1

2

Sign and speech share partially overlapping conceptual representations

Samuel Evans[*1,2], Cathy J. Price[3], Jörn Diedrichsen[4], Eva Gutierrez-Sigut[1,5] &
Mairéad MacSweeney[1,5]

[1]Institute of Cognitive Neuroscience, UCL, Alexandra House, 17-19 Queen Square,
London, WC1N 3AZ, UK
[2]Dept of Psychology, University of Westminster, 101 New Cavendish St, London,
W1W 6XH, UK
[3]Wellcome Centre for Human Neuroimaging, UCL, 12 Queen Square, London,
WC1N 3AR, UK
[4]Brain and Mind Institute, University of Western Ontario, London, Ontario, N6A 3K7,
Canada
[5]Deafness, Cognition and Language Research Centre (DCAL), 49 Gordon Square,
London, WC1H 0PD, UK

17

18

19

20

21

22

23

24

25

[*]Lead contact

Dr Samuel Evans, Institute of Cognitive Neuroscience, UCL, Alexandra House, 17
Queen Square, London WC1N 3AZ.

Email: S.Evans1@westminster.ac.uk

Summary

Conceptual knowledge is fundamental to human cognition. Yet the extent to which it is influenced by language is unclear. Studies of semantic processing show that similar neural patterns are evoked by the same concepts presented in different modalities (e.g. spoken words and pictures or text) [1–3]. This suggests that conceptual representations are 'modality independent'. However, an alternative possibility is that the similarity reflects retrieval of common spoken language representations. Indeed, in hearing spoken language users, text and spoken language are co-dependent [4,5] and pictures are encoded via visual and verbal routes [6]. A parallel approach investigating semantic cognition, shows that bilinguals activate similar patterns for the same words in their different languages [7,8]. This suggests that conceptual representations are 'language independent'. However, this has only been tested in spoken language bilinguals. If different languages evoke different conceptual representations, this should be most apparent comparing languages that differ greatly in structure. Hearing people with signing deaf parents are bilingual in sign and speech: languages conveyed in different modalities. Here we test the influence of modality and bilingualism on conceptual representation by comparing semantic representations elicited by spoken British English and British Sign Language in hearing early, sign-speech bilinguals. We show that representations of semantic categories are shared for sign and speech, but not for individual spoken words and signs. This provides evidence for partially shared representations for sign and speech, and shows that language acts as a subtle filter through which we understand and interact with the world.

Results

Hearing early, sign-speech bilinguals were presented with 9 conceptual items from 3

semantic categories: fruit, animals or transport, in a randomised event-related

functional Magnetic Resonance Imaging (fMRI) experiment. Each item was

presented as a sign (video) or as a spoken word (audio only, not audio-visual) and

was produced by a male or a female language model (Figure 1A). Participants were

highly accurate (mean = 97%) at performing a within scanner semantic monitoring

task (Figure 1B). Univariate GLM analyses indicated that speech and sign language

engaged similar fronto-temporal networks, consistent with previous studies [9–13]

(see Figure S2).

Shared semantic representations for speech and sign

Using a searchlight analysis, we first identified regions in which there were reliably

positive representational distances (see methods) between items ***within-modality***

(e.g. averaging speech-speech distances and sign-sign distances). We calculated

distances only between items from the different language models (e.g. different

speakers and signers respectively) to exclude similarities driven by low-level

perceptual properties. In these regions, we then tested for ***shared semantic***

representations using the following criteria: (A) a significant fit to the semantic

feature model in the ***within-modality*** distances (i.e. speech-speech across

speakers, and sign-sign across signers, see Figure 2B, red boxes) and (B) a

significant fit of the semantic feature model to the ***across-modality*** distances (i.e.

speech-sign and sign-speech, see Figure 2B, blue boxes). We also expected, (C)

no evidence of a difference in strength of fit to the semantic model between speech

3

74   and sign, (D) no evidence of low-level acoustic or visual sensitivity indicated by a fit

75   to a model predicting greater distances between items from a different, as compared

76   to the same speaker, in the speech-speech distances, or from a different, as

77   compared to the same signer, in the sign-sign distances and (E) no fit to a model

78   predicting sensitivity to the degree of iconicity of the signs, a perceptual feature

79   present in sign but not speech.

80   We found reliable within-modality distances in six clusters (Figure 2A): **(1)** in

81   bilateral V1-V3 and the LOC [-14 -96 10], **(2)** the right anterior superior temporal

82   gyrus [58 -4 -2], **(3)** the left anterior superior and middle temporal gyrus [-60 -10 -2],

83   **(4)** the right middle temporal gyrus and MT/V5 [52 -68 6], **(5)** the right insular [36 -12

84   14] and **(6)** the left posterior middle and inferior temporal gyrus (left pMTG/ITG) [-48 -

85   62 -6] (Figure 2, Table S2).

86   Three of these clusters showed a significant fit to the semantic model ***within-***

87   ***modality*** (after adjusting alpha to $p < 0.008$ for six tests/clusters). These were found

88   in the right middle temporal and V5/MT (cluster 4, $t(16) = 3.946$, $p = 5.78 \times 10^{-4}$, $d_z =$

89   0.957), the bilateral V1-V3 and LOC (cluster 1, $t(16) = 3.837$, $p = 7.28 \times 10^{-4}$, $d_z =$

90   0.931) and the left posterior middle and inferior temporal gyrus (left pMTG/ITG)

91   (cluster 6, $t(16) = 3.622$, $p = 0.001$, $d_z = 0.879$). However, the response in two of

92   these clusters was not consistent with shared semantic representations because the

93   fit to the semantic model was stronger for sign than for speech, after adjusting alpha

94   to $p < 0.017$ to account for 3 tests/clusters: right middle temporal and V5/MT cluster

95   ($t(16) = 2.842$, $p = 0.012$, $d_z = 0.689$) and the bilateral V1-V3/LOC cluster ($t(16) =$

96   4.630, $p = 2.78 \times 10^{-4}$, $d_z = 1.123$). In both areas, there was a significant fit to the

97   semantic feature model for sign (both $ps < 1.05 \times 10^{-4}$) but not speech (both $ps >$

4

98   0.110) and neither region showed a fit to the semantic model ***across-modality*** (both

99   ps > 0.046).

100   Only the response in the left posterior middle and inferior temporal gyri

101   (pMTG/ITG) was consistent with shared semantic representations (see Figure 2A,

102   cluster 6). In addition to (A) fitting the ***within-modality*** semantic feature model

103   (Figure 2D), the responses in this region showed (B**)** a significant fit to the ***across-***

104   ***modality*** semantic feature model (t (16) = 3.076, p = 0.004, d_z = 0.746, Figure 2D).

105   There was also (C) no evidence for differential sensitivity in the encoding of

106   semantics for speech and sign (t (16) = 0.400, p = 0.694, d_z = 0.097), (D) no

107   sensitivity to the acoustic or visual features associated with speaker (see model in

108   Figure 3E) or signer identity (see model in Fig. 4E), both ps > 0.060, and (E) no

109   influence of the iconicity structure of sign in the sign-sign or across-modality

110   distances, all ps > 0.106 (Figure S3).

111   The fit of the semantic feature model (Figure 1C) can be decomposed into

112   item-based dissimilarity (Figure 1D) and category-based dissimilarity (Figure 1E).

113   For ***within-modality*** distances, the left pMTG/ITG showed a significant fit to both the

114   semantic category (t (16) = 1.980, p = 0.033, d_z = 0.480) and item-based model (t

115   (16) = 4.185, p = 3.50 x 10^{-4}, d_z = 1.015). The critical analyses ***across-modality***,

116   indicated that the category-based model fit the data (t (16) = 2.509, p = 0.012, d_z =

117   0.608), but not the item-based model (t (16) = 0.475, p = 0.321, d_z = 0.115). There

118   was no evidence of a difference in strength of fit to the category model ***within-***

119   ***modality*** as compared to ***across-modality*** (t (16) = 0.135, p = 0.894, d_z = 0.033),

120   suggesting that semantic categories were represented robustly within- and across-

121   modality. By contrast, the item model was a better fit to the ***within-modality*** than

122   the ***across-modality*** distances (t (16) = 3.376, p = 0.004, d_z = 0.819, Figure 2F),

5

123   showing that item-based representations are less robustly encoded ***across-***

124   ***modality***.

125   Taken together, the results suggest that semantic category structure drives

126   similarity between sign and speech in left pMTG/ITG (see Figure 2C and 2E for the

127   MDS solution highlighting common category structure). As we did not observe the

128   same effects in anterior temporal lobe (ATL) regions that have previously been

129   associated with amodal semantic representations [14], we generated whole brain

130   tSNR maps to compare signal quality across regions. These indicated that tSNR

131   levels in the ATL were adequate and similar to the left pMTG/ITG (Figure S4).

132

133   **Modality specific representations**

134   In the absence of common category ***and*** item level representations, which would

135   have been supportive of fully shared semantic representations, we tested for

136   modality specific semantic representations. Using a searchlight approach, we

137   identified speech-specific and sign-specific regions by finding areas in which the

138   average of the speech-speech distances were greater than the sign-sign distances

139   and vice versa. In these regions, we tested for ***modality specific semantic***

140   ***representations,*** evidenced by a significant fit to (A) the full semantic feature model

141   (Figure 1C) and (B) to the semantic category model (Figure 1E) in the speech-

142   speech or sign-sign distances for speech or sign respectively, and (C) no evidence

143   of a fit to the speaker or signer identity model (see models in Figure 3E and Figure

144   4E) that would indicate a sensitivity to low level visual or auditory features.

145

146   ***Speech specific responses***

6

147   Four clusters showed greater representational distances for speech than sign: **(1)**

148   right anterior STG extending to the temporal pole [58 -4 -2], **(2)** left anterior STG [-56

149   -8 2], **(3)** right posterior STG/STS [58 -34 18] and **(4)** right putamen and insula [30 -

150   10 10] (Figure 3A, Table S2). None of the regions showed speech specific semantic

151   representations, as the category-based model (Figure 3D) was not a significant fit

152   (all ps > 0.110) after adjusting alpha to p < 0.013 to account for four clusters/tests. In

153   one of the clusters, the right anterior STG [58 -4 -2] (Figure 3A, cluster 1), there was

154   a significant fit to the semantic feature model (t (16) = 2.529, p = 0.011, d_z = 0.613,

155   Figure 3B and Figure 3H). However, this was driven by a fit to the item-level model

156   (t (16) = 5.229, p = 4.14 x 10^{-5}, d_z = 1.268, Figure 3C and Figure 3H) and was

157   accompanied by sensitivity to the acoustic differences between speakers (t (16) =

158   3.325, p = 0.002, d_z = 0.806, Figure 3E and Figure 3H). This pattern of response is

159   consistent with speech form representations rather than speech selective semantic

160   representations (Figure 3F and Figure 3G for MDS solution highlighting speaker-

161   based similarity). Identification of spoken word forms in the right anterior STG was

162   unexpected. This may reflect the greater involvement of the right hemisphere in

163   language processing in early bilinguals [15] or, given the reported greater importance

164   of the right hemisphere in sign processing in hearing native signers [16], may reflect

165   an effect more specific to early sign-speech bilinguals.

