Neural correlates of object and action naming practice

Ekaterina Delikishkina\textsuperscript{1,4}, Angelika Lingnau\textsuperscript{1,2,3}, Gabriele Miceli\textsuperscript{1,2,4,5*}

\textsuperscript{1}Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy
\textsuperscript{2}Department of Psychology & Cognitive Science, University of Trento, Italy
\textsuperscript{3}Institute of Psychology, University of Regensburg, Germany
\textsuperscript{4}International Doctorate for Experimental Approaches to Language and Brain (IDEALAB), Universities of Trento, Groningen, Potsdam, Newcastle and Macquarie University
\textsuperscript{5}Centro Linceo Interdisciplinare ‘Beniamino Segre’, Accademia dei Lincei, Rome, Italy

*Corresponding author: Prof. Gabriele Miceli, Center for Mind/Brain Sciences, University of Trento. Corso Bettini, 31, 38068 Rovereto (TN), Italy. E-mail: gabriele.miceli@unitn.it
Abstract

Word retrieval deficits are a common problem in patients with stroke-induced brain damage. While complete recovery of language in chronic aphasia is rare, patients’ naming ability can be significantly improved by speech therapy. A growing number of neuroimaging studies have tried to pinpoint the neural changes associated with successful outcome of naming treatment. However, the mechanisms supporting naming practice in the healthy brain have received little attention. Yet, understanding these mechanisms is crucial for teasing them apart from functional reorganization following brain damage. To address this issue, we trained a group of healthy monolingual Italian speakers on naming pictured objects and actions for ten consecutive days and scanned them before and after training. Although activity during object vs. action naming dissociated in several regions (lateral occipitotemporal, parietal and left inferior frontal cortices), training effects for the two word classes were similar and included activation decreases in classical language regions of the left hemisphere (posterior inferior frontal gyrus, anterior insula), potentially due to decreased lexical selection demands. Additionally, MVPA revealed training-related activation changes in the left parietal and temporal cortices associated with the retrieval of knowledge from episodic memory (precuneus, angular gyrus) and facilitated access to phonological word forms (posterior superior temporal sulcus).

Keywords

Language training, Object naming, Action naming, fMRI, MVPA
1. Introduction

Attempted naming can improve performance in aphasic individuals even in the absence of feedback or corrections (Howard, 2000; Nickels, 2002). Although the neural mechanisms underlying this improvement are not clear, it has been recently suggested that they may at least partially overlap with those that support naming facilitation in healthy controls (Heath et al., 2015; Kurland, Liu, & Stokes, 2018). Thus, identifying the neural changes induced by training in healthy participants is necessary to establish a “baseline”, against which the results of anomic patients could be compared.

Studies of incidental naming practice show that even a single instance of naming in the context of a picture naming task can facilitate subsequent processing of a stimulus for days and even weeks (Meister et al., 2005; van Turennout, Bielamowicz, & Martin, 2003; van Turennout, Ellmore, & Martin, 2000). While behaviorally this effect, known as repetition priming, manifests itself as shorter naming latencies, at the neural level it is reflected by decreased activity (or “repetition suppression”) in bilateral occipitotemporal and left prefrontal cortices, associated with facilitated perceptual/conceptual and linguistic processing of a stimulus respectively.

The effects of explicit naming practice were addressed by Basso et al. (2013), who used an intensive training paradigm, more closely resembling speech therapy in patients. Explicit training of object naming (ten repetitions per day over ten consecutive days) was associated with decreased BOLD response in the left inferior frontal cortex and the fusiform gyrus, in line with the studies on repetition suppression, and with increased response in the precuneus and the posterior cingulate cortex. Increased activation of the medial parietal areas, which are not involved in the classic language network, was attributed to retrieval of memories related to practiced items from long-term memory. Similar findings were reported by MacDonald et al. (2015) in healthy older adults, who showed increased activity in the precuneus and decreased activity in the inferior frontal and inferior temporal cortices bilaterally following two object naming sessions (three repetitions/session).

Practice-related activation changes could be modulated by a number of factors (intensity of practice, interval between stimulus repetitions, etc.). Yet, one critical factor, namely, the content of training, has received little attention. Most studies have focused on practiced naming of objects, that are referred to by nouns. Neuropsychological findings (for reviews, see Cappa & Perani, 2003; Mätzig, Druks, Masterson, & Vigliocco, 2009), as well as recent neuroimaging studies with healthy individuals (for reviews, see Crepaldi, Berlingeri, Paulesu, & Luzzatti, 2011; Vigliocco, Vinson, Druks, Barber, & Cappa, 2011) suggest that words belonging to different grammatical classes, such
as nouns and verbs, may have at least partially dissociable neural correlates. Thus, it seems reasonable to expect them to be differently affected by practice. A recent study by Kurland et al. (2018) attempted to address this question by including both nouns and verbs in their training protocol, but failed to find a significant interaction between training and word class. It is possible, however, that these null results were due to low intensity of practice (five repetitions a few days prior to the fMRI session, plus five repetitions immediately before the scanning).

In the present fMRI study we investigated differences in the magnitude and localization of training effects for nouns and verbs. Since subjects were instructed to remain silent inside the scanner during picture presentation (in order to avoid jaw movement artifacts) until they saw the next slide with a response cue, we gathered reaction time (RT) data from a separate group of volunteers who participated in an analogous study in which they were asked to produce a word as soon as they saw a picture. Healthy speakers of Italian practiced naming of objects and actions for ten consecutive days and were tested twice, on the days preceding and following the training. The two experimental sessions were identical and included trained items, as well as an equal number of untrained items that served to control for task habituation and priming effects. The use of this paradigm allowed us to investigate: (1) the putative distinctions in the neural representations of objects (nouns) and actions (verbs), (2) the effects of training and their potential interaction with word class, and (3) the reliability of long-term priming effects reported in previous studies (Meister et al., 2005; Meltzer, Postman-Caucheteux, McArdle, & Braun, 2009), separated from the effects of explicit training.

