
674  more complex referential domains, and domains are still available for inspection at the
675  time of comprehension. A second possibility is that multiple referential candidates were in
676  fact retrieved, eliciting medial parietal lobe activation. If this was the case, our results
677  would suggest that resolving reference elicits more activation than retrieving referential
678  candidates.

679  Like episodic memory tasks, reference resolution entails access to a contextually situated
680  cognitive representation, the referent. Unlike typical episodic memory studies, in our para-
681  digm, presentation of the referential domain and reference resolution were mere seconds
682  apart. Similarly, a temporally separated study phase is not a necessary precondition for
683  medial parietal involvement in processes involving the construction of episodic scenes

17

684  such as imagining future or fictitious events (Buckner & Carroll, 2007; Schacter & Addis,
685  2007). In a synthesis of the literature, Ranganath and Ritchey (2012) suggested a functional
686  explanation for these observations. They argued that retrosplenial and parahippocampal
687  cortex, which are functionally connected with precuneus and posterior cingulate cortex
688  (Kahn, Andrews-Hanna, Vincent, Snyder, & Buckner, 2008), are part of a memory system
689  for context memory. In the domain of language processing, this system is thought to be re-
690  sponsible for constructing discourse or situation models (Zwaan, 2015). Since reference
691  resolution involves locating referents in such context representations, our findings provide
692  more specific support for this hypothesis, which has the potential to connect research on
693  language and memory.

694  This interpretation is also consistent with the results from fMRI investigations of referen-
695  tial processing using text stimuli. One study found that stimuli containing conjoined refer-
696  ents (e.g., *Jeremy and Roger*) were associated with increased activity in the precuneus
697  (Boiteau et al., 2014). A similar parietal response, also extending to more dorsal areas, was
698  observed when a passage repeatedly referred to the same person by name instead of with a
699  pronoun (Almor et al., 2007). The authors argued that, because of to the strong expectation
700  that repeated references to the same person should use pronouns, readers temporarily in-
701  terpreted the repeated name as introducing a new referent, and that parietal regions were
702  involved in handling the additional discourse referent and resolving its coreference with
703  the first name.

704  Another study localized the response to sentences containing referentially ambiguous pro-
705  nouns, as in *Ronald told Frank that he ...*, and found a strong peak in the medial parietal lobe
706  (Nieuwland et al., 2007). Superficially, this is the opposite of our finding, which was more
707  medial parietal activity for unambiguous reference. However, our stimuli only contained
708  temporary ambiguities that were resolved after at most 3 additional words (*the grunt at*
709  *the beginning*), while the fMRI study involved permanent ambiguities that were never re-
710  solved. The low temporal resolution of fMRI cannot distinguish between a direct response
711  to the ambiguous pronouns and a later, possibly extended response to ambiguous sentenc-
712  es. Indeed, the BOLD signal was modeled as an extended response from the critical pro-
713  noun up to the onset of the sentence final word (p. 996). Rather than contradicting our re-
714  sults, this finding thus suggests that medial parietal lobe activation is not only involved
715  when a referent is found, but can also be recruited in situations that require dealing with
716  openly ambiguous referential expressions, which might reflect an extended search through
717  a situation model for possible referents. Finally, a study that focused on explicit pronoun
718  co-reference judgements implicated in particular the lateral frontal lobes for processing of
719  ambiguities, which could thus be related to higher level decision making demands
720  (McMillan, Clark, Gunawardena, Ryant, & Grossman, 2012). Together, these results suggest
721  a more complex picture of the processing of referential ambiguities with functionally sepa-
722  rable contributions from different brain regions.

723  Our design resembles short term memory tasks, in which a recognition task follows
724  presentation of study items after only seconds. This analogy suggests the alternative possi-
725  bility that the medial parietal lobe is involved merely in recalling the spatiotemporal prop-
726  erties of the referents. While our design by itself cannot exclude this possibility, it is ren-
727  dered less plausible when considering other research: short term memory tasks for visual

728  features tend to engage lateral rather than medial parietal regions (e.g. Bettencourt & Xu,
729  2015; Kawasaki, Watanabe, Okuda, Sakagami, & Aihara, 2008; Todd & Marois, 2004), nor
730  do auditory short term memory tasks tend to engage the medial parietal lobe (e.g. Kumar et
731  al., 2016). On the other hand, medial parietal regions are recruited by tasks that involve
732  judgements of complex spatial or temporal relations (Galati, Pelle, Berthoz, & Committeri,
733  2010; Kwok & Macaluso, 2015), which is more consistent with an involvement in relational
734  models as argued above. Furthermore, an explanation based solely on short term memory
735  access might have difficulty accounting for results from the hemodynamic studies dis-
736  cussed above, which implicated the medial parietal lobe in referential language processing
737  while using linguistic stimuli.