166

Sign specific responses

168   Five regions showed greater representational distances for sign than speech: **(1)** a

169   cluster spreading across left V1-V3 [-6 -98 16], **(2)** a cluster within right V1-V3 [22 -

170   90 16], **(3)** a cluster in the left LOC and MT/V5 [-44 -80 -6], **(4)** left superior occipital

171 gyrus and superior parietal lobule [-10 -84 42] and **(5)** left lingual gyrus spreading to

172 the cerebellum [-4 -48 -8] (Figure 4A, Table S2). Activity in these regions was not

173 consistent with sign-specific semantic representations, as the category-based model

174 was not a significant fit in any region (all ps > 0.037) after adjusting alpha to p <

175 0.010 for five clusters/tests. The response in the clusters in the left V1-V3 and right

176 V1-V3 were analogous to those for speech. Activity patterns were characterised by

177 a fit to the semantic feature model (both ps < 3.10×10^{-5}) but driven by item-based

178 encoding (ps < 1.34×10^{-7}) with additional sensitivity to signer identity (both ps <

179 3.07×10^{-6}, Figure 4), consistent with sign form representations.

180

181 **Discussion**

182 Our findings indicate that semantic representations for sign and speech are shared,

183 but only at a broad level of semantic specificity. In the left pMTG/ITG, both individual

184 items and categories were encoded within-modality, but across-modality, this was

185 true only for categories. Moreover, item-level encoding was significantly stronger

186 within- as compared to across-modality. In sign-specific and speech-specific

187 regions, we found item-based rather than category-based coding. These

188 representations retained sensitivity to auditory and visual features, suggestive of

189 phonological word and sign form representations rather than language specific

190 semantic representations.

191   Shared category representations for sign and speech in left pMTG/ITG is

192 consistent with studies showing common categories for items presented as pictures,

193 environmental sounds, and speech and text within this region [1,2]. Indeed,

194 activation of the left pMTG/ITG is associated with extraction of meaning from both

195 sound and vision. It is activated when reading words [17], perceiving semantically

196 ambiguous speech [18] and sign language [19–21]. However, the loci of shared

197 representation is more posterior than the more anterior temporal lobe regions

198 associated with amodal semantics predicted by the "hub and spokes" model of

199 semantic cognition [14]. Plausibly, the more posterior convergence identified in our

200 study may be influenced by visually derived language representations of sign that

201 may be found closer to the primary visual cortices. In contrast, amodal processing in

202 ATL has been observed in studies of spoken language, either in healthy individuals

203 or those with semantic dementia. Users of only spoken languages do not have

204 visually derived language representations in the same way that signers do. We learn

205 to read alphabetic scripts by making strong associations between orthography and

206 speech sounds [4]. Similarly, pictures likely activate dual visual-verbal processing

207 routes in spoken language users [6]. Our work highlights the unique contribution that

208 sign languages provide in understanding semantic cognition. Future studies, with

209 healthy sign language users, deaf and hearing, and those with semantic dementia

210 will contribute towards more complete models of semantic processing.

211   Common semantic coding was limited to category and not item level

212 representations. This subtle divergence between languages is consistent with the

213 notion that language influences, rather than determines, perception and thought

214 [22,23]. These data make a novel contribution, since we compared neural

215 responses to languages that differ substantially in their linguistic structure, using

216 sensitive multivariate statistical methods. However, we do not claim that our findings

217 are necessarily specific to the contrast between signed and spoken languages. Our

218 results are consistent with previous work that failed to show cross-decoding between

219 individual spoken and written words across languages in English-French bilinguals

9

220 [24], although that study did not test for category coding. Further work should

221 investigate whether similar mechanisms underlie both findings. Studies testing for

222 item and category-based similarity for text, speech and sign in sign-speech

223 bilinguals, and between stimuli in different modalities in spoken language bilinguals

224 using typologically close and distant spoken languages, will clarify the specificity of

225 our findings. Contrasts of representations of signs in deaf signers and speech in

226 hearing monolinguals will further clarify the influence of language experience on

227 such representations.

228   Why are conceptual representations shared at only a coarse level of semantic

229 specificity? Partially shared semantic representations between languages is

230 consistent with computational models of bilingualism, such as the Distributed

231 Feature Model [25]. These models predict a single semantic store, in which each

232 language weights semantic features independently [25–27]. One factor contributing

233 to differing weights between sign and speech may be the greater polysemy (lexical

234 items having more than one meaning) exhibited in signed languages [28]. Another

235 may be a consequence of differences in phonology. Studies of spoken language

236 show that lexical-semantic access is affected by the phonological structure of the

237 lexicon. Words from dense phonological neighbourhoods activate semantic

238 representations less strongly [29] due to cascading activation between phonology

239 and semantics [30]. Signed and spoken languages have very different phonologies

240 and therefore phonological neighbourhoods. This might affect the strength and

241 structure of semantic activation within sign and speech lexicons, reducing the

242 commonality of conceptual representations between the languages.

243   Another explanation is that the greater iconicity found in sign languages [31]

244 reduces the degree of similarity between sign and speech.   Although, we did not

10

245   observe an effect of iconicity in the response in the left pMTG/ITG, which would have

246   directly supported this explanation, there are also more opaque form-meaning links

247   that differ across speech and sign. For example, the handshape "I" (extension of the

248   little finger alone) denotes a number of BSL signs that have negative connotations:

249   bad, wrong, poison [32] . Similarly, English words beginning with "gl" are often

250   associated with light of low intensity: glow, glint, glimmer [31]. Some canonical signs

251   also carry additional layers of meaning that communicate size, location, movement

252   and other features of the referent; aspects of meaning that cannot be communicated

253   by the voice. These features may fundamentally differentiate semantic

254   representations for sign and speech. Given this, we might predict differences in the

255   representation of specific semantic categories. For example, representations for

256   tools might be expected to differ between unimodal (e.g. speech-speech) and

257   bimodal (e.g. sign-speech) bilinguals on the basis that signs for objects would evoke

258   greater specificity in the semantic features associated with how they are handled,

259   particularly in sign languages that emphasise the handling properties of objects [33].

260   To conclude, our results suggest that the language that we use to

261   communicate acts as a subtle filter through which we understand and interact with

262   the world. This finding is unexpected. Previous brain imaging studies showing

263   significant univariate overlap of activation for sign and speech [9–13] has led

264   researchers, including ourselves, to propose extensive similarity in the neural

265   processes underlying sign and speech [34]. Our findings suggest the need to rethink

266   this assumption and highlight the unique perspective that sign language can provide

267   on language processing and semantic representation more broadly.

268

11

269 **ACKNOWLEDGEMENTS**

277

278 **AUTHOR CONTRIBUTIONS**

279 S.E., M.M., J.D., C.P. & E.G. designed the study. S.E. collected the data. S.E., J.D.,

280 M.M. analysed the data. All authors contributed to writing the article.

281

282 **DECLARATION OF INTERESTS**
283 The authors declare no competing financial interests

284

285

286

287

288

289

290

291   **MAIN TEXT FIGURE/TABLE LEGENDS**

292

293   **Figure 1: Stimuli, experimental design and semantic models.**

294   (A) Hearing, early sign-speech bilinguals were presented with 9 conceptual items
295   that belonged to 3 semantic categories: fruit, animals and transport. Items were
296   presented as signs (videos) and spoken words (auditory presentation only) and were
297   produced by male and female language models.

298   (B) Within the scanner, participants attended to speech and sign and pressed a
299   button to identify items that were not in one of the three target categories (e.g.,
300   umbrella).

301   (C-E) The dissimilarity between neural patterns evoked by the signs and spoken
302   words were tau-a correlated with different theoretical models. The color bar reflects
303   the degree of predicted semantic dissimilarity between items. (C) A semantic feature
304   model derived from the CSLB concept property norms [35]. This model was
305   decomposed into two independent components: (D) An item-based model that
306   predicts that each item is uniquely represented, e.g., an 'apple' is more dissimilar to
307   other items than to itself and does not predict any broader semantic relatedness
308   between items and (E) a category-based model in which the between-item
309   similarities are predicted by the semantic feature model, but where the within-item
310   similarities are not tested. White squares in this model indicate comparisons that
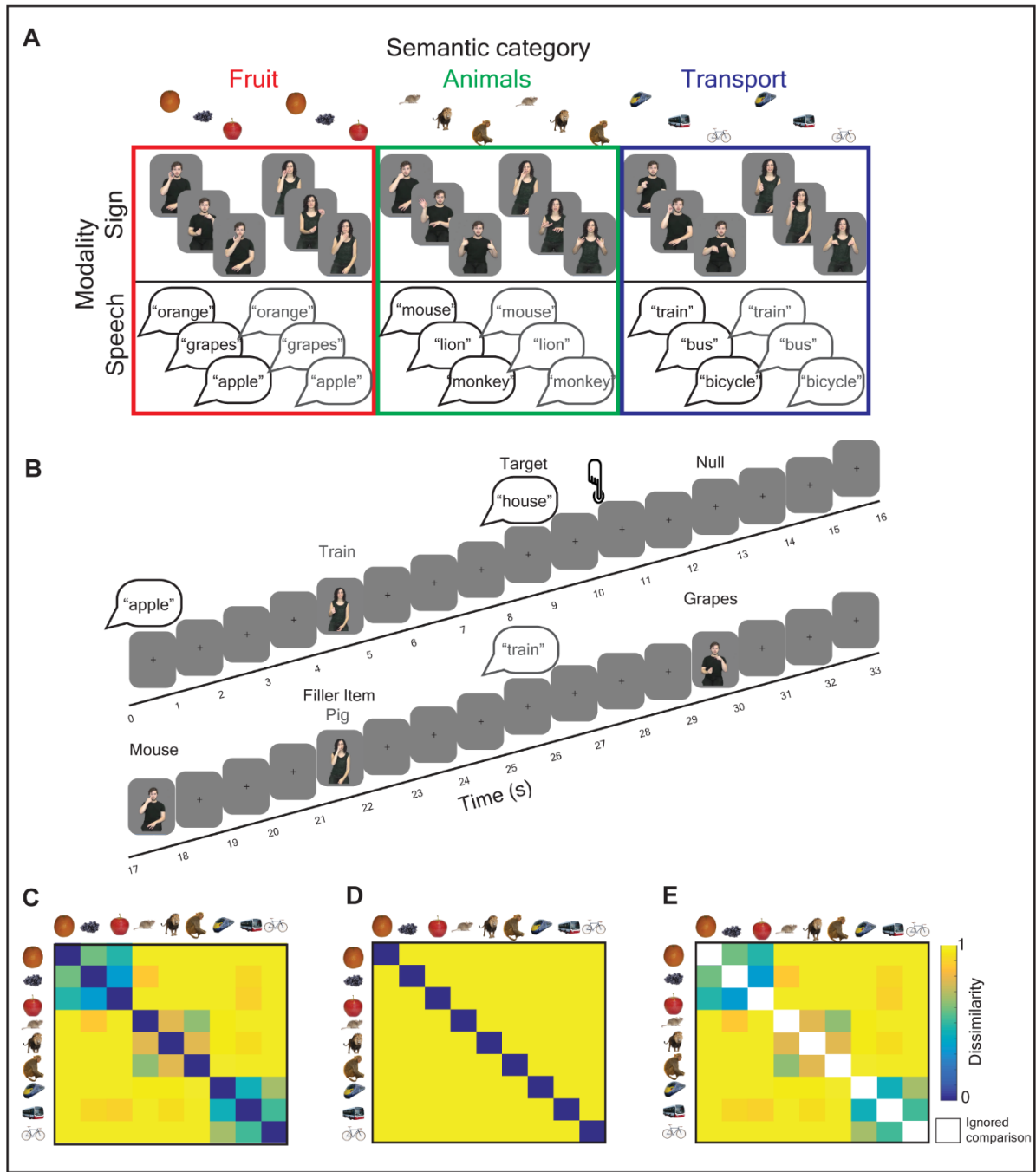311   were excluded.