2. Materials and Methods

2.1. Participants

A total of 35 native Italian speakers took part in this project—12 subjects (3 male, mean age: 23.3 ± 2.5 years, age range: 19-28 years) participated in the behavioral study, and 23 (9 male, mean age: 23.7 ± 3.3 years, age range: 19-32 years) in the fMRI study. Three subjects of the fMRI study were subsequently excluded from data analyses—two because of excessive head motion during scanning (more than 3 mm in one of the directions) and one due to non-compliance with the training protocol. All participants but one were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971); the remaining subject was a self-reported right-hander, but scored as ambidextrous on the Inventory. All participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric disease.
The study was conducted in compliance with the Declaration of Helsinki and was approved by the Human Research Ethics Committee of the University of Trento. All participants signed informed consent forms.

2.2. Stimuli

2.2.1. Preliminary naming task

All participants of the behavioral (N = 12) and the fMRI (N = 23) study were required to undergo the preliminary naming task prior to entering the two-session training study, in order to assure that they recognized objects and actions that would be presented in the two experimental sessions and retrieved their corresponding names. Stimuli consisted of line drawings of 80 objects and 80 actions that were presented at a comfortable pace using PowerPoint. Subjects were instructed to produce names of objects using an Italian noun in a singular form (without an article) and to name actions using a verb in the infinitive form. Part of the stimuli were specifically drawn for the present study, while others were selected from various sources, including the Verb and Action Test (VAT; Bastiaanse, Wieling, & Wolthuis, 2016), the Battery for the Analysis of Aphasic Deficits (BADA; Miceli, Laudanna, & Capasso, 2001), as well as the public domain (see examples of drawings in Fig. 1).

2.2.2. Experimental naming task

Of the 160 drawings presented in the preliminary task, 40 object and 40 action pictures were used in two identical experimental naming sessions. Half of the items in each set were included in the training protocol, while the other half were not explicitly trained and served as controls for potential stimulus priming/task habituation effects. The four resulting 20-item subsets—untrained nouns (NU), trained nouns (NT), untrained verbs (VU) and trained verbs (VT)—were matched for variables that reportedly affect word retrieval. Words in the four subsets were balanced for phonemic ($H(3) = .43, p = .934$) and syllabic ($H(3) = .804, p = .848$) length, as well as for relative lemma frequency ($H(3) = .006, p = .996$), based on a lexical database of written Italian (Corpus e Lessico di Frequenza dell’Italiano Scritto, CoLFIS; Bertinetto et al., 2005). Online questionnaires, created on the website SurveyMonkey.com, were delivered to separate groups of Italian native speakers in order to balance target words for familiarity (60 participants; $H(3) = 2.629, p = .452$),
imageability (38 participants; $H(3) = 6.549, p = .088$) and subjective age of acquisition (55 participants; $H(3) = 3.473, p = .324$). Another questionnaire was used to balance naming agreement of pictorial stimuli (48 participants; $H(3) = 2.158, p = .54$). Additionally, pictures were matched for objective visual complexity ($H(3) = .245, p = .97$) using the GIF lossless compression method (Forsythe, Mulhern, & Sawey, 2008). Nouns were selected from a broad range of semantic categories, including animals, professions, clothing, furniture, buildings, vehicles, fruit and vegetables. Verbs were roughly matched for transitivity (VT: 11 transitive, 9 intransitive; VU: 11 transitive, 9 intransitive) and instrumentality (VT: 10 instrumental, 10 non-instrumental; VU: 9 instrumental, 11 non-instrumental). Pictures were normalized with an average brightness of 128 cd/m$^2$. Fourier-transformed phase-scrambled images were additionally introduced into the experimental set as low-level controls. In the second experimental session all images were flipped horizontally in order to reduce potential effects of priming in early visual areas.

2.2.3. Training materials

Subjects were asked to practice overt naming of 20 objects in the NT subset and 20 actions in the VT subset using ten booklets with color photographs (one for each day of training). Photos were taken from the public domain and represented various depictions of to-be-trained objects and actions (see Fig. 1 for examples). Each booklet contained a different exemplar of the same concept, in order to tap into abstract structural representations rather than low-level perceptual features. A booklet was divided into two sections—“Objects” and “Actions”. Items within each section were presented in random order.

2.3. Procedure

The experimental paradigm is schematically depicted in Fig. 1. Subjects underwent intensive naming training for ten consecutive days (excluding weekends). The training material consisted of 20 objects (NT subset) and 20 actions (VT subset). Training was carried out at home, at a time comfortable for a subject. To make sure that participants complied with the training protocol, they were asked to record their responses with the help of a digital recorder. A daily training session consisted of naming all objects and actions in a given booklet for ten times.
All subjects completed two identical experimental sessions—one before and one after the training, either inside or outside of the MRI scanner, while their RTs were measured. In addition to the 40 trained objects and actions (NT and VT subsets), participants were presented with an equal number of untrained items (NU and VU subsets). Their task was to name a depicted object or action aloud, using a single Italian word (a noun without an article, a verb in the infinitive form). Whenever a scrambled image appeared, subjects were instructed to produce a pseudoword—/ber.'to:va/ (in session 1) or /sin.'to:ti/ (in session 2).