738  More generally, it could be argued that the medial parietal lobe is involved in encoding per-
739  ceptual as opposed to semantic relations. This explanation could account for fMRI studies
740  of language processing, assuming that participants in those studies maintained perceptual
741  models during language comprehension. This assumption is reasonable given evidence that
742  even purely linguistic stimuli seem to be encoded involving modality-specific systems (see
743  source cited in the Introduction). At least, the evidence presented here suggests that repre-
744  sentations in the medial parietal lobe are not purely visual, but multisensory. This issue al-
745  so highlights a larger question for future research concerning the extent to which sensory
746  representations are essential to discourse models, or whether they are epiphenomenal.

747  A further possibility is that the medial parietal lobe response reflects a priming effect of the
748  ambiguous targets, because ambiguous targets were applicable to two entities in the refer-
749  ential domain, whereas reference resolving words were applicable only to one. We consid-
750  er this explanation unlikely because semantic priming effects tend to start earlier, even
751  when sounds prime words (Vanpetten & Rheinfelder, 1995), and localize to temporal and
752  sometimes frontal regions (Lau, Phillips, & Poeppel, 2008; Lau et al., 2013, 2014). Both
753  characteristics apply to the N400-like responses we observed as main effects between tar-
754  get conditions, suggesting that we did have the power to detect such priming effects, had
755  they been present.

756  In sum, our results, together with the literature on the medial parietal lobe, suggest that
757  this region is involved in representing entities with contextual associations, and thus in
758  maintaining situation models for language comprehension.

## Activity in modality-specific brain systems

760  Our results suggest that resolving reference to a previously heard sound is associated with
761  an increase in activation in the vicinity of auditory cortex. This response occurred in the
762  absence of concurrent auditory stimulation, suggesting that it is associated with retrieval of
763  sensory properties of the referent. More detailed analysis suggested that reference to the
764  most recent sound was associated with faster activation, consistent with theories suggest-
765  ing that the most recent in a sequence of items stored in short term memory is more acces-
766  sible (McElree & Dosher, 1989).

767  The localization of this effect in the vicinity of auditory cortex is compatible with auditory
768  imagery, which is associated with activation posterior to Heschl's gyrus bilaterally

769 (McNorgan, 2012), although the same regions might also encode more abstract information
770 (Linke & Cusack, 2015). In our study the response was left-lateralized, paralleling neu-
771 roimaging (McNorgan, 2012) as well as behavioral (Prete, Marzoli, Brancucci, & Tommasi,
772 2016) evidence for a tendency towards functional left-lateralization of imagery.

773 Given that this response likely reflects activation of modality-specific representations, the
774 direction of the effect is interesting. Ambiguous nouns were compatible with two auditory
775 objects; activating two tokens as part of the referential search could lead to more activity
776 than activating a single token. Instead, we found that auditory cortex becomes more active
777 once a unique referent is found. This result suggests that auditory cortex becomes more
778 involved not during the search for a referent but when the referent is found, consistent
779 with the need to retrieve a more detailed representation of the referent to answer the
780 question.

781 The parallel test for activity associated with reference resolution to visual referents did not
782 reveal any significant clusters. While one possibility is that visual referential domains were
783 processed differently from the auditory domains, there are other possible explanations for
784 this null result. Processing the visually presented words could have interfered with, or
785 overshadowed a response associated with the referential domain. Reference resolution in
786 auditory domains was not associated with such a competing process, as words were pre-
787 sented visually. Furthermore, previous results suggest that reference resolution in visual
788 domains is associated with a brain response that depends on the spatial position of the ref-
789 erent along the horizontal axis (Brodbeck et al., 2015), consistent with results indicating
790 that visual short-term memory access is associated with a brain response that depends on
791 the horizontal position of the item that is accessed (Hopf et al., 2000; Hopf, Boelmans,
792 Schoenfeld, Luck, & Heinze, 2004). However, in the present study, referents were arranged
793 vertically, and to our knowledge, location-dependent activity in the vertical dimension is
794 not an established phenomenon.

795 While the behavioral results paralleled the recency-effect observed in the auditory do-
796 mains, with better performance when the referent was the most recently presented sound,
797 the location of the referent did not influence behavioral performance in visual domains.
798 This could reflect the fact that in visual domains all referents were presented concurrently,
799 contrasting with the serial presentation of the auditory referents.

## Models of language comprehension

801 The response in auditory cortex started at practically the same time as the medial parietal,
802 modality-general response; if considering reference to the most recent sound, it preceded
803 it. This observation constrains the functional interpretation of the two responses. Specifi-
804 cally, the medial parietal response does not seem to be the earliest response reflecting ref-
805 erence resolution. On the other hand, we cannot exclude the possibility that reference reso-
806 lution involves modality-specific systems, while the medial parietal response reflects a
807 subsequent process such as integrating the referent in a situation model.

808 In the context of models of visual word perception (e.g. Pylkkänen & Marantz, 2003; Grain-
809 ger & Holcomb, 2009) the onset around 400 ms puts our effects in a post-lexical time win-

20

810 dow. This is in agreement with a sequential model in which reference resolution follows
811 lexical processing. Consistent with this, we observed an N400-like response with an onset
812 around 320 ms. Given that the N400 is thought to reflect access to lexical information (Lau
813 et al., 2014) this suggests that activation of lexical information preceded reference resolu-
814 tion.