13

Figure 2: Shared semantic representations for speech and sign.

(A) A searchlight analysis identified brain regions containing positive *within-modality* representational distances, thresholded at p < 0.005 peak level, FDR corrected at q < 0.05 at the cluster level (Extent threshold, k = 172 voxels). Clusters are numbered according to the text in the results section. Table S2 details the local maxima from this analysis. See Figure S2 for the univariate overlap between sign and speech and see Figure S4 for tSNR maps showing how signal quality varied across the brain.

(B) Representational distances in these regions were Tau-a correlated with the semantic feature model within-modality and across-modality. The red boxes illustrate the within-modality distances, with the upper red box testing for abstracted speech

14

323　representations (e.g. from speaker 1 to 2), and the lower red box testing for
324　abstracted representations for sign (e.g. from signer 1 to 2). The blue box contains
325　all across-language distances. Each 9x9 submatrix of dissimilarities is predicted
326　from the semantic feature model (Figure 1C). White boxes are comparisons
327　excluded from the analysis. The color bar reflects the predicted strength of
328　dissimilarity.

329　(C-F) Plots show the response in cluster 6, the left pMTG/ITG [-48 -62 -6]. In this
330　region, there was a fit to the semantic feature model within-modality and across-
331　modality. However, when item-based and category-based representations were
332　differentiated, this showed that the semantic category model (Figure 1E) was a fit
333　within-modality and across-modality, but the item-based model (Figure 1D) was only
334　a fit within-modality. Further, the item based model was a better fit within-modality
335　than across-modality. (C) Shows the non-metric MDS representation of the response
336　in this region: the left panel shows within sign distances magnified to make the
337　representational structure clearer and the right panel shows the equivalent speech
338　representations. In these magnified images, lines connect the same conceptual item
339　produced by each speaker or signer, marked as speaker/signer 1 or speaker/signer
340　2 on the figure. (D) Plot shows the significant fit to the semantic feature model both
341　within-modality and across-modality. Violin plots show distributions and individual
342　data points for the z transformed values, including the 90% confidence interval and
343　the noise ceiling (grey rectangle). (E) The non-metric MDS representation showing
344　the mean centroid of each category within each modality for fruit (red), animals
345　(green), blue (transport), with dashed line connecting centroids across-modality.
346　Note the similar ordering of the category centroids both within and across each
347　modality. (F) Plot shows the difference in fit to the item model within-modality and
348　across-modality. See Figure S3 for the influence of sign iconicity on the left
349　pMTG/ITG and Figure S1A for the definition of leave-one-out ROIs for testing
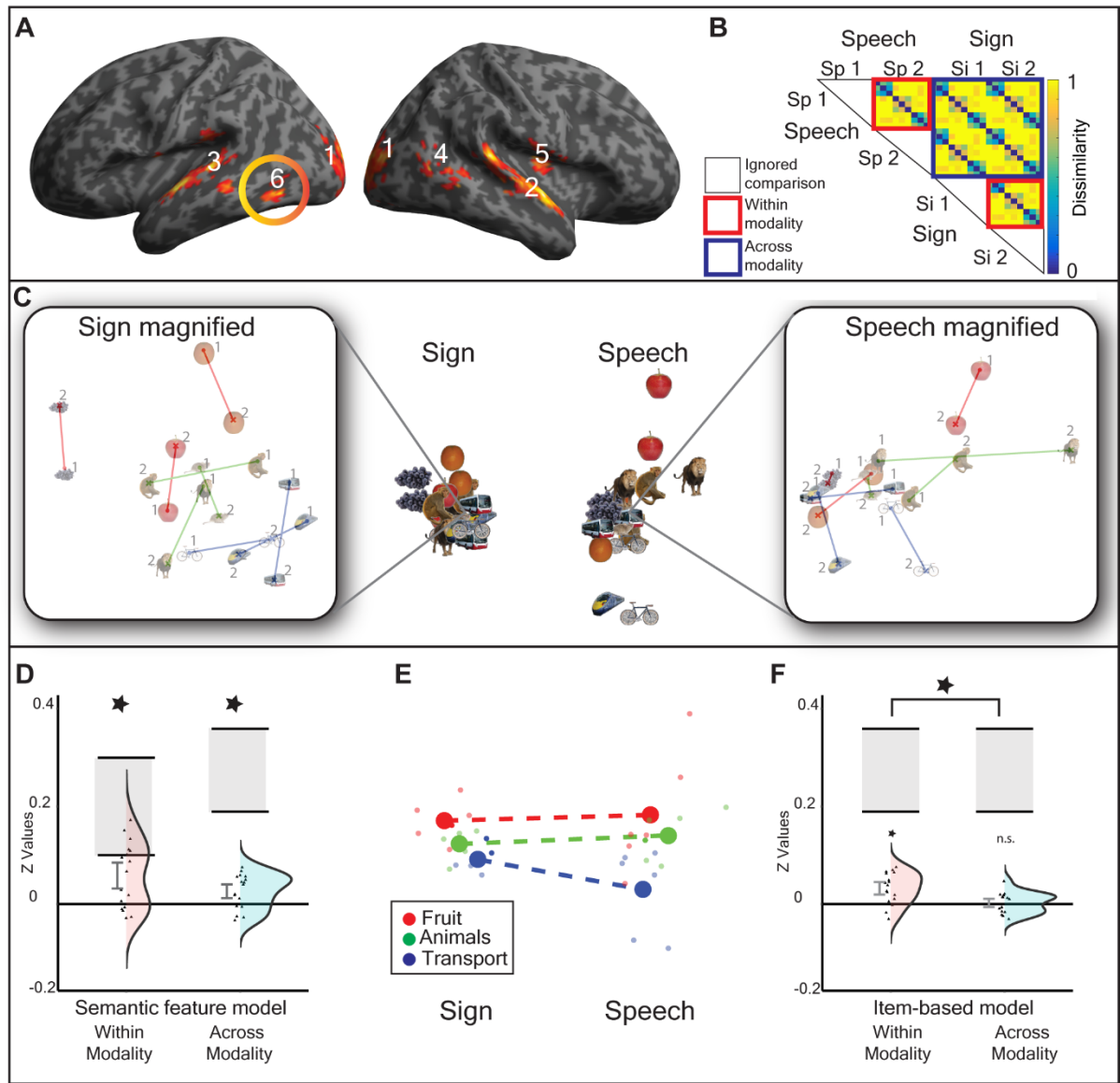350　sensitivity to speaker and signer identity in this region.

15

351

Figure 3: Speech-specific neural responses.

(A) A searchlight analysis identified regions with greater representational distances for speech compared to sign, thresholded at $p < 0.005$ peak level, FDR corrected at $q < 0.05$ at the cluster level (Extent threshold, $k = 146$ voxels). Clusters are numbered according to the text in the results section. Table S2 details the local maxima from this analysis.

(B-E) Show the within-speech models that were tested: (B) Within-speech semantic feature model, (C) Within-speech item-based model, (D) Within-speech category-based model and (E) Between-speaker model. All models test dissimilarities across speaker (e.g. from speaker 1 to 2) in order to identify representations abstracted from perceptual features. Color bar reflects predicted strength of dissimilarity. White boxes are comparisons excluded from analysis.

(F-H) Show the response in cluster 1, the right anterior STG [58 -4 -2]. In this region, there was a significant fit to the semantic feature model, driven by item-based rather than category-based similarity and additional sensitivity to speaker identity. This is

16

367 consistent with abstract spoken word form representations rather than modality
368 specific semantic processing. (F) Shows the non-metric MDS solution. (G) Illustrates
369 speaker identity encoding in leave-one-participant-out ROIs (see Figure S1B). Large
370 circles represent the centroids for items from speaker 1 (red) and speaker 2 (blue).
371 Smaller circles represent the observed response for each item. Grey lines connect
372 each item to centroid. (H) Violin plots show model fits for z transformed values for
373 each model, with distributions and individual data points and 90% confidence
374 intervals and noise ceiling (grey box shown).

375

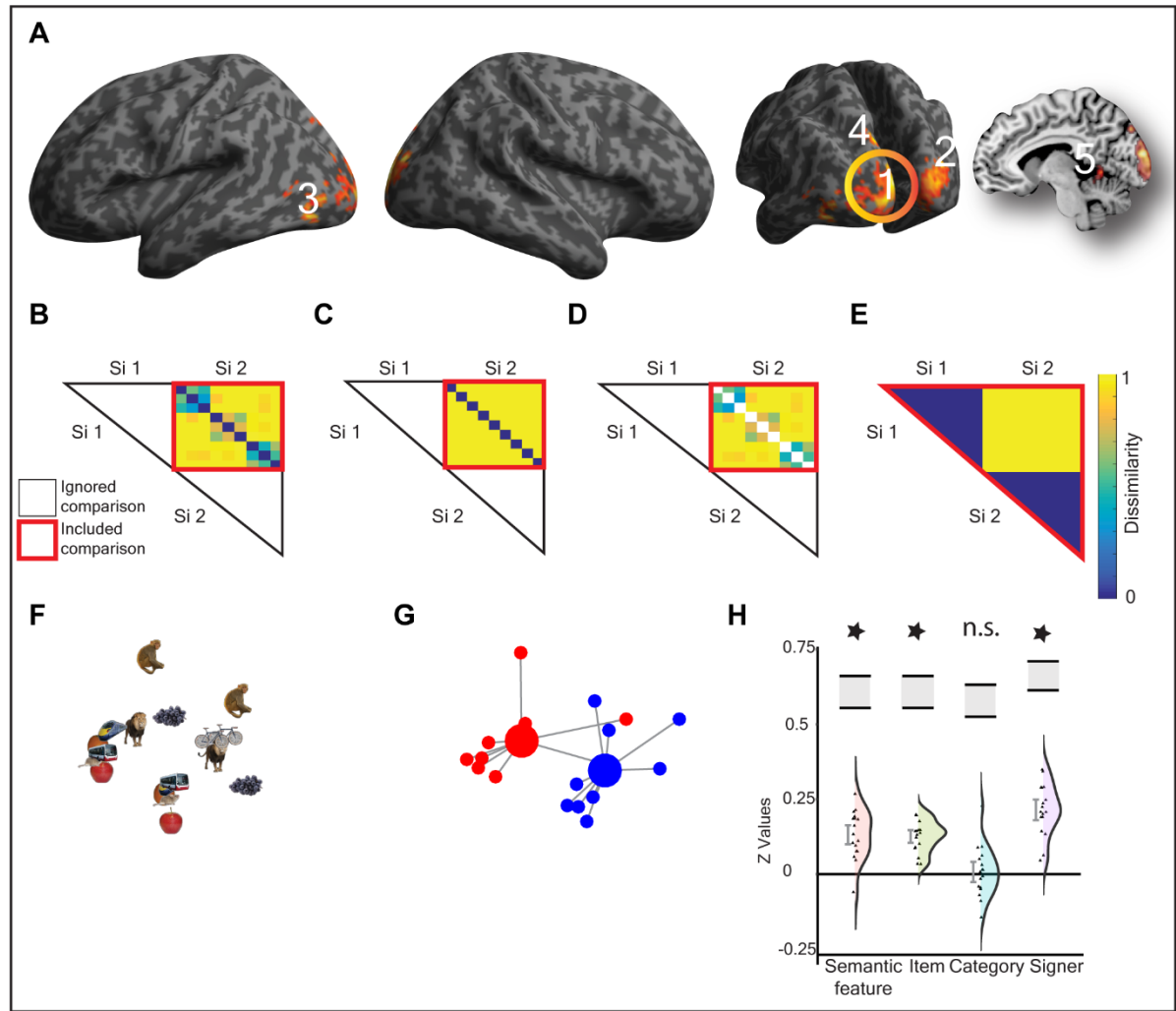**Figure 4: Sign-specific neural responses.**

377 (A) A searchlight analysis identified regions with greater representational distances
378 for sign compared to speech, thresholded at p < 0.005 peak level, FDR corrected at
379 q < 0.05 at the cluster level (Extent threshold, k = 116 voxels). Clusters are
380 numbered according to the text in the results section. Table S2 details the local
381 maxima from this analysis.