![Fig. 1 — The experimental paradigm. Participants in the behavioral and the fMRI studies named color photographs of 20 objects (NT) and 20 actions (VT) for ten consecutive days. Before and after the training program, they underwent two identical experimental naming sessions that included trained items (NT and VT), as well as an equal number of untrained nouns (NU) and untrained verbs (VU), used as controls for potential stimulus/task habituation effects. Fourier-transformed phase-scrambled images (Control), in response to which subjects were instructed to produce a pseudoword, were also presented in both studies.](image-url)
Stimuli were delivered in blocks. A run consisted of four blocks—NU, NT, VU and VT—presented in random order. Each block included five items belonging to one of the four experimental conditions, as well as two randomly interspersed scrambled images. Word class was cued by a colored frame around an image: a red frame for nouns (NU and NT blocks), a blue frame for verbs (VU and VT blocks). A frame was also placed around scrambled images that appeared within a noun or a verb block. Subjects were instructed to produce the same pseudoword irrespective of frame color. All stimuli (N = 80) were presented within four experimental runs, and subsequently repeated within four additional runs in a different order. Due to technical reasons, for one of the participants of the fMRI study only four out of eight runs were acquired in the first experimental session.

Prior to each experimental session, participants received written instructions and underwent short practice. Stimulus presentation and response collection were controlled with ASF (Schwarzbach, 2011), a toolbox based on Psychtoolbox-3 (Brainard, 1997) for MATLAB (MathWorks, Natick, MA, U.S.A.).

2.3.1. Behavioral naming study (N = 12 participants)

The first experimental session took place on the day after the preliminary naming task. Training started 1 to 3 days after the first experimental session (mean: 1.9 days) and finished on the day preceding the second experimental session. Subjects were allowed to refrain from training during weekends (mean: 2 days). Each trial started with a 2 s black fixation cross followed by a 3 s picture presentation. The inter-trial interval (ITI) was set to 1 s. Blanks of 5 s were introduced between blocks, and at the beginning and the end of each run. Subjects were asked to reply as soon as they saw a picture. Stimuli were presented on an LCD screen with the resolution of 1920 × 1080 pixels and the frame rate of 60 Hz.

2.3.2. fMRI naming study (N = 23 participants)

The preliminary naming task was administered 1 to 4 days prior to the first fMRI session (mean interval: 1.5 days). The training procedure started 1 to 4 days after the first fMRI session (mean: 1.9 days) and finished on the day preceding the second fMRI session. Subjects were allowed to take 1-4 days of rest from training (mean: 2.2 days). Each trial started with a black fixation cross whose
presentation lasted between 2 and 5 s. The duration of the initial fixation was chosen from a geometric distribution ($p = .4$; in steps of 1 s). The fixation cross was followed by a picture, presented for 2 s. Subjects were instructed to withhold overt responses while viewing the picture and to respond when a green fixation cross following the picture appeared (3.5 s). The ITI was jittered between 0.5 and 1 s (in steps of 0.25 s). Blanks with a duration of 6 s were introduced between blocks. Each run started and ended with a 12 s blank. In the scanner, stimuli were back-projected onto a screen (frame rate: 60 Hz, screen resolution: $1024 \times 768$ pixels) via a liquid crystal projector (OC EMP 7900, Epson, Nagano, Japan). Participants viewed the screen binocularly through a mirror mounted on the head coil.

2.4. Data acquisition

2.4.1. Behavioral data acquisition

Vocal responses of the participants of the behavioral study were collected using the Samson Q4 microphone with a low-noise microphone cable (Thomann, Burgebrach, Germany). RTs were measured automatically using the voice key function supplied with ASF. Recordings were digitized at a sampling rate of 44.1 kHz.

2.4.2. MR data acquisition

Neuroimaging data were collected at the Functional Neuroimaging Laboratories (LNiF) of the Center for Mind/Brain Sciences (CIMeC) at the University of Trento, Italy, using a 4T Bruker MedSpec MR scanner with an 8-channel birdcage head coil. Functional images were acquired using a T2*-weighted gradient echo-planar imaging (EPI) sequence with fat suppression. Scanning was performed continuously during a functional run with the following parameters: repetition time (TR) = 2.2 s, echo time (TE) = 33 ms, flip angle (FA) = 75°, field of view (FOV) = 192 × 192 mm, matrix size = 64 × 64, voxel resolution = $3 \times 3 \times 3$ mm. We acquired 31 slices in ascending-interleaved odd-even order, with a thickness of 3 mm and a 15% gap (0.45 mm). Slices were aligned to the AC-PC plane. An imaging volume was positioned to cover the entire temporal lobe; as a result, a small portion of the superior parietal cortex was not captured in most subjects. The number of volumes in a functional run varied (range: 130-142) as a result of temporal jittering introduced into trials. Before each run we performed an additional scan measuring the point-spread function (PSF) of the...
acquired sequence, in order to correct the distortion in geometry and intensity expected with high-field imaging (Zaitsev, Hennig, & Speck, 2004; Zeng & Constable, 2002). A T1-weighted structural scan at the beginning of each scanning session served as reference for coregistration of functional data. Structural images were acquired using a magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence (TR = 2.7 s, TE = 4.18 ms, FA = 7°, FOV = 256 × 224 mm, 176 slices, inversion time (TI) = 1020 ms), with generalized autocalibrating partially parallel acquisition (GRAPPA) with an acceleration factor of 2.

2.5. Data analysis

2.5.1. Behavioral data analysis

Voice onset intensity threshold was calibrated for each subject based on the visual inspection of the wave plots of vocal responses with displayed RTs at a given threshold produced by the ASF software for each trial. RTs deviating from a subject’s mean by more than two standard deviations were considered outliers and removed from the analysis (5.2% of the data removed, including 3.4% of object trials, 9.9% of action trials and 1.4% of control trials). After calculating individual descriptive statistics in MATLAB R2015b, the data were submitted to inferential analysis with repeated-measures ANOVAs and paired-samples t-tests in SPSS 24.