815 By describing a neural response to reference resolution, our results add a critical compo-
816 nent to the sequence of computational steps in language comprehension that can be
817 tracked with electrophysiology. Our results thus offer not only novel insights into the neu-
818 ral basis of reference resolution, but also new tools to study language processing.

## Conclusions

820 Our results provide evidence against a model of referential language processing in which
821 semantic language representations interact with representations of the referential domain
822 exclusively in modality specific brain systems. Instead, our findings suggest that a brain
823 system including a medial parietal region supports referential language processing with an
824 increase in activity when the referent of an expression can be resolved. This finding pro-
825 vides a crucial bridge between language processing and the memory literature which at-
826 tributes context representations based on situation models to medial parietal brain regions
827 (Ranganath & Ritchey, 2012). In addition, our results provide a possible explanation for the
828 consistent observation of medial parietal activity during tasks involving coherent language
829 (Ferstl et al., 2008; Boiteau et al., 2014; Nieuwland et al., 2007). While our finding thus pro-
830 vides a considerable advance in region-function mapping, reference resolution is not a
831 monolithic process, and it will be a task for future research to clarify the precise computa-
832 tional steps that engage medial parietal and modality-specific regions.

## References

834 Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-

835 periodic environmental magnetic noise in MEG measurement by continuously adjusted

836 least squares method. *IEEE Transactions on Applied Superconductivity*, *11*(1), 669–672.

837 https://doi.org/10.1109/77.919433

838 Almor, A., Smith, D. V., Bonilha, L., Fridriksson, J., & Rorden, C. (2007). What is in a name?

839 Spatial brain circuits are used to track discourse references. *Neuroreport*, *18*(12), 1215–

840 1219. https://doi.org/10.1097/WNR.0b013e32810f2e11

841 Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world:

842       the "blank screen paradigm." *Cognition*, *93*(2), B79–B87.

843       https://doi.org/10.1016/j.cognition.2004.02.005

844 Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between lan-

845       guage and the visual world: eye movements and mental representation. *Cognition*, *111*,

846       55–71. https://doi.org/10.1016/j.cognition.2008.12.005

847 Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*,

848       *22*(4), 637–660. https://doi.org/10.1017/S0140525X99532147

849 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models

850       Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

851       https://doi.org/10.18637/jss.v067.i01

852 Bemis, D. K., & Pylkkänen, L. (2011). Simple Composition: A Magnetoencephalography Inves-

853       tigation into the Comprehension of Minimal Linguistic Phrases. *The Journal of Neurosci-*

854       *ence*, *31*(8), 2801–2814. https://doi.org/10.1523/JNEUROSCI.5003-10.2011

855 Bettencourt, K. C., & Xu, Y. (2015). Decoding the content of visual short-term memory under

856       distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150–157.

857       https://doi.org/10.1038/nn.4174

858 Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception.

859       *Nature Reviews Neuroscience*, *14*(10), 693–707. https://doi.org/10.1038/nrn3565

860 Boiteau, T. W., Bowers, E., Nair, V. A., & Almor, A. (2014). The neural representation of plural

861       discourse entities. *Brain and Language*, *137*, 130–141.

862       https://doi.org/10.1016/j.bandl.2014.08.003

22

863    Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some

864         investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Be-*

865         *havior*, *11*(6), 717–726. https://doi.org/10.1016/S0022-5371(72)80006-9

866    Brodbeck, C., Gwilliams, L., & Pylkkänen, L. (2015). EEG can track the time course of success-

867         ful reference resolution in small visual worlds. *Frontiers in Psychology*, *6*, 1787.

868         https://doi.org/10.3389/fpsyg.2015.01787

869    Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of

870         current word frequency norms and the introduction of a new and improved word frequen-

871         cy measure for American English. *Behav Res Methods*, *41*, 977–90.

872         https://doi.org/10.3758/BRM.41.4.977

873    Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sci-*

874         *ences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

875    Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., &

876         Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG

877         for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67.

878         https://doi.org/10.1016/S0896-6273(00)81138-1

879    Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability

880         effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103.

881         https://doi.org/10.1016/j.brainres.2006.02.010

882    Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany,

883         R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on

884         MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980.

885         https://doi.org/10.1016/j.neuroimage.2006.01.021

886    Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listen-

887        ing to competing speakers. *Proc Natl Acad Sci U S A*, *109*, 11854–9.

888        https://doi.org/10.1073/pnas.1205381109

889    Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye

890        movements as a window into real-time spoken language comprehension in natural con-

891        texts. *Journal of Psycholinguistic Research*, *24*(6), 409–436.

892        https://doi.org/10.1007/BF02143160

893    Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language

894        network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain*

895        *Mapping*, *29*(5), 581–593. https://doi.org/10.1002/hbm.20422

896    Ferstl, E. C., & von Cramon, D. Y. (2002). What Does the Frontomedian Cortex Contribute to

897        Language Processing: Coherence or Theory of Mind? *NeuroImage*, *17*(3), 1599–1612.

898        https://doi.org/10.1006/nimg.2002.1247

899    Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed).