382 (B-E) show the within-sign models that were tested: (B) Within-sign semantic feature
383 model, (C) Within-sign item-based model, (D) Within-sign category-based model and
384 (E) Between-signer model. All models test dissimilarities across signer (e.g. from
385 signer 1 to 2) in order to identify representations abstracted from perceptual features.

386 Color bar reflects predicted strength of dissimilarity. White boxes are comparisons
387 excluded from analysis.

388 (F-H) shows response in cluster 1, the left V1-V3 [-6 -98 16]. In this region, there
389 was a significant fit to the semantic feature model, driven by item-based rather than
390 category-based similarity structure and an additional sensitivity to signer identity,
391 consistent with abstract sign form representations rather than modality specific
392 semantic processing. (F) Shows the non-metric MDS solution. (G) Illustrates signer
393 identity encoding in leave-one-participant-out ROIs (see Figure S1C). Large circles
394 represent the centroids for items from signer 1 (red) and signer 2 (blue). Smaller
395 circles represent the observed response for each item. Grey lines connect each item
396 to centroid. (H) Violin plots show model fits for z transformed values for each model
397 fit, with distributions and individual data points and 90% confidence intervals and
398 noise ceiling (grey box shown).

399

400

401

402

403

404

405

STAR METHODS

407

LEAD CONTACT AND MATERIALS AVAILABILITY

409 Further information and requests for resources and reagents should be directed to

410 and will be fulfilled by the Lead Contact, Dr Samuel Evans

411 (S.Evans1@westminster.ac.uk). All materials are available upon request.

412

EXPERIMENTAL METHODS AND SUBJECT DETAILS

Participants

415 Ethical approval was granted by the UCL ethics committee and informed consent

416 was obtained from all participants. Data were collected from 18 right handed early

417 sign-speech bilinguals with no known neurological, hearing or language learning

418 impairments. One participant's data was removed from the set due to an incidental

419 finding, leaving a final data set of 17 participants (Mean age=33; range 20-52 years;

420 female=12). All of the participants were born and educated in the UK, except for one

421 who was born in Australia and another who was born in a non-English speaking

422 country, but moved to the UK at the age of three. Fifteen participants learned British

423 Sign Language (BSL) from a deaf parent and two from an older deaf sibling. Two of

424 the participants who learned sign language from a deaf parent did not learn BSL

425 from birth; one, learned AUSLAN from birth and learned BSL from the age of twenty-

426 one, the other, was exposed to another sign language from birth, before learning

427 BSL from 3 years of age. Participants judged themselves to have excellent BSL

428 skills on a self-report scale (1 poor - 7 excellent): mean = 6.3/7, SD= 0.86, range = 4-

429 7. Six participants had previously worked as a BSL interpreter or were currently

430 training to be an interpreter. One was a BSL teacher and three had worked or were

431 working as Communication Support Workers (CSWs). All participants reported

432 having previously interpreted in an informal capacity for a family member.

433

434 **METHOD DETAILS**

435 ***Speech and sign stimuli***

436 Stimuli consisted of nine core items for which neural responses were analysed.

437 Each core item was presented 48 times across the whole experiment, in different

438 modalities (sign/ speech) and by different models (male/ female) (see 'fMRI

439 paradigm' for more details). These nine items belonged to three categories: fruit

440 (orange, grapes and apple), animals (mouse, lion and monkey) and transport (train,

441 bus and bicycle). Items within each category were similar and were distinct from

442 other categories on the basis of their semantic features, as evidenced by the CSLB

443 concept property norms [35] (see Figure 1C). Items were chosen to ensure that the

444 categories were matched for age of acquisition (fruit M = 3.78; animals M = 4.52;

445 transport = 4.04), imageability (fruit M = 618; animals M = 610; transport M = 622),

446 familiarity (fruit M = 566; animals M = 521; transport M = 551) and the number of

447 syllables and phonemes in spoken English [36–38] (see Table S1 for full details). In

448 addition, we ensured that the BSL equivalents of the spoken words were matched

449 across category for handshape, location, movement and handedness, and that

450 iconicity [39] was similar across categories (fruit M = 3.80; animals M = 3.92;

451 transport M = 4.23; 1 low - 7 high iconicity). Iconicity ratings from the participants'

452 were significantly correlated with those collected from deaf BSL users by Vinson et.

453 al. [39] (n=18, r= 0.917, p = 2.22 x 10^{-07}).

454   Speech samples were recorded by a male and female Southern British

455 English (SBE) speaker in an acoustically shielded booth with 16-bit quantisation and

456 a sampling rate of 22050 Hz using Adobe Audition. These were auditory only, rather

457 than the auditory-visual presentations typically used in studies comparing speech

458 and sign language processing [19]. Auditory only speech presentations ensured that

459 speech and sign were maximally different from each other and that any observed

460 commonalities could not be attributed to common visual features. Auditory

461 recordings were excised at the zero crossing point. They were then filtered to

462 account for the frequency response of the Sensimetric headphones used in the

463 scanner (http://www.sens.com/products/model-s14/) and the overall amplitude was

464 Root Mean Square (RMS) equalised to ensure a similar perceived loudness. The

465 mean duration of the auditory stimuli for the core items was 558ms (range = 323-865

466 ms), these sounds were similar in duration across semantic categories (fruit M = 573

467 ms; animals M = 575 ms; transport M = 533 ms) and gender of the speaker (male M

468 = 557 ms; female M = 564 ms). The phonological distance between each of the

469 spoken words was calculated using the Levenshtein distance [40]. This was

470 achieved by calculating the number of phoneme insertions, deletions and/or

471 substitutions necessary to turn one word into the other, divided by the number of

472 phonemes in the longest word. The absolute value of the difference in Levenshtein

473 distance between each item was calculated. These distances did not correlate with

474 the semantic feature distances (r = 0.063, n = 36, p = 0.713), hence semantic

475 structure was not confounded with phonemic structure.

21

476   The BSL signs were all common variants in southern England as shown in the

477 BSL SignBank [41] (http://bslsignbank.ucl.ac.uk/dictionary/). Signs were recorded

478 with a Sony Handycam HDR-CX130 on a blue background by a male and a female

479 deaf native signer with a sampling rate of 50 fps and an aspect ratio of 1920x1080.

480 The blue background was keyed out and replaced with a dark grey background.

481 Videos were down-sampled to 30 frames per second and a resolution of 960 x 540

482 with Adobe Premiere for presentation in the scanner. All signs were produced with

483 corresponding BSL mouthing. The signs were recorded in isolation such that the

484 hands returned to a neutral position resting on the knees between each sign. During

485 editing, the start and end-points of a sign were identified as a 'hold' (very brief pause

486 in movement of the hands) to remove the transitional movement into and out of the

487 neutral hands on the lap. Still frames of the hold points at the beginning and end of

488 each sign, with duration of 333ms, were inserted to ensure that the signs were easily

489 perceived in the scanner. The mean duration of the sign stimuli was 1107ms (range

490 = 867-1400ms). The signs were similar in duration as a function of semantic

491 category (fruit M = 1079ms; animals M = 1055ms; transport M = 1128ms) and

492 gender of the signer (male M = 1087ms; female M = 1086ms).

493   An iconicity dissimilarity measure [39] for the signs was calculated by taking

494 the absolute value of the difference between ratings of each item with every other.

495 These distances did not correlate with semantic feature similarity (r = -0.126, n =36,

496 p=0.465), hence semantic structure was not confounded with iconicity.

497   Participants were shown 36 additional items in the scanner to facilitate a

498 semantic monitoring task (see Figure 1B) for which neural activity was not analysed.

499 The additional items consisted of 18 items from outside the categories of fruit, animal

500 and transport, e.g. buildings, clothes, furniture and tools, which were included as

22

501 target filler trials. Plus, an additional 18 non-target filler trials, 6 per category, of

502 other types of fruit, animals or transport that were included to reduce habituation to

503 the nine core items (see 'fMRI Paradigm' below for details of number of

504 presentations). Each individual filler item was produced by only one of the speakers

505 or signers, with the number of items from each speaker and signer balanced.

506   Prior to scanning, participants were familiarised with the signs and spoken

507 words used in the study. Participants saw each sign stimulus produced by both sign

508 models and were required to translate the word into spoken English. They also heard

509 each word produced by both speech models and were required to repeat the spoken

510 word aloud. They were shown all core items, target and non-target fillers. Sign

511 recognition was high (core items: mean = 17/18, min = 15/18, max = 18/18; filler

512 items: mean = 32/36, min = 21/36, max = 35/36). On very few occasions participants

513 interpreted a sign as a non-intended English word. Typically when this occurred

514 participants provided a translation that reflected their regional variant of BSL. When

515 participants were asked if they knew any other meanings of the sign, they were

516 usually able to provide the target translation. They were then asked to interpret the

517 sign, on this occasion, as the target translation for the study. They were then

518 retested on all the items in the experiment to ensure retention. Seventeen out of 18

519 participants required one round of correction, the remaining participant required a

520 second round. Participants practiced a mock version of the within scanner task on a

521 laptop prior to scanning.

522 ***fMRI task***

523 In the scanner, participants were required to attend to the signed and spoken stimuli

524 and to press a button when they encountered an item from outside the categories of

23

525 fruit, animals or transport, e.g. a target filler item (see Figure 1B). The handedness

526 of the button press was counterbalanced across participants. On average 97% of

527 outside category target items were identified (mean 35/36 correct, SD = 1.45, min =

528 31, max = 36) and accuracy was significantly greater than chance (mean d' score =

529 4.56), t (16) = 42.74, p = 6.37 x 10^{-18}, indicating that participants were fully engaged

530 with the task.

531   Data were collected in 6 runs. In each run, each of the 9 core items were

532 presented twice in each of the following formats: sign and speech; male and female

533 model. Therefore, each core item was presented 8 times in each run (2x2x2), with

534 72 core trials in total (9 items x 8 instances). Within each run, core items were

535 presented as two concatenated mini blocks of 36 trials. Within each mini block items

536 were randomised with the constraint that the same concept (e.g., 'orange') could not

537 be presented consecutively, regardless of modality, to reduce habituation.

538   In addition, in each run there were 6 target filler trials (non-fruits, transport or

539 animals) for which participants were required to press a button and 6 non-target

540 fillers ('other' fruits, transport or animal items). The total number of trials was

541 balanced within run for modality (e.g. whether sign or speech) and language model

542 (e.g. speaker and signer). The filler trials (target and non-target fillers) were

543 interspersed within each run regularly but unpredictably. An additional, seven null

544 trials lasting 4 seconds were regularly but unpredictably interspersed within the each

545 run. During these trials a white fixation cross was presented on a grey background

546 in the absence of sound or additional visual stimulation for 4 seconds.