2.5.2. fMRI data analysis

Neuroimaging data were preprocessed and analyzed using BrainVoyager QX 2.8.4 (Brain Innovation, Maastricht, the Netherlands) in combination with the NeuroElf toolbox (v. 1.1) and in-house software written in MATLAB. The first three volumes of a functional run were discarded to avoid T1 saturation. For each run, we performed slice timing correction (cubic spline interpolation), followed by 3D motion correction (trilinear interpolation for estimation and sinc interpolation for resampling, all functional volumes acquired in a session realigned to the first volume of the first run) and temporal high-pass filtering with linear trend removal (cut-off frequency of 3 cycles per run). For univariate analyses, functional data were spatially smoothed with a Gaussian filter of 6 mm full width at half maximum (FWHM), in order to reduce noise and minimize intersubject anatomical differences. Functional and structural data were aligned in several steps, using the rigid-body transformation with 6 parameters (3 translations, 3 rotations); the first volume of the first functional
run in a session was coregistered to an anatomical image for the corresponding session; then, anatomical scans obtained in the two sessions with a participant were aligned to each other; finally, functional data from both sessions were coregistered to one of the anatomical images using the transformation parameters obtained during the intersession anatomical alignment. For group analysis, structural and functional data were standardized to the Talairach stereotactic space (Talairach & Tournoux, 1988), using sinc interpolation.

Statistical analyses were performed with a general linear model (GLM), as implemented in BrainVoyager. A trial was modeled as an epoch lasting from the onset to the offset of a picture (2 s). Regressors included predictors of the 10 experimental conditions (2 sessions × 5 conditions: S1_NU, S1_NT, S1_VU, S1_VT, S1_Control, S2_NU, S2_NT, S2_VU, S2_VT, S2_Control). Additionally, 6 parameters resulting from head motion correction were included in the model as regressors of no-interest. Each predictor was convolved with a dual-gamma hemodynamic response function (HRF; Friston et al., 1998). The resulting reference time courses were used to fit the signal time courses in each voxel.

Analyses were performed on the cortical surface, with the help of cortex-based alignment (CBA) as implemented in BrainVoyager. This procedure enables better alignment of structural and functional data across subjects by taking into account individual variability in gyral and sulcal folding patterns. To this end, we segmented the white/gray matter boundary on individual Talairach-transformed T1-weighted anatomical scans and reconstructed 3D hemispheric meshes for each participant. Then we inflated each mesh to a sphere with cortical curvature maps projected onto it (with four coarse-to-fine levels of smoothing) and aligned it to a standard spherical surface using a coarse-to-fine moving target approach (Fischl, Sereno, Tootell, & Dale, 1999; Goebel, Esposito, & Formisano, 2006). The resulting transformation matrices were used to create group-averaged surface meshes for the left and the right hemisphere. Statistical analyses were performed separately for each hemisphere. Thresholded statistical maps obtained for a group were projected onto the group-averaged hemispheric meshes for visualization and described using the CBA-transformed macroanatomical surface atlases supplied with BrainVoyager.

2.5.2.1. Univariate analysis

For the univariate analysis, we created mesh time courses for each run by sampling the spatially smoothed functional data from -1 to 2 mm from the reconstructed white/gray matter boundary. For
the first level analysis (prior to CBA), we ran individual fixed-effects (FFX) GLMs on the subject data collapsed across runs and obtained $t$-statistics for main effects of the experimental conditions. These $t$-maps were subsequently aligned to group-averaged meshes using the aforementioned transformation matrices. At the group level, individual CBA-transformed $t$-maps were submitted to nonparametric permutation testing (Nichols & Holmes, 2002), in combination with Threshold-Free Cluster Enhancement (TFCE; Smith & Nichols, 2009) as implemented in the CoSMoMVPA toolbox (Oosterhof, Connolly, & Haxby, 2016). A total of 1000 Monte Carlo simulations and a corrected cluster threshold $\alpha = .05$ (two-tailed; $z > 1.96$) were used.

2.5.2.2. MVPA

In addition to the standard GLM, we carried out multivariate pattern analysis (MVPA; Haxby et al., 2001). Rather than contrasting the amplitude of the BOLD response in individual voxels/surface vertices, MVPA is based on comparing spatial patterns of activation in response to different experimental conditions (for review, see Haxby, 2012). Specifically, we employed a whole-brain searchlight analysis, a recently developed MVPA technique for identifying locally informative areas of the brain (Etzel, Zacks, & Braver, 2013; Kriegeskorte, Goebel, & Bandettini, 2006), which may outperform mass-univariate analyses due to its greater sensitivity to distributed coding of information (Davis et al., 2014; Jimura & Poldrack, 2012). We performed a searchlight analysis on the brain surface (Oosterhof, Wiestler, Downing, & Diedrichsen, 2011), using a linear discriminant analysis (LDA) classifier as implemented in CoSMoMVPA. We aimed to examine in which areas the classifier could reliably (i.e., significantly above chance) distinguish: (1) nouns vs. verbs (using data from the pre-training fMRI session), and (2) trained vs. untrained items (based on data from the post-training fMRI session). To this end, we ran single-study GLMs separately for each run of each subject, using unsmoothed mesh time courses. At the single-subject level (prior to CBA), maps containing $t$-statistics for the main effects of experimental conditions for each run in a session were stacked together and submitted to the searchlight analysis across the entire cortex using spheres with an 8-mm radius. Classification accuracies were obtained using a leave-one-out cross-validation method with an 8-fold partitioning scheme. The dataset was split into 8 chunks (each corresponding to one experimental run), and the classifier was trained on the data from 7 chunks and tested on the remaining one. The procedure was repeated for 8 iterations, using all possible train/test partitions, and the average decoding accuracies across these iterations were calculated. Decoding accuracies
obtained for a given searchlight were assigned to its central vertex. Individual surface maps containing average decoding accuracies were aligned to the group-averaged mesh using the transformation matrices obtained during CBA. At the group level, a two-tailed one-sample $t$-test across individual maps identified vertices where classification was significantly above chance (50%, since our classifiers were binary). The resulting maps were corrected using TFCE with 1000 Monte Carlo simulations (corrected cluster threshold $\alpha = .05$, two-tailed; $z > 1.96$).