900        Hoboken, N.J: Wiley.

901    Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second). Thousand

902        Oaks CA: Sage. Retrieved from

903        http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

904    Galati, G., Pelle, G., Berthoz, A., & Committeri, G. (2010). Multiple reference frames used by

905        the human brain for spatial perception and memory. *Experimental Brain Research*,

906        *206*(2), 109–120. https://doi.org/10.1007/s00221-010-2168-8

907    Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2013). Abstract En-

908        coding of Auditory Objects in Cortical Activity Patterns. *Cerebral Cortex*, *23*(9), 2025–

909        2037. https://doi.org/10.1093/cercor/bhs162

910    Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review*

911        *of Psychology*, *48*(1), 163–189. https://doi.org/10.1146/annurev.psych.48.1.163

912    Grainger, J., & Holcomb, P. J. (2009). Watching the Word Go by: On the Time-course of Com-

913        ponent Processes in Visual Word Recognition. *Language and Linguistics Compass*, *3*(1),

914        128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x

915    Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämä-

916        läinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neu-*

917        *roscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267

918    Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., … Hämä-

919        läinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*,

920        446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

921    Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., &

922        Dale, A. M. (2002). N400-like Magnetoencephalography Responses Modulated by Se-

923        mantic Context, Word Frequency, and Lexical Class in Sentences. *NeuroImage*, *17*(3),

924        1101–1116. https://doi.org/10.1006/nimg.2002.1268

925    Hauk, O., Coutout, C., Holden, A., & Chen, Y. (2012). The time-course of single-word reading:

926        Evidence from fast behavioral and brain responses. *NeuroImage*, *60*(2), 1462–1477.

927        https://doi.org/10.1016/j.neuroimage.2012.01.061

928    Hauk, O., Davis, M. H., Kherif, F., & Pulvermüller, F. (2008). Imagery or meaning? Evidence

929        for a semantic origin of category-specific brain activity in metabolic imaging. *European*

930 *Journal of Neuroscience*, *27*(7), 1856–1866. https://doi.org/10.1111/j.1460-

931 9568.2008.06143.x

932 Hauk, O., Wakeman, D. G., & Henson, R. N. (2011). Comparison of noise-normalized minimum

933 norm estimates for MEG analysis using multiple resolution metrics. *NeuroImage*, *54*(3),

934 1966–1974. https://doi.org/10.1016/j.neuroimage.2010.09.053

935 Holmes, A. P., Blair, R. C., Watson, G., & Ford, I. (1996). Nonparametric Analysis of Statistic

936 Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow & Me-*

937 *tabolism*, *16*(1), 7–22. https://doi.org/10.1097/00004647-199601000-00002

938 Hopf, J. M., Boelmans, K., Schoenfeld, M. A., Luck, S. J., & Heinze, H. J. (2004). Attention to

939 features precedes attention to locations in visual search: Evidence from electromagnetic

940 brain responses in humans. *J Neurosci*, *24*, 1822–1832.

941 https://doi.org/10.1523/Jneurosci.3564-03.2004

942 Hopf, J. M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., & Heinze, H. J.

943 (2000). Neural sources of focused attention in visual search. *Cereb Cortex*, *10*, 1233–

944 1241. https://doi.org/10.1093/cercor/10.12.1233

945 Kahn, I., Andrews-Hanna, J. R., Vincent, J. L., Snyder, A. Z., & Buckner, R. L. (2008). Distinct

946 Cortical Anatomy Linked to Subregions of the Medial Temporal Lobe Revealed by In-

947 trinsic Functional Connectivity. *Journal of Neurophysiology*, *100*(1), 129–139.

948 https://doi.org/10.1152/jn.00077.2008

949 Kawasaki, M., Watanabe, M., Okuda, J., Sakagami, M., & Aihara, K. (2008). Human posterior

950 parietal cortex maintains color, shape and motion in visual short-term memory. *Brain Re-*

951 *search*, *1213*, 91–97. https://doi.org/10.1016/j.brainres.2008.03.037

952    Krause, B. J. (1999). Episodic retrieval activates the precuneus irrespective of the imagery con-

953        tent of word pair associates: A PET study. *Brain*, *122*(2), 255–263.

954        https://doi.org/10.1093/brain/122.2.255

955    Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016). A

956        Brain System for Auditory Working Memory. *Journal of Neuroscience*, *36*(16), 4492–

957        4505. https://doi.org/10.1523/JNEUROSCI.4341-14.2016

958    Kwok, S. C., & Macaluso, E. (2015). Immediate memory for "when, where and what": Short-

959        delay retrieval using dynamic naturalistic material: Neural Correlates of Immediate Epi-

960        sodic Retrieval. *Human Brain Mapping*, *36*(7), 2495–2513.

961        https://doi.org/10.1002/hbm.22787

962    Lau, E. F., Gramfort, A., Hamalainen, M. S., & Kuperberg, G. R. (2013). Automatic Semantic

963        Facilitation in Anterior Temporal Cortex Revealed through Multimodal Neuroimaging.

964        *Journal of Neuroscience*, *33*(43), 17174–17181.

965        https://doi.org/10.1523/JNEUROSCI.1018-13.2013

966    Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing

967        the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933.