547   In summary, each of 6 runs consisted of 91 trials (72 core trials, 6 target filler

548 trials, 6 non-target filler trials, 7 null trials). The order of modality of presentation of

24

549 the items (speech/sign) was counter balanced across pairs of participants, such that

550 items presented as signs to participant 1 were presented as speech to participant 2,

551 and vice versa. Each stimulus was presented for its natural duration and was

552 followed by a fixation cross lasting 3 seconds, before the start of the next trial.

553   After scanning, participants provided iconicity ratings on the sign stimuli that

554 they had viewed in the scanner using the technique described by Vinson et al. [39].

555 They then took part in a multiple arrangement task in which they arranged pictures of

556 the core and filler items "based on their similarity" using a drag and drop interface

557 [42]. The Euclidean distances derived from this arrangement correlated highly with

558 the CSLB concept property norms for the core items ($r = 0.904$, $n = 36$, $p = 4.42 \times 10^{-14}$

559 $^{-14}$), suggesting that the semantic feature norms provided a good summary of the

560 semantic space of our participant group.

561 *MRI Data Acquisition*

562 Data was acquired with a 3-Tesla scanner using a Magnetom TIM Trio systems

563 (Siemens Healthcare, Erlangen, Germany) with a 32 channel headcoil. A 2D epi

564 sequence was used comprising forty 3mm thick slices using a continuous ascending

565 sequence (TR=2800ms, TA=2800ms, FA= 90°, TE=30ms, matrix size= 64x64, in-

566 plane resolution: 3mm x 3mm, interslice gap = 1mm). Six runs of data were acquired

567 each lasting ~6-7 minutes with around 136 brain volumes collected per run; the exact

568 number of volumes was dependent on the stimuli included in each run. EPI data

569 collection lasted around 45 minutes. This was followed by a fieldmap, acquired

570 using a double-echo FLASH gradient echo sixty-four slice sequence (TE1=10ms,

571 TE2=12.46ms, in-plane view 192x192 mm, in-plane resolution: 3mm x 3mm,

572 interslice gap = 1mm). At the end of the session a high-resolution T1 weighted

25

573   structural image was collected using a 3D Modified Driven Equilibrium Fourier

574   Transform (MDEFT) sequence (TR=1393ms, TE=2.48ms, FA= 16°, 176 slices, voxel

575   size = 1 × 1 × 1 mm).

576   In the scanner, stimuli were presented using the COGENT toolbox

577   (http://www.vislab.ucl.ac.uk/cogent.php) running in MATLAB. Auditory stimuli were

578   presented at the same comfortable listening level for all participants. Visual images

579   were presented using a JVC DLA-SX21 projector, with a screen resolution of

580   1024x768 and frame rate of 60Hz, using back projection onto a within bore screen at

581   a distance of 62cm from the participants' eyes.

582

583   **QUANTIFICATION AND STATISTICAL ANALYSIS**

584   ***Univariate Analysis***

585   Data were analysed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/) using MATLAB.

586   The first six images of each run were removed to account for T1 equilibrium effects.

587   The structural and functional images were centred at the anterior commissure.

588   Functional scans were slice time corrected to the middle slice, realigned to the first

589   image and unwarped using field maps. The structural image was co-registered to

590   the mean functional image. The parameters derived from segmentation, using the

591   revised SPM12 segmentation routines, were applied to normalise the functional

592   images that were re-sampled to 2x2x2mm. The normalized images were then

593   smoothed with a Gaussian kernel of 6-mm full-width half maximum. Data were

594   analyzed using a general linear model with a 360 second high-pass filter and AR1

595   correction for auto-correlation. In the first level design matrices, events were

596   modelled with a canonical hemodynamic response function marking the onset of the

26

597   stimulus and duration in seconds. The design matrices included a regressor for the

598   onset of the speech trials, sign trials, filler target and non-target trials in each

599   modality (4 regressors), button presses when the target was present in each

600   modality (e.g. hits) (2 regressors) and button presses when the target trials were

601   absent for each modality (e.g. false alarms) (2 regressors), six movement regressors

602   of no interest and the session means.   The rest condition constituted an implicit

603   baseline. Contrast images of [speech > rest] and [sign > rest] were taken to the

604   second level to conduct one sample t-tests.

605   ***Representational similarity analysis (RSA)***

606   At the first level, data were analysed with SPM12. Analyses were conducted in

607   native space. Images were slice time corrected to the middle slice, realigned to the

608   first image and unwarped using fieldmaps, but were not normalised or smoothed.

609   The images were segmented, using the revised SPM12 segmentation routine, to

610   estimate the transformation from native space to MNI space and vice versa. In the

611   first level model in native space, the two repetitions of each core item presented in

612   each modality and by each speaker and signer were modelled as a separate

613   regressor (36 regressors: 9 core items x 2 modalities x 2 language models).

614   Additional regressors were included modelling the onset of filler target and filler non-

615   target trials for each modality (4 regressors), plus button presses when the target

616   was present in each modality (e.g. hits) (2 regressors) and button presses when the

617   target trials were absent for each modality (e.g. false alarms) (2 regressors). This

618   constituted 42 regressors per run, plus 6 motion parameter regressors and 6 session

619   means. A high pass filter set at 360 seconds and AR(1) correction was applied.

620   RSA analysis was conducted with the latest version of the RSA toolbox

621   (https://github.com/rsagroup/rsatoolbox) [43]. The representational distances

27

622 estimated from the first level betas were used to calculate the cross-validated

623 Mahalanobis (crossnobis) distances using the RSA toolbox [43]. These crossnobis

624 distances employ multivariate noise normalisation that down-weight correlated noise

625 across voxels, thereby increasing sensitivity to experimental effects [44]. The cross-

626 validation across imaging runs ensures that the estimated distances between neural

627 patterns are not systematically biased by run-specific noise, which allows us to test

628 the distances directly against zero (as one would test cross-validated classification

629 accuracy against chance). Therefore, the crossnobis distance provides a

630 measurement on a ratio scale with an interpretable zero value that reflects an

631 absence of distance between items.

Searchlight RSA analyses

633 A volumetric searchlight analysis [45] was conducted using a spherical 8mm

634 searchlight containing 65 voxels, consistent with the parameters used in previous

635 studies of language processing [46]. In the searchlight analysis, the crossnobis

636 distance between each core stimulus and every other was calculated to generate a

637 Representational Dissimilarity Matrix (RDM) for every voxel and its surrounding

638 neighbourhood. The resulting RDM reflected sign-sign, speech-speech or speech-

639 sign distances, that constitute within and across-modality dissimilarities. In the

640 searchlight analyses, the average of speech-speech and sign-sign distances (e.g.

641 combined within-modality distances) and the average of the speech-speech and

642 sign-sign distances separately were returned to the voxel at the centre of each

643 sphere in three separate searchlight analyses. Within-modality distances were

644 calculated only between items from the different language models (e.g. different

645 speakers and signers respectively) to exclude similarities driven by low-level

646 perceptual properties. Each participants' native space whole brain searchlight map

647 was normalised to MNI space. These maps were inclusively masked with a >20%

648 probability grey matter mask, using the canonical MNI brain packaged with SPM12.

649 The resulting normalised, masked images were submitted to SPM12 for one sample

650 t-tests testing for greater than zero within-modality distances and paired t-tests

651 testing for differences between the speech-speech and sign-sign distances at the

652 second level. All statistical maps are presented at an uncorrected peak level

653 threshold of p < 0.005, FDR cluster corrected at q < 0.05 to identify regions of

654 interest for subsequent analysis. Extent thresholds were as follows: within-modality

655 distances (k = 172 voxels), speech > sign distances (k = 146 voxels) and sign >

656 speech distances (k = 116 voxels).

657 ***Regions of Interest (ROI) Analyses***

658 The clusters identified from the searchlight analyses were used as Regions of

659 Interest (ROIs) in which to test theoretical models of brain function. Note that ROI

660 analyses are advised when testing special populations in which sample sizes are

661 necessarily restricted [47]. Using ROIs that contain reliable representational

662 structure, e.g. greater than zero distances, provides an additional protection against

663 spurious distance-model correlations in regions in which there is no reliable

664 representational structure. This approach is an efficient and statistically powerful

665 way to generate ROIs as it uses all the data [48].

666 As each cluster contains multiple RDMs, one for each searchlight contained

667 within the cluster, the RDMs were averaged, to provide a single representative RDM

668 for each cluster, and each participant. These distances were then used to test

669 hypothetical models of brain function (described below). The non-parametric Tau-a

670 correlation was used in preference to Pearson or Spearman correlation as the

29

671 models contained tied ranks [43]. The resulting correlation coefficient was converted

672 to a Pearson's r value, then to a Fisher-transformed Z value, to permit parametric

673 statistical analysis [49]. Noise ceilings [43] were estimated within-modality and

674 across-modality separately as appropriate for each model. The lower bound was

675 estimated by calculating the mean z converted Tau-a correlation coefficient between

676 each participant's RDM and the average RDM for the group excluding that

677 participant (e.g. leaving one participant out). This is an estimate of the fit that should

678 be achieved if the theoretical model captures all systematic variation in the RDM

679 across subjects in this region. The upper bound was estimated by calculating the

680 mean z converted, Tau-a correlation between each participant's RDM and the

681 average RDM for the group including that participant. This value constitutes a

682 theoretical maximum of the best possible fit that can be achieved between the data

683 and a model with this region. These limits provide a benchmark against which to

684 assess the quality of model fit as they reflect the bounds of the best possible model

685 fit that could be expected given the noise in the data.

686 ***RSA Models***

687 A semantic model was tested using the CSLB concept property norms [35] (Figure

688 1C). This kind of feature-based semantic model can account for the ability to

689 categorize by semantic group, e.g. a zebra is an animal, and to tell-apart unique

690 items, e.g. that a zebra differs from a horse. As such, the similarities expressed by

691 the model can be decomposed into two independent components. One, an **item-**

692 **based** model that predicts that each item is uniquely represented, e.g., an 'orange' is

693 more dissimilar to all other items than to itself, and does not predict any other

694 relatedness between items (Figure 1D). The other, a model in which item-to-item

695 similarities are not tested, but category structure is predicted (Figure 1E) – referred

696   to as a **category-based** model. An additional model testing for dissimilarities based

697   on speaker (Figure 3E) and signer identity (Figure 4E) was also tested, e.g. models

698   predicting trials from speaker/signer 1 to be more dissimilar than trials from

699   speaker/signer 2, and vice versa. The purpose of this model was to test for neural

700   dissimilarities based on lower level acoustic and visual features.

701   These models can be tested ***within-modality***, e.g. correlated within speech-

702   speech and sign-sign distances combined or separately, or ***across-modality***, e.g.

703   correlated with speech-sign distances. The testing of models using ***across-modality***

704   distances is equivalent to cross decoding representational structure between speech

705   and sign, positive evidence provides support for common representational structure

706   across languages [50]. Note that we only test for ***across-modality*** semantic

707   representations in areas in which there is evidence of ***within-modality***

708   representational structure. As negative correlations are not plausible, greater than 0

709   model fits were assessed with one-tailed, one sample t-tests. Two-tailed paired t-

710   tests were used to assess differences in fit between models. Multidimensional

711   Scaling (MDS) was conducted to visualise the similarity structure of the RDMs by

712   calculating the averaged participant RDM and applying non-metric MDS, consistent

713   with the non-parametric correlational approach.