3. Results

3.1. Behavioral results

Average response latencies of the 12 participants in the behavioral experiment are presented in Fig. 2. A three-way repeated-measures ANOVA with session (first, second), word class (noun, verb) and training (trained, untrained) as within-subject factors was carried out. Significant main effects of word class (faster RTs to nouns than verbs; $F(1, 11) = 68.89, p < .001$), training (faster RTs to trained than untrained items; $F(1, 11) = 41.22, p < .001$) and session (faster RTs in session 2 than in session 1; $F(1, 11) = 6.32, p = .029$) were found. The session-by-training interaction was also significant ($F(1, 11) = 10.57, p = .008$)—an expected outcome, considering that prior to training untrained and to-be-trained items were indistinguishable. The lack of significant training-by-word-class ($p = .199$, ns) and session-by-word-class ($p = .285$, ns) interactions suggests the magnitude of training and session effects was similar for words belonging to both classes.
Fig. 2 – Behavioral results. Average reaction times for each stimulus category in session 1 (S1; light gray bars) and session 2 (S2; dark gray bars). NU = untrained nouns, NT = trained nouns, VU = untrained verbs, VT = trained verbs. Error bars reflect standard errors of the mean after removing between-subject variability (Cousineau, 2005); * denotes significant effects at $p_{FDR} < .05$; ** denotes significant effects at $p_{FDR} < .005$.

To examine whether shorter RTs in the second session were driven mainly by session or by training effects, we conducted six paired-samples $t$-tests comparing trained and untrained items belonging to the same word class within and across sessions. In addition, to rule out significant differences between subsets prior to training, we carried out a $t$-test comparing items from the trained and the untrained subset in session 1 for each word class. The resulting $p$-values were corrected using the false discovery rate (FDR) method for the overall number of comparisons ($N = 8$; Benjamini & Yekutieli, 2001). The $t$-tests designed to compare the two subsets of nouns and verbs prior to training failed to distinguish RTs to to-be-trained and not-to-be-trained items (nouns: S1_NT vs. S1_NU: $p = .154$, ns; verbs: S1_VT vs. S1_VU: $p = .611$, ns). Significant effects of training, for both nouns and verbs, were observed when comparing responses to trained items before and after training (S2_NT vs. S1_NT: $p_{FDR} = .026$; S2_VT vs. S1_VT: $p_{FDR} = .01$), and responses to trained and untrained items after training (S2_NT vs. S2_NU: $p_{FDR} = .003$, S2_VT vs. S2_VU: $p_{FDR} = .004$). However, no significant session effects were found for untrained nouns (S2_NU vs. S1_NU: $p_{FDR} =$...
and verbs (S2_VU vs. S1_VU: \( p_{FDR} = .645, \text{ns} \)), suggesting that the significant main effect of session had been actually driven by the training effect.

### 3.2. fMRI results

#### 3.2.1. Word class effects

##### 3.2.1.1. Univariate results

Naming of pictured objects and actions activated similar brain networks (Fig. 3 A), as shown by contrasting nouns/verbs with phase-scrambled controls in the first fMRI session (S1_NU + S1_NT > S1_Control; S1_VU + S1_VT > S1_Control). Activations were observed bilaterally in ventral and lateral occipitotemporal areas, including inferior and middle occipital gyri, fusiform and parahippocampal gyri, i.e., in the ventral visual processing stream involved in object and shape recognition (Goodale & Milner, 1992, 2018; Ungerleider & Mishkin, 1982). They extended bilaterally into the posterior portion of the superior parietal lobule (SPL) and anterior insular cortices. Additionally, picture naming recruited most of the left inferior frontal gyrus (IFG), as well as the ventral precuneus and pre-supplementary motor area (pre-SMA). Detailed information about the areas activated in this and other univariate contrasts is provided in Supplementary Table 1.

Inspection of Fig. 3 A suggests that bilateral activations associated with verb production were more extensive than those associated with noun production in posterior middle temporal gyri (pMTG) and SPL. The direct contrast of responses to nouns and verbs in session 1 (S1_VU + S1_VT > S1_NU + S1_NT; Fig. 3 B) revealed that verb naming engaged to a significantly greater extent the lateral occipitotemporal cortices (LOTC), as well as regions in the left SPL and the intraparietal sulcus. The opposite contrast (S1_NU + S1_NT > S1_VU + S1_VT) detected a stronger BOLD response for nouns than verbs only in a small cluster in the posterior portion of the medial fusiform/extrastriate cortex.
Fig. 3 – Word class effects, revealed by the univariate RFX GLM contrasts of stimuli presented in the pre-training fMRI session. **A.** Object (blue) and action (red/orange) naming networks, identified by contrasting all nouns/verbs with meaningless scrambled images (negative tail not shown). The logical conjunction of the two maps (i.e., areas recruited both during object and action naming) is indicated in purple. **B.** Areas showing increased BOLD response to verbs as compared to nouns (red/orange) and to nouns as compared to verbs (blue). The statistical group maps ($N = 20$) for each hemisphere were corrected for multiple comparisons using TFCE ($\alpha = .05$, two-tailed) and projected onto the group-averaged surface meshes for visualization.
3.2.1.2. MVPA results

We trained the classifier on data from the pre-training fMRI session to identify the areas in which it would reliably distinguish nouns and verbs. The whole-brain searchlight analysis (Fig. 4) showed that patterns of t-scores for nouns and verbs were decodable in a number of bilateral areas, including the LOTC and the SPL, confirming findings of the univariate analysis. Nouns and verbs were also distinguishable on the basis of their patterns of activation in ventral occipitotemporal cortices, early visual areas and the precuneus. Finally, we were able to decode nouns and verbs in the left IFG, extending dorsally into the middle frontal gyrus and caudally into the premotor cortex. Detailed information about the areas identified by the multivariate analyses is provided in Supplementary Table 2.