968        https://doi.org/10.1038/nrn2532

969    Lau, E. F., Weber, K. P., Gramfort, A., Hamalainen, M. S., & Kuperberg, G. R. (2014). Spatio-

970        temporal Signatures of Lexical-Semantic Prediction. *Cereb Cortex*.

971        https://doi.org/10.1093/cercor/bhu219

972    Linke, A. C., & Cusack, R. (2015). Flexible Information Coding in Human Auditory Cortex dur-

973        ing Perception, Imagery, and STM of Complex Sounds. *Journal of Cognitive Neurosci-

974        ence*, *27*(7), 1322–1333. https://doi.org/10.1162/jocn_a_00780

975 Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs.

976 *Psychonomic Bulletin & Review*, *1*(4), 476–490. https://doi.org/10.3758/BF03210951

977 Lundstrom, B. N. (2003). Isolating the retrieval of imagined pictures during episodic memory:

978 activation of the left precuneus and left prefrontal cortex. *NeuroImage*, *20*(4), 1934–

979 1943. https://doi.org/10.1016/j.neuroimage.2003.07.017

980 Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior

981 frontal cortex during source memory episodic retrieval. *NeuroImage*, *27*(4), 824–834.

982 https://doi.org/10.1016/j.neuroimage.2005.05.008

983 Maguire, E. A. (2001). The retrosplenial contribution to human navigation: a review of lesion

984 and neuroimaging findings. *Scandinavian Journal of Psychology*, *42*(3), 225–238.

985 https://doi.org/10.1111/1467-9450.00233

986 Margulies, D. S., Vincent, J. L., Kelly, C., Lohmann, G., Uddin, L. Q., Biswal, B. B., …

987 Petrides, M. (2009). Precuneus shares intrinsic functional architecture in humans and

988 monkeys. *Proceedings of the National Academy of Sciences*, *106*(47), 20069–20074.

989 https://doi.org/10.1073/pnas.0905314106

990 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.

991 *Journal of Neuroscience Methods*, *164*(1), 177–190.

992 https://doi.org/10.1016/j.jneumeth.2007.03.024

993 McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The

994 time course of recognition. *Journal of Experimental Psychology: General*, *118*(4), 346–

995 373. https://doi.org/10.1037/0096-3445.118.4.346

996 McMillan, C. T., Clark, R., Gunawardena, D., Ryant, N., & Grossman, M. (2012). fMRI evi-

997      dence for strategic decision-making during resolution of pronoun reference. *Neuropsy-*

998      *chologia*, *50*(5), 674–687. https://doi.org/10.1016/j.neuropsychologia.2012.01.004

999 McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural cor-

1000      relates of modality-specific and modality-general imagery. *Frontiers in Human Neuro-*

1001      *science*, *6*. https://doi.org/10.3389/fnhum.2012.00285

1002 Nichols, T. E., Brett, M., Andersson, J., Wager, T. D., & Poline, J. B. (2005). Valid conjunction

1003      inference with the minimum statistic. *NeuroImage*, *25*, 653–60.

1004      https://doi.org/10.1016/j.neuroimage.2004.12.005

1005 Nieuwland, M. S., Petersson, K. M., & Van Berkum, J. J. A. (2007). On sense and reference: ex-

1006      amining the functional neuroanatomy of referential processing. *NeuroImage*, *37*(3), 993–

1007      1004. https://doi.org/10.1016/j.neuroimage.2007.05.048

1008 Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The

1009      representation of semantic knowledge in the human brain. *Nat Rev Neurosci*, *8*, 976–87.

1010      https://doi.org/10.1038/nrn2277

1011 Prete, G., Marzoli, D., Brancucci, A., & Tommasi, L. (2016). Hearing it right: Evidence of hemi-

1012      spheric lateralization in auditory imagery. *Hearing Research*, *332*, 80–86.

1013      https://doi.org/10.1016/j.heares.2015.12.011

1014 Pulvermuller, F. (2013). How neurons make meaning: brain mechanisms for embodied and ab-

1015      stract-symbolic semantics. *Trends Cogn Sci*, *17*, 458–70.

1016      https://doi.org/10.1016/j.tics.2013.06.004

29

1017 Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG.

1018      *Trends in Cognitive Sciences*, *7*(5), 187–189. https://doi.org/10.1016/S1364-

1019      6613(03)00092-5

1020 R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria:

1021      R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

1022 Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Na-

1023      ture Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

1024 Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: re-

1025      membering the past and imagining the future. *Philosophical Transactions of the Royal

1026      Society B: Biological Sciences*, *362*(1481), 773–786.

1027      https://doi.org/10.1098/rstb.2007.2087

1028 Sedivy, J. C., K. Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incre-

1029      mental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–

1030      147. https://doi.org/10.1016/S0010-0277(99)00025-6

1031 Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory:

1032      eye movements to absent objects. *Psychological Research-Psychologische Forschung*,

1033      *65*, 235–241. https://doi.org/10.1007/s004260100059

1034 Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and

1035      spoken language comprehension: Effects of visual context on syntactic ambiguity resolu-

1036      tion. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-

1037      0285(02)00503-0

1038  Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration

1039       of visual and linguistic information in spoken language comprehension. *Science*,

1040       *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

1041  Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior

1042       parietal cortex. *Nature*, *428*(6984), 751–754. https://doi.org/10.1038/nature02466

1043  Van Essen, D. C. (2005). A Population-Average, Landmark- and Surface-based (PALS) atlas of

1044       human cerebral cortex. *NeuroImage*, *28*(3), 635–662.