714   It is important that the RSA models were evaluated within regions of interest

715   that were defined in a manner that is statistically unbiased [51]. We tested RSA

716   models in regions identified as having positive within modality distances or larger

717   relative distances for speech than sign, and vice versa. The between speaker and/or

718   between signer distances were used to define ROIs. Analyses that evaluate models

719   that use only the between speaker and signer distances are orthogonal to ROI

720   selection. This is because the mean distance is implicitly subtracted out in the

31

721 correlation between the model and the distances [52]. This is true of all the models

722 tested in this study except the speaker and signer identity models. These models

723 predict larger distances for the between speaker/signer than the within

724 speaker/signer distances. As the ROIs are defined on the basis that they show non-

725 zero across speaker/signer distances, the testing of these models would not be

726 orthogonal to ROI selection. Therefore, for these models, to ensure that ROI

727 selection was orthogonal, we generated leave-one-participant-out ROIs to evaluate

728 the fit of the speech and signer identity models [53]. That is, to identify an ROI for

729 Participant 1, we re-estimated the random effects t-test using the whole-brain

730 searchlight maps for the within modality, speech > sign and sign > speech distances,

731 with Participants 2 to 17, and so forth for all participants. We thresholded these maps

732 at $p < 0.001$ (uncorrected) to extract the clusters. This threshold identified discrete

733 clusters, in the same regions as the full group model in all leave-one-out

734 permutations. This generated 17 subtly different ROIs, that were statistically

735 independent, which were used to evaluate the model fit of the speaker/signer identity

736 models (see Figure S1 for the location and overlap between these ROIs).

737

DATA AND CODE AVAILABILITY

739 Anonymised group level data and stimulus materials are available at Mendeley Data

740 (DOI: 10.17632/3d983g83v5.1). The raw MRI data supporting the current study

741 have not been deposited in a public repository, as the participants did not consent to

742 sharing their data publicly. However, these data are available upon request.

743

REFERENCES

744

745　1.　Simanova, I., Hagoort, P., Oostenveld, R., and van Gerven, M.A.J. (2014).

746　　　Modality-independent decoding of semantic information from the human brain.

747　　　Cereb. Cortex *24*, 426–34. Available at:

748　　　http://www.ncbi.nlm.nih.gov/pubmed/23064107 [Accessed May 12, 2015].

749　2.　Fairhall, S.L., and Caramazza, A. (2013). Brain Regions That Represent

750　　　Amodal Conceptual Knowledge. J. Neurosci. *33*, 10552–10558. Available at:

751　　　http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0051-13.2013.

752　3.　Shinkareva, S. V, Malave, V.L., Mason, R. a, Mitchell, T.M., and Just, M.A.

753　　　(2011). Commonality of neural representations of words and pictures.

754　　　Neuroimage *54*, 2418–25. Available at:

755　　　http://www.ncbi.nlm.nih.gov/pubmed/20974270 [Accessed January 9, 2015].

756　4.　Anthony, J., and Francis, D. (2005). Development of Phonological Awareness

757　　　skill. Curr. Dir. Psychol. Sci. *14*, 255–259.

758　5.　Araújo, S., Fernandes, T., and Huettig, F. (2018). Learning to read facilitates

759　　　the retrieval of phonological representations in rapid automatized naming:

760　　　Evidence from unschooled illiterate, ex-illiterate, and schooled literate adults.

761　　　Dev. Sci., e12783.

762　6.　Paivio, A. (1991). Canadian Journal of Psychology. Can. J. Psychol. *45*, 255–

763　　　287.

764　7.　Correia, J., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., and Bonte,

765　　　M. (2014). Brain-Based Translation: fMRI Decoding of Spoken Words in

766　　　Bilinguals Reveals Language-Independent Semantic Representations in

767　　　Anterior Temporal Lobe. J. Neurosci. *34*, 332–338. Available at:

33

768   http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.1302-13.2014.

769   8.   Buchweitz, A., Shinkareva, S. V., Mason, R.A., Mitchell, T.M., and Just, M.A.

770   (2012). Identifying bilingual semantic neural representations across languages.

771   Brain Lang. *120*, 282–289. Available at:

772   http://dx.doi.org/10.1016/j.bandl.2011.09.003.

773   9.   MacSweeney, M., Woll, B., Campbell, R., Calvert, G.A., McGuire, P.K., David,

774   A.S., Simmons, A., and Brammer, M.J. (2002). Neural correlates of British sign

775   language comprehension: spatial processing demands of topographic

776   language. J. Cogn. Neurosci. *14*, 1064–1075. Available at:

777   http://dx.doi.org/10.1162/089892902320474517.

778   10.   Petitto, L.A., Zatorre, R.J., Gauna, K., Nikelski, E.J., Dostie, D., and Evans,

779   A.C. (2000). Speech-like cerebral activity in profoundly deaf people processing

780   signed languages: Implications for the neural basis of human language. Proc.

781   Natl. Acad. Sci. *97*, 13961–13966. Available at:

782   http://www.pnas.org/cgi/doi/10.1073/pnas.97.25.13961.

783   11.   Emmorey, K., McCullough, S., and Weisberg, J. (2014). Neural correlates of

784   fingerspelling, text, and sign processing in deaf American Sign Language–

785   English bilinguals. Lang. Cogn. Neurosci. *30*, 749–767. Available at:

786   http://www.tandfonline.com/doi/full/10.1080/23273798.2015.1014924.

787   12.   Sakai, K.L., Tatsuno, Y., Suzuki, K., Kimura, H., and Ichida, Y. (2005). Sign

788   and speech: Amodal commonality in left hemisphere dominance for

789   comprehension of sentences. Brain *128*, 1407–1417.

790   13.   Soderfeldt, B., Ronnberg, J., and Risberg, J. (1994). Regional Cerebral Blood

791   Flow in Sign Language Users. Brain Lang. *46*, 59–68.

34

792 14. Lambon Ralph, M., Jefferies, E., Patterson, K., and Rogers, T.T. (2016). The

793   neural and computational bases of semantic cognition. Nat. Rev. Neurosci. *18*,

794   42–55. Available at: http://dx.doi.org/10.1038/nrn.2016.150.

795 15. Połczyńska, M.M., Japardi, K., and Bookheimer, S.Y. (2017). Lateralizing

796   language function with pre-operative functional magnetic resonance imaging in

797   early proficient bilingual patients. Brain Lang. *170*, 1–11.

798 16. Newman, A.J., Bavelier, D., Corina, D., Jezzard, P., and Neville, H.J. (2002). A

799   critical period for right hemisphere recruitment in American Sign Language

800   processing. Nat. Neurosci. *5*, 76–80.

801 17. Price, C.J., Wise, R.J.S., and Frackowiak, R.S.J. (1996). Demonstrating the

802   implicit processing of visually presented words and pseudowords. Cereb.

803   Cortex *6*, 62–70.

804 18. Vitello, S., Warren, J.E., Devlin, J.T., and Rodd, J.M. (2014). Roles of frontal

805   and temporal regions in reinterpreting semantically ambiguous sentences.

806   Front. Hum. Neurosci. *8*, 1–14. Available at:

807   http://journal.frontiersin.org/article/10.3389/fnhum.2014.00530/abstract.

808 19. Macsweeney, M., Woll, B., Campbell, R., Mcguire, P.K., David, A.S., Williams,

809   S.C.R., Suckling, J., Calvert, G.A., and Brammer, M.J. (2002). Neural systems

810   underlying British Sign Language and audio-visual English processing in native

811   users. Brain *125*, 1583–1593.

812 20. MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A.S.,

813   McGuire, P.K., Calvert, G. a, and Brammer, M.J. (2004). Dissociating linguistic

814   and nonlinguistic gestural communication in the brain. Neuroimage *22*, 1605–

815   18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15275917 [Accessed

35

816   May 29, 2015].

817   21.   MacSweeney, M., Campbell, R., Woll, B., Brammer, M.J., Giampietro, V.,

818   David, A.S., Calvert, G. a, and McGuire, P.K. (2006). Lexical and sentential

819   processing in British Sign Language. Hum. Brain Mapp. *27*, 63–76. Available

820   at: http://www.ncbi.nlm.nih.gov/pubmed/15966001 [Accessed June 1, 2015].

821   22.   Sapir, E. (1929). The Status of Linguistics as a Science. Linguist. Soc. Am. *5*,

822   207–214.

823   23.   Whorf, B.L. (1956). Language, thought and reality: Essays by B. L. Whorf J. B.

824   Carroll, ed. (Cambridge M.A.: MIT Press).

825   24.   Van de Putte, E., De Baene, W., Price, C.J., and Duyck, W. (2018). "Neural

826   overlap of L1 and L2 semantic representations across visual and auditory

827   modalities: a decoding approach." Neuropsychologia *113*, 68–77. Available at:

828   https://doi.org/10.1016/j.neuropsychologia.2018.03.037.

829   25.   de Groot, A.M.B. (1992). Determinants of Word Translation. J. Exp. Psychol.

830   Learn. Mem. Cogn. *18*, 1001–1018.

831   26.   Duyck, W., and Brysbaert, M. (2004). Forward and backward number

832   translation requires conceptual mediation in both balanced and unbalanced

833   bilinguals. J. Exp. Psychol. Hum. Percept. Perform. *30*, 889–906.

834   27.   Van Hell, J., and De Groot, A. (2003). Conceptual representation in bilingual

835   memory: Effects of concreteness and cognate status in word association.

836   Biling. Lang. Cogn. *1*, 193–211. Available at:

837   http://www.journals.cambridge.org/abstract_S1366728998000352.

838   28.   Zwitserlood, I., Kristoffersen, J., and Troelsgard, T. (2013). Issues in Sign

839   Language Lexicography. In The Bloomsbury Campanion to Lexicography, H.

840   Jackson, ed. (London: Bloomsbury), pp. 259–283.

841   29. Marslen-Wilson, W. (1995). Activation, Competition, and Frequency in Lexical

842   Access. Cogn. Model. Speech Process. Psycholinguist. Comput. Perspect.,

843   148–172. Available at:

844   http://books.google.com/books?id=2R7XTlebSF8C&printsec=frontcover%5Cn

845   papers://5860649b-6292-421d-b3aa-1b17a5231ec5/Paper/p8360.

846   30. Apfelbaum, K.S., Blumstein, S.E., and Mcmurray, B. (2011). Semantic priming

847   is affected by real-time phonological competition: Evidence for continuous

848   cascading systems. Psychon. Bull. Rev. *18*, 141–149.

849   31. Perniss, P., Thompson, R.L., and Vigliocco, G. (2010). Iconicity as a general

850   property of language: Evidence from spoken and signed languages. Front.

851   Psychol. *1*, 1–15.

852   32. Marshall, C., Rowley, K., and Atkinson, J. (2013). Modality-Dependent and -

853   Independent Factors in the Organisation of the Signed Language Lexicon:

854   Insights From Semantic and Phonological Fluency Tasks in BSL. J.

855   Psycholinguist. Res. *43*, 587–610.

856   33. Padden, C.A., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., and Sampson, T.

857   (2014). Patterned iconicity in sign language lexicons. Gesture *13*, 287–308.