![Fig. 4 – Decoding of word classes (nouns vs. verbs), based on the pre-training fMRI session. A. Mean accuracy maps of the searchlight MVPA. Individual accuracy maps (N = 20) were averaged and projected onto a flattened group-averaged hemispheric surfaces. Chance accuracy is 50%. B. Statistical group maps, corrected using TFCE at α = .05 (two-tailed).](https://doi.org/10.1101/590026)
3.2.2. Training effects

3.2.2.1. Univariate results

To identify the neural correlates of intensive naming practice, we compared responses to trained and untrained items. No significant differences were found when contrasting these items in the first, pre-training fMRI session (S1_NT > S1_NU; S1_VT > S1_VU), attesting to the fact that the two subsets of nouns and the two subsets of verbs had been matched for the relevant variables. In the second, post-training fMRI session, contrasting the same items (S2_NT > S2_NU; S2_VT > S2_VU) revealed significant BOLD amplitude changes in several brain regions (Fig. 5 A). When compared to untrained items in the same post-training session, both trained nouns (Fig. 5 A, left panel) and trained verbs (Fig. 5 A, right panel) yielded a significantly reduced BOLD response in anterior regions of the left hemisphere, including the posterior IFG (pars opercularis and pars triangularis) and the adjacent frontal operculum/anterior insula. Even though deactivations seemed more extensive for verbs than for nouns, the compound contrast (S2_VT > S2_VU) > (S2_NT > S2_NU) failed to reach significance.
Fig. 5 – Training effects, revealed by the univariate RFX GLM analysis (N = 20), projected onto the group-averaged mesh of the left hemisphere. **A.** Areas showing significant deactivation following training of nouns (left panel) and verbs (right panel), as evidenced by contrasting words from trained and untrained subsets in the post-training fMRI session (corrected using TFCE at α = .05, two-tailed). **B.** Comparison of training and session effects across the two fMRI sessions (corrected using TFCE at α = .05, one-tailed), separately for nouns (left panel) and for verbs (right panel).

Additionally, we examined the effects of incidental word repetition (i.e., session effects), in order to subsequently compare them with the effects of explicit naming practice (i.e., training effects). By contrasting untrained items in the post- and pre-training sessions (Supplementary Fig. 1), we found that the mere exposure to the same stimuli and the same task (in the same scanner environment) twice over the course of two weeks yielded significant repetition suppression in early visual areas and the fusiform, both for nouns (S2_NU > S1_NU; Supplementary Fig. 1 A) and for verbs (S2_VU
> S1_VU; Supplementary Fig. 1 B), in line with reports on priming of low-level features and amodal structural representations respectively (for reviews, see Henson, 2003; Schacter & Buckner, 1998). Decreased activity in the SPL (bilateral for nouns, right-lateralized for verbs) was not expected, but could be explained by facilitated visuospatial processing of familiar stimuli (Beauchamp, Petit, Ellmore, Ingeholm, & Haxby, 2001; Corbetta & Shulman, 1998; Nobre et al., 1997). Additionally, a significantly reduced BOLD response for nouns was observed in the left posterior superior frontal gyrus (on the lateral surface, adjacent to the precentral gyrus). The compound contrast (S2_VU > S1_VU) > (S2_NU > S1_NU) failed to reveal significantly different session effects for the two word classes.

Finally, we directly compared the across-session training and session effects (Fig. 5 B). Since we hypothesized that the BOLD amplitude would decrease more for trained items in the areas associated with explicit practice, we used a one-tailed \( t \)-test \( (z < -1.65) \). Indeed, verb training was associated with significantly stronger decrease of the BOLD signal in the left posterior IFG, including a cluster encompassing most of the pars opercularis and a portion of the pars triangularis, as well as clusters in the pars orbitalis and the anterior insula ((S2_VT > S1_VT) > (S2_VU > S1_VU); Fig. 5 B, right panel). An analogous contrast for nouns ((S2_NT > S1_NT) > (S2_NU > S1_NU); Fig. 5 B, left panel) revealed greater training-related reductions of the BOLD response in the left pars triangularis and the adjacent frontal operculum.

3.2.2.2. MVPA results

To further localize areas sensitive to training, we performed searchlight pattern classification analysis on data from the post-training fMRI session. As a first step, the two classifiers learned to distinguish between trained and untrained items, separately for nouns and verbs. Average decoding accuracies did not go beyond chance in any brain region, possibly due to insufficient statistical power. In order to increase the power, the data were collapsed across word classes and a binary classifier was trained to distinguish between trained and untrained items, irrespective of word class. Results (Fig. 6) show that several areas in the left hemisphere (posterior superior temporal sulcus, angular gyrus, precuneus) were sensitive to training. Decoding was also significantly above chance in the left anterior insula (replicating the univariate results) and in two small clusters close to the right calcarine sulcus.
Fig. 6 – Decoding of trained vs. untrained items, based on the data from the post-training fMRI session. **A.** Mean accuracy maps of the searchlight MVPA. Individual accuracy maps (N = 20) were averaged and projected onto a flattened group-averaged surface. Decoding accuracy at chance is 50%. **B.** Statistical group maps, corrected using TFCE at α = .05 (two-tailed).
4. Discussion

We examined the mechanisms underlying naming practice of object and action pictures in healthy participants. We found that, although naming of objects and actions evoked significantly different responses in temporal, parietal and frontal regions, training effects for the two word classes were statistically indistinguishable. Below we will discuss these observations in more detail.

4.1. Word class effects

Verbs were produced significantly slower than nouns, in line with the previous reports (Kurland et al., 2018; Vigliocco, Vinson, Lewis, & Garrett, 2004). Converging evidence from univariate and multivariate fMRI analyses points to the bilateral LOTC and predominantly left parietal regions as the potential neural loci of word class effects.