1045       https://doi.org/10.1016/j.neuroimage.2005.06.058

1046  Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequen-

1047       cyinevent-related brainpotentials. *Memory & Cognition*, *18*(4), 380–393.

1048       https://doi.org/10.3758/BF03197127

1049  Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and

1050       closed-class words. *Memory & Cognition*, *19*(1), 95–112.

1051       https://doi.org/10.3758/BF03198500

1052  Vanpetten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and

1053       environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*(4),

1054       485–508. https://doi.org/10.1016/0028-3932(94)00133-A

1055  Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent

1056       features of word, sentence, and narrative comprehension. *NeuroImage*, *25*(3), 1002–

1057       1015. https://doi.org/10.1016/j.neuroimage.2004.12.013

1058  Zwaan, R. A. (2015). Situation models, mental simulations, and abstract concepts in discourse

1059       comprehension. *Psychon Bull Rev*. https://doi.org/10.3758/s13423-015-0864-x

1060

## Figure legends

*Figure 1: Modality-specific and -general representations: Possible flow of information between modality-general and modality-specific representations of both linguistic meaning and referents. Information originates at the top in linguistic form information, i.e., written or spoken words. This input triggers retrieval of meanings, which in turn are used to find referents. Some theories suggest the existence of a "semantic hub" that is involved in processing semantic representations regardless of the modality of their content (Patterson, Nestor, & Rogers, 2007), while others suggest that meaning is exclusively represented in modality-specific brain systems (Pulvermuller, 2013). Both of these theories are potentially compatible with a model in which referents are represented exclusively in modality-specific brain systems (blue and red arrows, respectively). This hypothesis would suggest that the "referents" box at the bottom left can be removed from the diagram, since referential processing is constrained to modality-specific representations.*

*Figure 2: Design for auditory referential domains. Time proceeds from left to right, with vertical offset indicating elements that differ between conditions. The arrow indicates the target word for analysis. The upper sequence illustrates a trial in which the target word grunt resolves reference, whereas the lower sequence illustrates a trial in which it does not, and reference is resolved by adding the prepositional phrase in the middle. Below the displays, presentation time of each frame is indicated in ms.*

*Figure 3: Design for visual referential domains. Time proceeds from left to right, with vertical offset indicating elements that differ between condition. The arrow indicates the target word for analysis. For both trials with noun targets and trials with adjective targets, the upper sequence illustrates a trial in which the target resolves reference, and the lower sequence illustrates a trial in which reference is resolved later in the sentence. Below the displays, presentation time of each frame is indicated in ms.*