858   34. MacSweeney, M., Capek, C.M., Campbell, R., and Woll, B. (2008). The

859   signing brain: the neurobiology of sign language. Trends Cogn. Sci. *12*, 432–

860   40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18805728 [Accessed

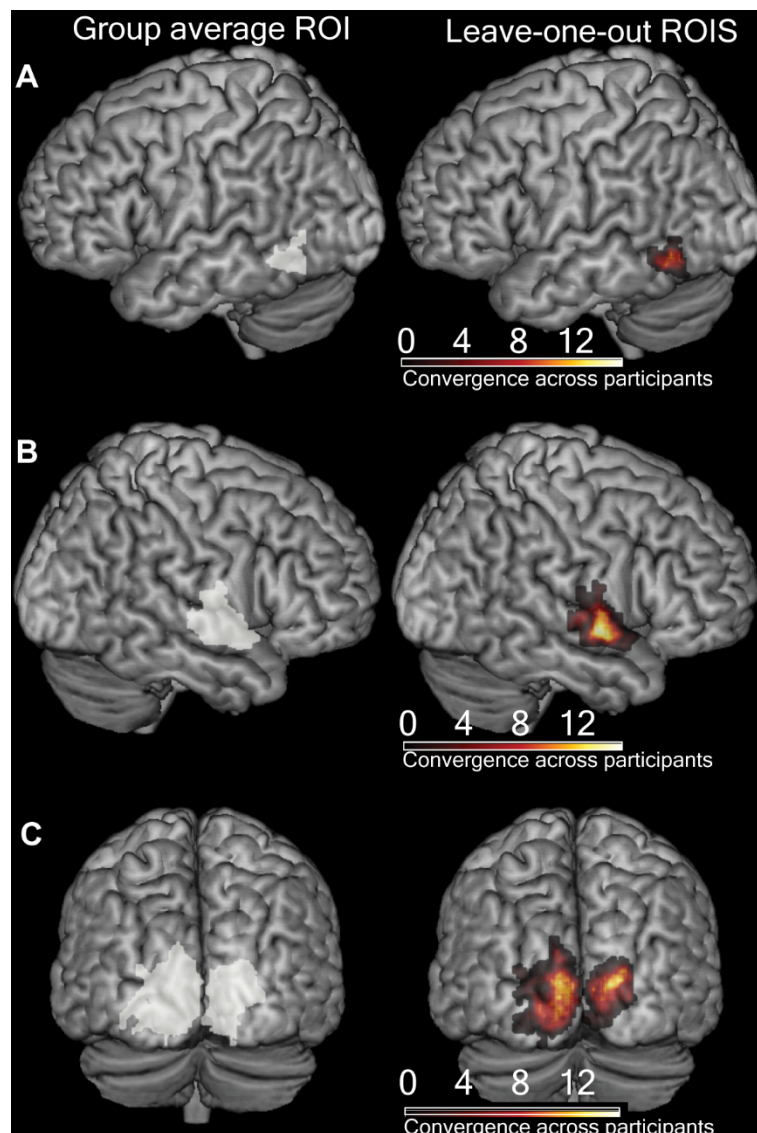861   August 29, 2015].

37

862   35.   Devereux, B.J., Tyler, L.K., Geertzen, J., and Randall, B. (2014). The Centre

863   for Speech, Language and the Brain (CSLB) concept property norms. Behav.

864   Res. Methods *46*, 1119–1127.

865   36.   Davis, C.J. (2005). N-Watch: A program for deriving neighbourhood size and

866   other pyscholinguistic statistics. Behav. Res. Methods *37*, 65–70.

867   37.   Kuperman, V., Stadthagen-Gonzalez, H., and Brysbaert, M. (2012). Age-of-

868   acquisition ratings for 30,000 English words. Behav. Res. Methods *44*, 978–

869   990.

870   38.   Wilson, M. (1988). The MRC Psycholinguistic Database: Machine Readable

871   Dictionary, Version 2. Behav. Res. Methods, Instruments Comput. *20*, 6–11.

872   39.   Vinson, D.P., Cormier, K., Denmark, T., Schembri, A., and Vigliocco, G.

873   (2008). The British Sign Language (BSL) norms for age of acquisition,

874   familiarity, and iconicity. Behav. Res. Methods *40*, 1079–1087.

875   40.   Levenshtein, V.I. (1966). Binary codes capable of correcting deletions,

876   insertions, and reversals. Sov. Phys. Dokl. *10*, 707–710.

877   41.   Fenlon, J., Cormier, K., and Schembri, A. (2015). Building BSL SignBank: The

878   lemma dilemma revisited. Int. J. Lexicogr. *28*, 169–206.

879   42.   Kriegeskorte, N., and Mur, M. (2012). Inverse MDS: Inferring dissimilarity

880   structure from multiple item arrangements. Front. Psychol. *3*, 1–13.

881   43.   Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., and

882   Kriegeskorte, N. (2014). A toolbox for representational similarity analysis.

883   PLoS Comput. Biol. *10*, e1003553. Available at:

884   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3990488&tool=pmc

885   entrez&rendertype=abstract [Accessed May 23, 2014].

886   44.   Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J.

887   (2015). Reliability of dissimilarity measures for multi-voxel pattern analysis.

888   Neuroimage.

889   45.   Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based

890   functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. *103*, 3863–3868.

891   Available at: http://www.pnas.org/content/103/10/3863.abstract.

892   46.   Evans, S., and Davis, M.H. (2015). Hierarchical organization of auditory and

893   motor representations in speech perception: Evidence from searchlight

894   similarity analysis. Cereb. Cortex *25*.

895   47.   Poldrack, R., Baker, C.I., Durnez, J., Gorgolewski, K., Matthews, P.M.,

896   Munafo, M., Nichols, T., Poline, J.-B., Vul, E., and Yarkoni, T. (2017).

897   Scanning the Horizon: challenges and solutions for neuroimaging research.

898   Nat. Rev. Neurosci. *18*, 115–126. Available at:

899   http://biorxiv.org/lookup/doi/10.1101/059188.

900   48.   Friston, K.J., Rotshtein, P., Geng, J.J., Sterzer, P., and Henson, R.N. (2006). A

901   critique of functional localisers. Neuroimage *30*, 1077–1087. Available at:

902   http://www.sciencedirect.com/science/article/pii/S1053811905006026.

903   49.   Walker, D.A. (2017). JMASM9: Converting Kendall's Tau For Correlational Or

904   Meta-Analytic Analyses. J. Mod. Appl. Stat. Methods *2*, 525–530.

905   50.   Kaplan, J.T., Man, K., and Greening, S.G. (2015). Multivariate cross-

906   classification: applying machine learning techniques to characterize

907   abstraction in neural representations. Front. Hum. Neurosci. *9*, 151. Available

908   at: /pmc/articles/PMC4373279/?report=abstract.

909   51.   Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S.F., and Baker, C.I. (2009).

910   Circular analysis in systems neuroscience: the dangers of double dipping. Nat.

911   Neurosci. *12*, 535–540. Available at: isi:000265575400006.

912   52.   Diedrichsen, J., and Kriegeskorte, N. (2017). Representational models: A

913   common framework for understanding encoding, pattern-component, and

914   representational-similarity analysis. PLoS Comput. Biol. *13*, 1–33.

915   53.   Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.-C., and Yantis, S. (2010).

916   Avoiding non-independence in fMRI data analysis: leave one subject out.

917   Neuroimage *50*, 572–6. Available at:

918   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2823971&tool=pmc

919   entrez&rendertype=abstract [Accessed September 22, 2014].

920

921

922

923

924

925

926

927

928

929

40

930 **SUPPLEMENTARY MATERIALS**

931

932 **Figure S1: Definition of leave-one-participant-out regions of Interest, Related to STAR**
933 **Methods, Figure 2, Figure 3 and Figure 4**

934 A leave-one-participant-out procedure [S1] was used to test for speaker and signer identity
935 models to ensure that the evaluation of the model was statistically independent of the
936 process used to generate the ROIs. Rendered with MRICRON on the Ch2better brain.

937 (A-C) Left side panels show the group average ROI using the data from all participants.
938 Right side panels show the overlap of the leave-one-participant-out ROIs across
939 participants.

940 (A) ROI in the left pMTG/ITG [-48 -62 -6] generated by a searchlight analysis testing for > 0
941 within modality distances and associated leave-one-participant-out ROIs.

942 (B) ROI in the right anterior STG [58 -4 -2] generated by a searchlight analysis testing for
943 speech > sign distances and associated leave-one-participant-out ROIs.

41

944   (C) ROIs in the left V1-V3 [-6 -98 16] and the right V1-V3 [22 -90 16] generated by a
945   searchlight analysis testing for sign > speech distances and associated leave-one-
946   participant-out ROIs.
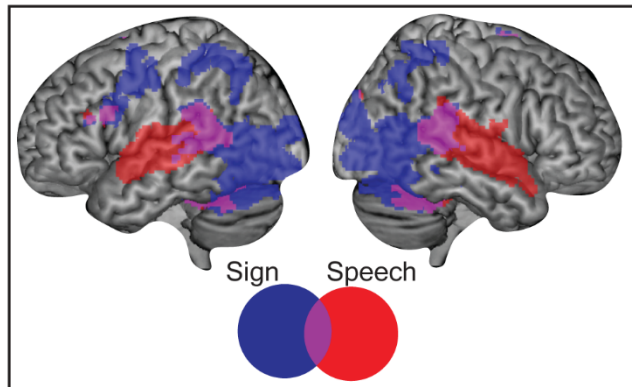
947

948   **Figure S2: Univariate overlap between sign and speech, Related to Figure 2**

949   Areas responding to speech (red) and sign (blue) compared to rest and their overlap (pink),
950   thresholded at $p < 0.005$ peak level, $q < 0.05$ FDR corrected at the cluster level. Rendered
951   with MRICRON on the Ch2better brain. As expected, areas of shared univariate activity for
952   sign and speech were found in the bilateral posterior superior and the middle temporal
953   gyrus, the left inferior frontal gyrus and bilateral cerebellum.
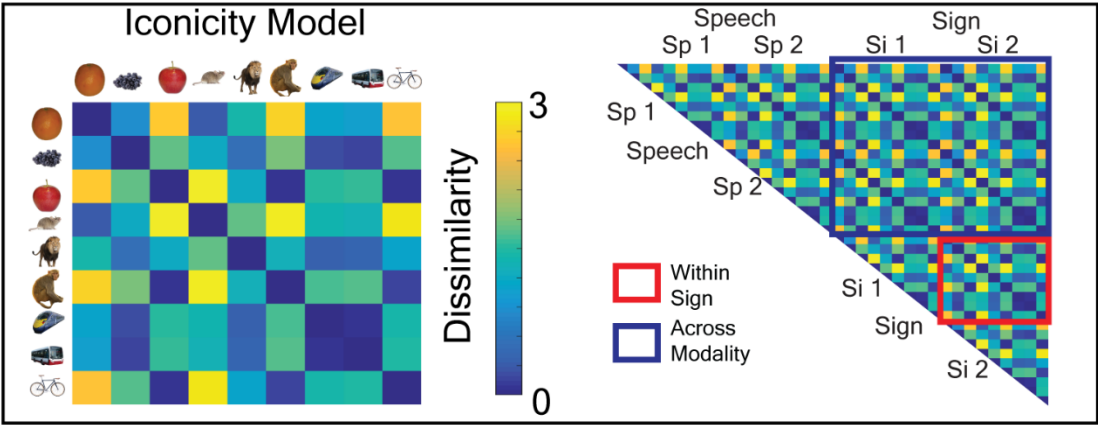
954