Although the lateral occipital regions surrounding the middle temporal (MT) area are primarily associated with perceptual processing of basic motion (Lingnau, Ashida, Wall, & Smith, 2009; Wall, Lingnau, Ashida, & Smith, 2008), they are also sensitive to motion implied by static images (e.g., Kable, Lease-Spellmeyer, & Chatterjee, 2002; Kourtzi & Kanwisher, 2000), accounting for why these areas may be activated by action drawings. It has recently been proposed that more anterior, temporal regions of the LOTC store conceptual, rather than perceptual, representations of actions (for review, see Lingnau & Downing, 2015). Several findings support this claim. First, stronger activation for verbs than nouns occurs in the lateral posterior temporal regions in response not only to pictures/videos, but also to written and spoken words (Kable, Kan, Wilson, Thompson-Schill, & Chatterjee, 2005; Kable et al., 2002; Papeo & Lingnau, 2015). Second, action-related information in the anterior LOTC does not rely on visual experience, as it is sensitive to both motion verbs and mental state verbs (Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Bedny, Caramazza, Pascual-Leone, & Saxe, 2012), and is activated during action processing in congenitally blind individuals (Bedny et al., 2012). Third, experiments employing MVPA successfully decoded representations of observed actions in portions of the LOTC anterior to MT that generalize across object exemplars and motion kinematics (Wurm, Ariani, Greenlee, & Lingnau, 2016; Wurm & Lingnau, 2015).
Notably, verb-preferring LOTC activations in our experiment extended into the mid-portion of the left middle/superior temporal cortex, a region implicated in retrieval of lexical and grammatical (rather than conceptual/semantic) information about verbs (Crepaldi et al., 2011; Willms et al., 2011). While the conceptual and linguistic accounts cannot be teased apart in the context of our experiment (since verbs referred to actions and nouns referred to objects), studies investigating the nature of verb-related LOTC activations (Bedny, Dravida, & Saxe, 2014; Peelen, Romagno, & Caramazza, 2012) confirm that whereas posterior lateral temporal regions store conceptual representations of actions, a distinct cluster in the left pMTG may be specialized for processing verbs as a grammatical class.

Several clusters in the posterior parietal lobe (on the left in the univariate analysis; bilateral, but mainly on the left in the searchlight analysis) were recruited during action, as compared to object naming. The parietal cortex is the endpoint of the dorsal visual processing stream (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982), mediating spatial perception and visually-guided reaching and grasping movements. In addition to action planning and execution, it is involved in action-related perceptual tasks, such as action observation, and in processing of tools and other action attributes (for review, see Culham & Valyear, 2006). Activation of the posterior parietal cortex has also been reported in language tasks, including verb production (Marangolo, Piras, Galati, & Burani, 2006; Saccuman et al., 2006; Shapiro, Moo, & Caramazza, 2006; Warburton et al., 1996) and comprehension (Tsigka, Papadelis, Braun, & Miceli, 2014). The linguistic nature of parietal activations in our experiment is indirectly supported by their being left-lateralized. While there is no consensus on the role of parietal regions in verb processing, recent studies suggest that they may be crucial for thematic role assignment (Finocchiaro, Capasso, Cattaneo, Zuanazzi, & Miceli, 2015; Thothathiri, Kimberg, & Schwartz, 2012). On an alternative view, increased parietal activations during action naming may be explained by greater complexity of action drawings that for transitive verbs included not only agents, but also undergoers and instruments (Liljeström, Hultén, Parkkonen, & Salmelin, 2009; Liljeström et al., 2008).

In addition, the searchlight analysis revealed that nouns and verbs showed significantly different activation patterns in virtually all left inferior frontal regions engaged in picture naming. The association of verb processing with left IFG activation has a long-standing tradition, corroborated by neuropsychological findings. The fronto-temporal dichotomy hypothesis (FTDH; Damasio & Tranel, 1993) linked selective verb impairment to lesions in the left posterior inferior frontal and precentral gyri, and selective noun retrieval deficits to middle and inferior temporal damage.
Consistent evidence was reported in lesion studies (Bak, O’Donovan, Xuereb, Boniface, & Hodges, 2001; Daniele, Giustolisi, Silveri, Colosimo, & Gainotti, 1994; Shapiro & Caramazza, 2003; Tranel, Adolphs, Damasio, & Damasio, 2001; for review, see Cappa & Perani, 2003), cortical mapping in glioma patients (Havas et al., 2015) and neuroimaging experiments (Bedny et al., 2008; Berlingeri et al., 2008; Palti, Ben Shachar, Hendler, & Hadar, 2007; Perani et al., 1999; Tyler, Bright, Fletcher, & Stamatakis, 2004; Tyler et al., 2003). However, evidence from some reports is at odds with the FTDH (e.g., Aggujaro et al., 2006; De Renzi & di Pellegrino, 1995; Silveri & Di Betta, 1997; Silveri, Perri, & Cappa, 2003; Tranel et al., 2008), suggesting that noun/verb dissociations result not from a single factor, but rather from disruption at different levels of word processing, including semantic, phonological and syntactic (Black & Chiat, 2003; Cappa & Perani, 2003). According to an alternative account, activation differences in the verb vs. noun contrast are incidental and are attributable to the difference in morphosyntactic and task demands, rather than the presence of any verb-specific linguistic information in this area (Berlingeri et al., 2008; Siri et al., 2008; Vigliocco et al., 2011). On this view, increased prefrontal response to verbs as compared to nouns in the context of a picture naming task (and during word production in general) may reflect differences in lexical selection demands, as verbs typically have more synonyms/hyponyms/hyperonyms than nouns, and thus place more load on selection processes (Kan & Thompson-Schill, 2004).

Using MVPA, we were able to reliably decode nouns and verbs in posterior ventral occipitotemporal cortices bilaterally, consistent with the role attributed to this area in representation of object concepts (for review, see Martin, 2007). The univariate analysis detected only a small noun-prefering cluster in the left posterior fusiform, which may have to do with the fact that our stimuli were selected from a wide range of semantic categories whose conceptual representations are distributed along the ventral occipitotemporal cortex.