*Figure 4: Full model ANOVA results. The ANOVA analysis revealed a significant spatio-temporal cluster for the main effect of reference resolution in the medial parietal lobe of the left hemisphere, indicating an activity increase associated with reference resolution starting at 415 ms after target word onset.* **Top left**: *Anatomical extent of the cluster, shown on a lateral and a medial view of the left hemisphere. Each source that was part of the cluster at any time point is color coded, indicating the sum of F-values over time, i.e., how much that particular source contributed to the cluster's mass.* **Top right**: *The anatomical region included in the test is indicated in blue; the left hemisphere is shown for illustration, but the test also included the corresponding area of the right hemisphere.* **Bottom**: *The time course of activation in the ROI defined by the cluster in response to the target words, shown separately for auditory and visual referential domains for illustration purposes. Time 0 ms corresponds to the onset of the visual presentation of the target word. The time course plots indicate within-subject standard errors. The time region covered by the cluster is indicated with gray shading (415-600 ms).*
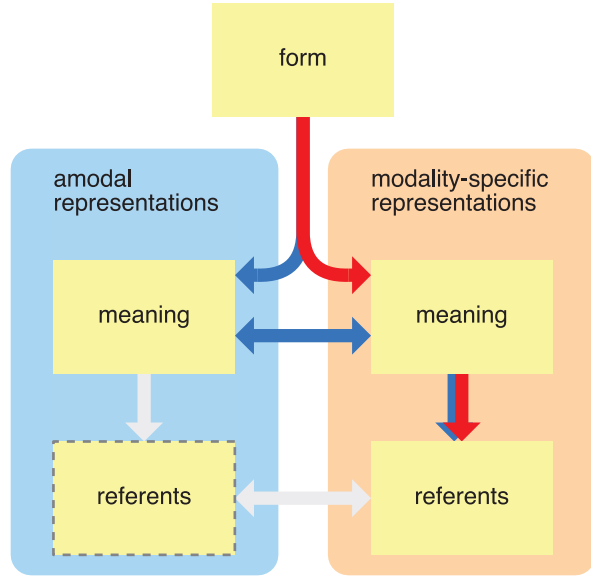
*Figure 5: Test for conjunction of reference resolution in auditory and visual referential domains.* **Top**: *Anatomical extent of the cluster, showing for each anatomical source the sum of the t-values over time, i.e., indicating how much each source contributes to the cluster mass statistic (details analogous to Figure 4).* **Below**: *activation in the ROI defined by the cluster in*
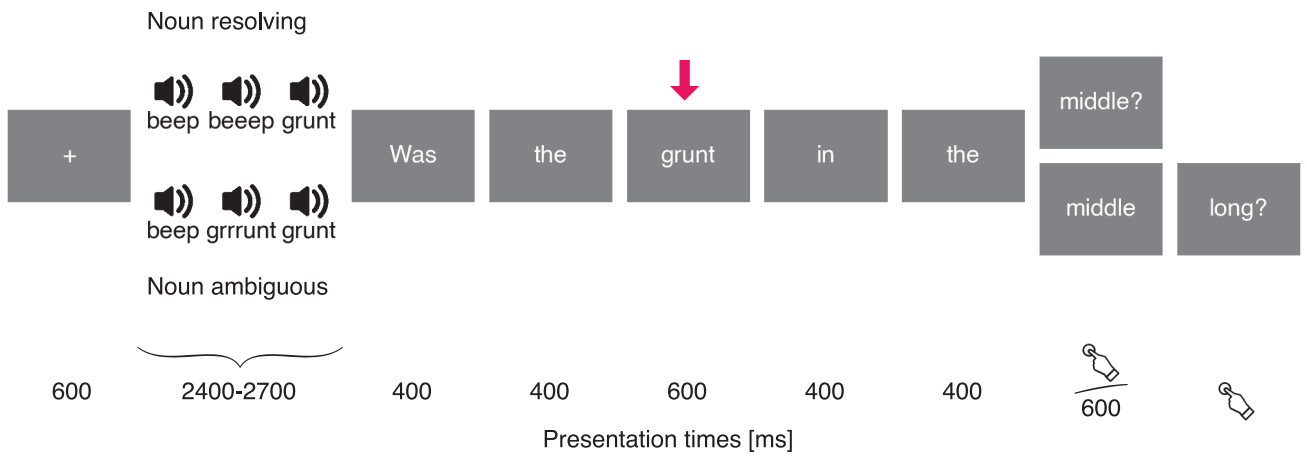
*the different conditions. **Left**: schematic depiction of the referential domains for each condi-
tion, and a corresponding example target word. Referents are color-coded for clarity of the
results only – referents in the visual referential domains that were presented to subjects were
always black and white. **Middle**: the time course of activation in response to the target word
presentation in the region of interest defined by the spatial extent of the cluster. The time re-
gion covered by the cluster is indicated with gray shading. **Right**: bar-plots show the average
activation for each condition in the spatio-temporal region covered by the cluster. In both
time- and bar-plots, variability is indicated with the within-subject standard error.*

*Figure 6: Results in auditory cortex. A significant cluster indicated an activation increase
when reference was resolved to auditory objects. The time course of activation in the ROI de-
fined by the cluster is plotted separately according to the sequential position of the referent in
the auditory domain. For each plot, the relevant position of the referent is indicated in the
schematic of the referential domain. Other details are analogous to previous figures.*

Noun resolving

🔊 🔊 🔊
beep beeep grunt

🔊 🔊 🔊
beep grrrunt grunt

Noun ambiguous

+ Was the grunt in the middle? middle long?

600    2400-2700    400    400    600    400    400    600

Presentation times [ms]

## Noun target

Noun resolving

Noun ambiguous

| + | | Was | the | bird | in | the | middle? |
| | | | | | | | middle | black? |

| 600 | 300/300 ISI | 400 | 400 | 600 | 400 | 400 | 600 |

Presentation times [ms]

## Adjective target

Adjective resolving

Adjective ambiguous

| + | | Was | the | black | bird | in | the | middle? |

| 600 | 300/300 ISI | 400 | 400 | 600 | 600 | 400 | 400 |

Presentation times [ms]

## Main effect of reference resolution



Sum of F-values

250
200
150
100
50
0

Region for spatio-temporal clusters:

Time-course in cluster-based ROI:

## Auditory Domains



Activation [dSPM]

Reference resolving
Reference ambiguous

## Visual Domains



time [ms]