955

956

957

958

959

960

961

962

Figure S3: Effects of iconicity, Related to Figure 2

An iconicity model (left) was derived from the group average iconicity ratings for each sign. This model was created by taking the absolute value from the subtraction of the average iconicity value of each sign from every other sign. The model was tested (right) on the sign-sign distances (red box), e.g. within sign, and the speech-sign distances (blue box), e.g. across-modality. There was no significant fit to the within sign ($t(16) = 0.382$, $p = 0.354$, $d_z = 0.093$) or across-modality distances ($t(16) = 1.298$, $p = 0.106$, $d_z = 0.315$) in the left pMTG/ITG. An additional iconicity model was tested that used each individuals' iconicity ratings for each exemplar of each sign. As with the model using the group averaged iconicity values, there was no significant fit in the within sign ($t(16) = 0.588$, $p = 0.282$, $dz = 0.143$) or across-modality distances ($t(16) = 0.277$, $p = 0.393$, $dz = 0.067$).
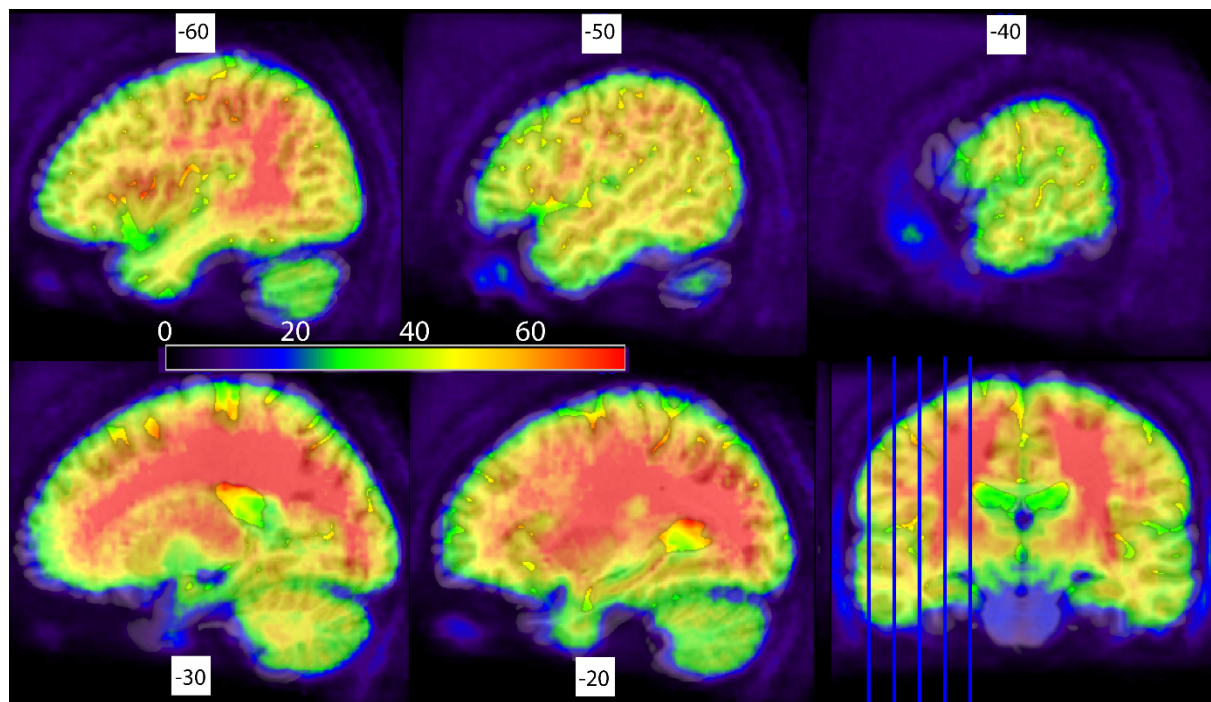
43

Figure S4: Signal quality in the temporal lobe, Related to Figure 2

Whole brain tSNR maps for the group. Sagittal slices of the left temporal lobe are shown. The mid-anterior temporal lobe has been ascribed an important role in 'amodal' semantic cognition. Within this area, a gradient of function from posterior-anterior has been suggested that reflects a wider-to-narrower window of semantic specificity, e.g. from categories to items and individual exemplars [S2,S3]. This region is particularly susceptible to signal drop out [S4,S5]. However, tSNR maps indicated relatively good signal quality in the mid-anterior inferior temporal cortex and drop out that was similar to that found in the left pMTG/ITG. We chose not to use a dual echo sequence to mitigate against drop out, as our sequence was optimised for signal quality in the posterior temporal cortex, the region most consistently activated by both sign and speech in previous univariate studies. The absence of shared item-level correspondences might also reflect the fact that participants were asked to monitor for category rather than item-level distinctions [S6]. We decided to use a category-based task to maximise the likelihood of finding commonality between the languages, which we assumed would be more robust at a broader level of semantic specificity. Future studies using dual echo sequences and item-level discriminative tasks are necessary to exclude the possibility that these methodological details obscured identification of item-level correspondences.

44

1007

Item	Age of acquisition	Imageability	Familiarity	Syllables	Phonemes	Iconicity
Orange	3.26	626	567	2	5	2.56
Grapes	3.94[a]	591[a]	532[a]	1	5	3.50
Apple	4.15	637	598	2	3	5.35
Mean	**3.78**	**618**	**566**	**2**	**4**	**3.80**
Mouse	4.94	615	520	1	3	2.24
Lion	4.42	626	511	2	4	4.09
Monkey	4.21	588	531	2	5	5.44
Mean	**4.52**	**610**	**521**	**2**	**4**	**3.92**
Train	4	593	548	1	4	**3.74**
Bus	3.85	624[c]	513[c]	1	3	**3.68**
Bicycle	4.26	649[b]	591[b]	3	6	**5.26**
Mean	**4.04**	**622**	**551**	**2**	**4**	**4.23**

1008

1009 **Table S1: The psycholinguistic properties of the core items, Related to STAR**
1010 **Methods.**

1011 Imageability (Bristol/MRC), subjective familiarity (MRC database), number of syllables and
1012 phonemes extracted from the N-Watch program [S7], age of acquisition was extracted from
1013 Kuperman et al. [S8] and iconicity values were acquired directly from the participants. Note
1014 that [a]the term "grape" was used in the absence of the term "grapes" for age of acquisition,
1015 familiarity and imageability ratings, [b]the term "bike" was used in the absence of the term
1016 "bicycle" for familiarity and imageability ratings and [c]the term coach was used in the absence
1017 of "bus" for familiarity and imageability ratings.

1018

1019

1020

1021

Region	x	y	z	Extent	Z Value
Within-modality representational structure					
Right superior temporal gyrus	58	-4	-2	1545	5.283
Right inferior parietal lobule	64	-30	14		4.968
Right superior temporal gyrus	52	-2	-8		4.861
Left superior occipital gyrus	-14	-96	10	2629	4.677
Right superior occipital gyrus	14	-100	16		4.479
Right cuneus	6	-92	22		4.226
Left superior temporal gyrus	-60	-10	-2	1276	4.500
Left middle temporal gyrus	-64	-30	6		4.476
Left middle temporal gyrus	-64	-44	2		4.175
Left inferior temporal gyrus	-48	-62	-6	172	4.361
Left middle occipital gyrus	-42	-64	0		3.122
Right insula	36	-12	14	194	4.178
Right putamen	30	-8	10		4.160
Right middle temporal gyrus	52	-68	6	279	3.954
Right middle temporal gyrus	56	-48	0		3.748
Right middle temporal gyrus	54	-54	6		3.574
Speech > Sign					
Right superior temporal gyrus	58	-4	-2	754	4.877
Right superior temporal gyrus	52	0	-8		4.779
Right superior temporal gyrus	60	-12	4		3.590
Left superior temporal gyrus	-56	-8	2	743	4.484
Left superior temporal gyrus	-62	-30	10		4.253
Left superior temporal gyrus	-62	-2	0		3.720

46

Right Putamen	30	-10	10	146	4.364
Right Insular	40	-12	10		3.354
Right superior temporal gyrus	58	-34	18	285	4.160
Right superior temporal gyrus	66	-32	14		3.763
Right superior temporal gyrus	56	-26	0		3.722
Sign > Speech					
Left cuneus	-6	-98	16	1145	4.623
Left middle occipital gyrus	-12	-102	4		4.019
Left cuneus	-8	-94	28		3.830
Right superior occipital gyrus	22	-90	16	969	4.375
Right lingual gyrus	16	-84	-4		3.976
Right cuneus	16	-100	12		3.655
Left inferior occipital gyrus	-44	-80	-6	264	4.107
Left middle occipital gyrus	-50	-72	-2		3.937
Left middle occipital gyrus	-42	-80	4		3.449
Left cerebellum	-4	-48	-8	116	3.808
Left lingual gyrus	-10	-56	-2		3.767
Left cerebellum	-4	-50	0		3.102
Left superior occipital gyrus	-10	-84	42	127	3.781
Left superior occipital gyrus	-16	-78	40		3.396
Left superior parietal lobule	-26	-80	48		3.172

1022

Table S2: **MNI coordinates for RSA searchlight analyses, Related to Figure 2, Figure 3 and Figure 4.**

3 local maxima more than 8 mm apart

1026

47

1027

1028 **SUPPLEMENTAL REFERENCES**

1029 S1. Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.-C., and Yantis, S. (2010).

1030   Avoiding non-independence in fMRI data analysis: leave one subject out.

1031   Neuroimage *50*, 572–6. Available at:

1032   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2823971&tool=pmc

1033   entrez&rendertype=abstract [Accessed September 22, 2014].

1034 S2. Lambon Ralph, M., Jefferies, E., Patterson, K., and Rogers, T.T. (2016). The

1035   neural and computational bases of semantic cognition. Nat. Rev. Neurosci. *18*,

1036   42–55. Available at: http://dx.doi.org/10.1038/nrn.2016.150.

1037 S3. Clarke, A., and Tyler, L.K. (2015). Understanding What We See: How We

1038   Derive Meaning From Vision. Trends Cogn. Sci. *19*, 677–687. Available at:

1039   http://dx.doi.org/10.1016/j.tics.2015.08.008.

1040 S4. Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E.,

1041   Matthews, P.M., and Tyler, L.K. (2000). Susceptibility-induced loss of signal:

1042   Comparing PET and fMRI on a semantic task. Neuroimage *11*, 589–600.

1043   Available at: isi:000087963600002.

1044 S5. Halai, A.D., Welbourne, S.R., Embleton, K., and Parkes, L.M. (2014). A

1045   comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal

1046   lobe. Hum. Brain Mapp. *35*, 4118–28. Available at:

1047   http://www.ncbi.nlm.nih.gov/pubmed/24677506 [Accessed July 22, 2015].

1048 S6. Bonte, M., Hausfeld, L., Scharke, W., Valente, G., and Formisano, E. (2014).

1049   Task-dependent decoding of speaker and vowel identity from auditory cortical

48

1050 response patterns. J. Neurosci. *34*, 4548–57. Available at:

1051 http://www.ncbi.nlm.nih.gov/pubmed/24672000 [Accessed June 3, 2014].

1052 S7. Davis, C.J. (2005). N-Watch: A program for deriving neighbourhood size and

1053 other pyscholinguistic statistics. Behav. Res. Methods *37*, 65–70.

1054 S8. Kuperman, V., Stadthagen-Gonzalez, H., and Brysbaert, M. (2012). Age-of-

1055 acquisition ratings for 30,000 English words. Behav. Res. Methods *44*, 978–

1056 990.

1057

1058