4.2. Training effects

Naming latencies were significantly reduced for trained nouns and verbs, both when comparing trained and untrained items in the post-training session and when contrasting the same items from each trained subset across sessions. At the neural level, intensive training for ten days yielded similar activation changes for nouns and verbs, that were separable from the effects of incidental naming practice and involved both anterior and posterior regions of the left hemisphere.
4.2.1. Training-related activation decreases in left anterior regions

Training of nouns and verbs was associated with significant decreases of the BOLD response in the left IFG and anterior insula, replicating findings of studies on repeated object naming (Basso et al., 2013; MacDonald et al., 2015; Meister et al., 2005; Meltzer et al., 2009; van Turennout et al., 2003, 2000). Notably, activation decreases in these regions were significantly greater after the explicit training as compared to a single word repetition over the course of two weeks, attesting to the cumulative nature of practice.

Activation of the left posterior IFG during naming tasks has been ascribed to a number of linguistic functions, including both phonological and semantic processing (Poldrack et al., 1999; Vigneau et al., 2006). Although the practice-related decreases of the BOLD signal in this region may be attributed to facilitation at any level(s) of language processing, they may also be explained by decreased reliance on executive mechanisms, such as response selection and inhibition of competing responses (Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997; Thompson-Schill, D’Esposito, & Kan, 1999). Indeed, subjects were encouraged to settle on target words from the beginning and stick to them throughout the training. This may have increased name agreement, thus decreasing left prefrontal activation (Kan & Thompson-Schill, 2004).

Activation decreases in the left anterior insula support findings of Basso et al. (2013), who compared the BOLD responses to low-frequency nouns before and after training with response to high-frequency nouns that were not involved in practice. While prior to training low-frequency items yielded greater activations in the insula bilaterally, in the post-training session they were indistinguishable from high-frequency items. Hence, training effects in the left anterior insula may mimic the frequency effects, as the usage frequency of trained items was manipulated by intensive repetition. Supporting evidence comes from previous studies implicating insula in processing of low-frequency words (Binder, Medler, Desai, Conant, & Liebenthal, 2005; Carreiras, Mechelli, & Price, 2006; de Zubicaray, McMahon, Eastburn, Finnigan, & Humphreys, 2005; Graves, Grabowski, Mehta, & Gordon, 2007).
4.2.2. Training-related activation changes in left temporal and parietal regions

MVPA distinguished activation patterns associated with naming of trained and untrained words in several posterior regions of the left hemisphere, including the precuneus, angular gyrus and posterior superior temporal sulcus (pSTS).

As discussed above, posterior lateral temporal cortices are implicated in storage of conceptual and lexical representations. Specifically, activation in pSTS is often attributed to processing of lexical word forms (Hickok & Poeppel, 2007; Indefrey & Levelt, 2004). In our study, activation changes in this region might reflect facilitated access to phonological representations of the trained words. Interestingly, sparing of this region was the sole predictor of anomia recovery in a lesion-symptom mapping study (Fridriksson, 2010; Supplementary Fig. 2, blue sphere), attesting to its role in word retrieval. Notably, as shown by diffusion-weighted imaging (DWI) studies in healthy subjects, posterior lateral temporal regions are extensively connected to the eloquent areas in the left IFG by the arcuate fasciculus, both directly and indirectly, via the inferior parietal lobule (Catani, Jones, & ffytche, 2005).

Practice-related increase of the BOLD signal in the precuneus has been reported in several studies of repeated object naming in healthy populations (Basso et al., 2013; Kurland et al., 2018; MacDonald et al., 2015) and in patients with aphasia (Fridriksson, 2010; Fridriksson et al., 2007; Heath et al., 2015). Increased response of the precuneus and the angular gyrus was also reported following sentence repetition (Hasson, Nusbaum, & Small, 2006; Poppenk, McIntosh, & Moscovitch, 2016). Kurland et al. (2018) documented practice effects following several repetitions of nouns and verbs in a portion of the inferior parietal lobule, closely overlapping with the angular gyrus cluster identified by our study (Supplementary Fig. 2, yellow sphere). The above-mentioned studies linked activation of the parietal regions to the explicit memory of practiced items. Evidence consistent with this view was reported by Schott et al. (2005), who found that only conscious recognition of previously studied items (but not priming in the absence of explicit memory) yielded increased response in the precuneus and the inferior parietal lobule.

Whereas the role of the precuneus in mediating episodic memory has been long established, the supporting role of the angular gyrus in this process was highlighted relatively recently (Seghier, 2013; Yazar, Bergström, & Simons, 2012). Importantly, fibers from the angular gyrus project both to the domain-general regions implicated in long-term memory, including the precuneus, and to the IFG (Seghier, 2013), which makes this area well-suited to mediate language learning.
While at first glance training-related changes outside of the classic language circuit may seem surprising, our results map well onto the studies in anomic patients that report a positive relationship between treatment-induced naming improvement and modulation of activity in the lateral and medial parietal cortices (Fridriksson, 2010; Fridriksson et al., 2007; Menke et al., 2009). Although functional changes in parietal regions did not receive much attention in the clinical literature (that mostly focuses on the perisylvian areas and their right-hemispheric homologues), they may support naming recovery in some patients with chronic aphasia. Thus, the intactness and potential functional reorganization of these areas following practice should be considered in naming treatment studies with patients.
**Funding**

Funding for this work was provided by the European Commission within the action 2014—0685/001-001-EMJD, Framework Partnership Agreement 2012-2025 (to ED), and by the Fondazione CaRiTRo (to GM). AL is supported by a grant from the German Research Foundation (Heisenberg-Professorship, Li 2840/2-1).

**Disclosures**

The authors declare no competing financial interests.

**Acknowledgements**

The authors thank UNITN students Irene Graziosi and Mariapia Piccinini for assistance with translation and subject recruitment.
References


Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of