Reference resolving
Reference ambiguous
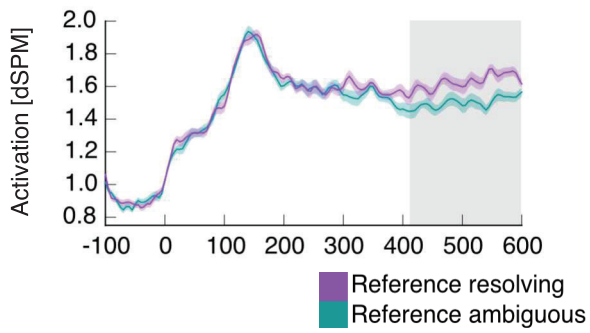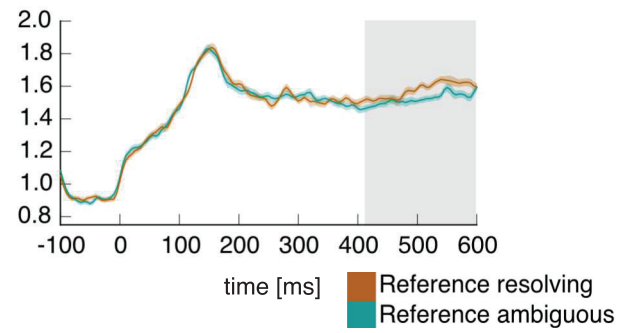
Reference resolution in visual domains ∩ reference resolution in auditory domains



Region for spatio-temporal clusters:
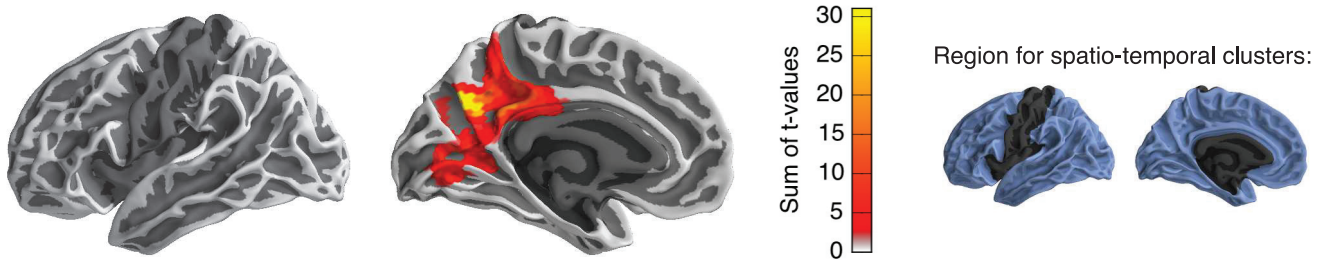
Auditory domains, noun targets:

grunt

Time-course in cluster-based ROI:
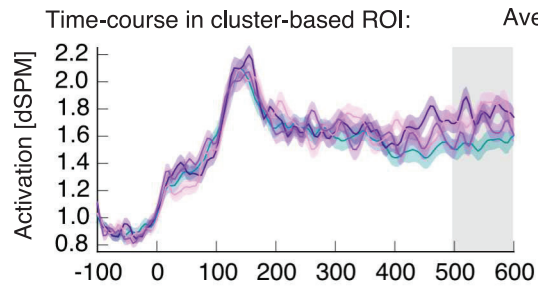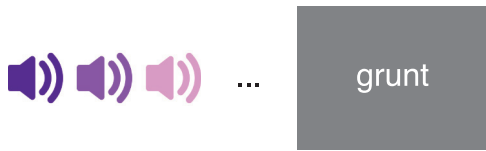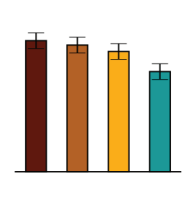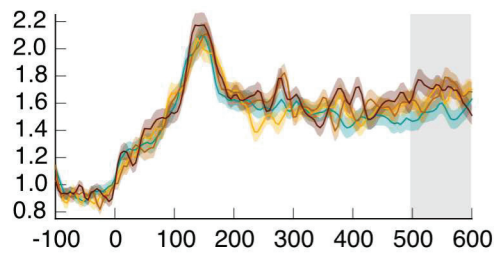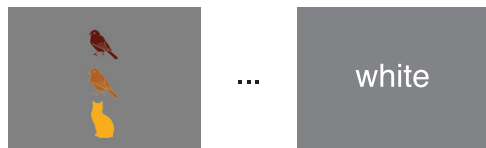
Average in cluster:
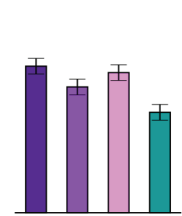
- Reference to first sound
- Reference to second sound
- Reference to third sound
- Reference ambiguous

Visual domains, noun targets:

cat
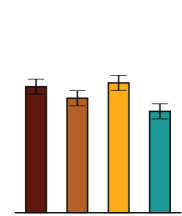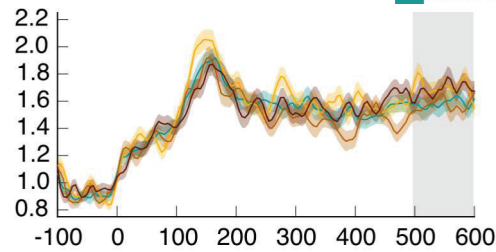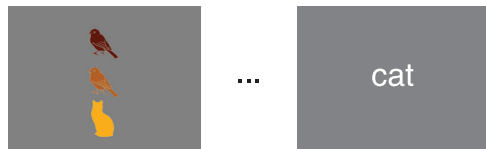
Visual domains, adjective targets:

white

Time [ms]

- Reference to upper item
- Reference to middle item
- Reference to lower item
- Reference ambiguous