Research Articles: Behavioral/Cognitive

Converging evidence for differential specialisation and plasticity of language systems

https://doi.org/10.1523/JNEUROSCI.0851-20.2020

Cite as: J. Neurosci 2020; 10.1523/JNEUROSCI.0851-20.2020
Received: 11 April 2020
Revised: 23 September 2020
Accepted: 24 September 2020

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

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

Copyright © 2020 Gurunandan et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.
Converging evidence for differential specialisation and plasticity of language systems

Abbreviated title: Specialisation and plasticity of language systems

Kshipra Gurunandan*,1, Jaione Arnaez-Telleria1, Manuel Carreiras1,2,3 and Pedro M. Paz-Alonso*

1 BCBL Basque Center on Cognition, Brain and Language, 20009 Donostia-San Sebastian, Spain
2 IKERBASQUE Basque Foundation for Science, 48013 Bilbao, Spain
3 Department of Basque Language and Communication, EHU/UPV, 48015 Bilbao, Spain

*Correspondence should be addressed to: Kshipra Gurunandan, k.gurunandan@bcbl.eu and Pedro M. Paz-Alonso, p.pazalonso@bcbl.eu

K.G., J.A.-T., M.C. and P.M.P-A. designed research; K.G. and J.A.-T. performed research; K.G. and P.M.P-A. analysed data and wrote the paper; M.C. provided comments on the manuscript.

The manuscript includes 5 figures and 1 table.

Number of words: Abstract (221 words), Introduction (676 words), Discussion (1961 words).

The authors declare no competing interests.

Kshipra Gurunandan received support from “la Caixa” Foundation (ID 100010434) through the fellowship LCF/BQ/II17/11620005 and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 713673. Jaione Arnaez-Telleria was supported by a Basque Government predoctoral grant [PRE_2015_1_028]. Manuel Carreiras was supported by project APCIN-2015-061-MultiLateral funded by the Spanish Ministry of Economy and Competitiveness (MINECO) [FLAG-ERA JTC 2015]. Pedro Paz-Alonso was supported by grants from MINECO [RYC-2014-15440, PGC2018-093408-B-I00], and the Neuroscience Research Projects programme from the Fundacion Tatiana Perez de Guzman el Bueno. The research was also supported by funding from the Basque Government [BERC 2018–2021] and the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation [SEV-2015-0490]. We thank Oihana Vadillo and Diego López-Zuazo for their assistance with data collection.
Abstract

Functional specialisation and plasticity are fundamental organising principles of the brain. Since the mid-1800s, certain cognitive functions have been known to be lateralised, but the provenance and flexibility of hemispheric specialisation remain open questions. Language is a uniquely human phenomenon that requires a delicate balance between neural specialisation and plasticity, and language learning offers the perfect window to study these principles in the human brain. In the current study, we conducted two separate functional MRI experiments with male and female language learners, one cross-sectional and one longitudinal, involving distinct populations and languages, and examined hemispheric lateralisation and learning-dependent plasticity of three language systems: reading, speech comprehension and verbal production. A multi-pronged analytic approach revealed a highly consistent pattern of results across the two experiments, showing that (i) in both native and non-native languages, while language production was left-lateralised, lateralisation for language comprehension was highly variable across individuals, and (ii) with increasing non-native language proficiency, reading and speech comprehension displayed substantial changes in hemispheric dominance, with languages tending to lateralise to opposite hemispheres; while production showed negligible change and remained left-lateralised. These convergent results shed light on the long-standing debate of neural organisation of language by establishing robust principles of lateralisation and plasticity of the main language systems. Findings further suggest involvement of the sensorimotor systems in language lateralisation and its plasticity.
Significance Statement

The human brain exhibits a remarkable ability to support a vast variety of languages that may be acquired at different points in the lifespan. Language is a complex construct involving linguistic as well as visual, auditory, and motor processes. Using functional MRI, we examined hemispheric specialisation and learning-dependent plasticity of three language systems — reading, speech comprehension and verbal production — in cross-sectional and longitudinal experiments in language learners. A multi-pronged analytic approach revealed converging evidence for striking differences in hemispheric specialisation and plasticity between the language systems. The results have major theoretical and practical implications for our understanding of fundamental principles of neural organisation of language, language testing and recovery in patients, and language learning in healthy populations.
Introduction

Functional specialisation in the brain is a well-established principle of neural organisation, but studies of atypical development suggest dramatic potential for neural plasticity (Payne and Lomber 2001; Bavelier and Neville 2002). While the capacity for neural reorganisation decreases with age, it does not disappear completely, and adult neural plasticity is essential for learning and maintaining new information or behaviours (Kleim and Jones 2008). The human propensity for language requires a delicate balance between neural specialisation and capacity for re-organisation, making language learning the ideal candidate for examination of specialisation and plasticity in the human brain.

Language typically activates a fronto-temporo-parietal network (Skeide and Friederici, 2016; Hagoort, 2019), and has long been thought to be predominantly left-lateralised (Broca, 1863; Dax, 1863). However, the right hemisphere appears to be capable of taking over or supporting language function if needed, as seen in cases of language recovery after left-hemisphere damage (Papanicolaou et al., 1987; Boatman et al., 1999; Duffau et al., 2002, 2003; Hope et al., 2017) and in language learning (Vingerhoets et al., 2003; Park et al., 2012). It is thus unclear whether the left hemisphere is indeed specialised for language as is broadly accepted, with the right hemisphere playing at best a supporting role (Vigneau et al., 2010), or whether hemispheric dominance is more variable across individuals, as suggested by the larger than expected prevalence of language deficits following right hemisphere brain surgery (Vilasboas et al., 2017).

Language is a complex construct involving multi-level representations that can be processed visually (reading), auditorily (listening) or by motor production (speaking/writing), and cumulative evidence points to these functions lateralising differently. Auditory language has been found to be bilateral in infants (Dehaene-Lambertz et al., 2002; Perani et al., 2011), with either no increase in lateralisation from childhood to adulthood (Lidzba et al., 2011), increasing left-lateralisation (Ahmad et al., 2003), or increasing right-hemisphere involvement (Booth et al.,
2000), and a meta-analysis of auditory comprehension studies suggested that any left-lateralisation from childhood to adulthood increases more slightly and gradually than previously thought (Enge et al., 2020). On the other hand, there is little evidence to suggest that language production is anything but left-lateralised (Gaillard et al., 2003; Szaflarski et al., 2006; Lidzba et al., 2011).

Language learning is known to change the pattern of neural activation for language. Studies comparing bilinguals and monolinguals consistently find differences in activation between them, with bilinguals typically exhibiting greater right hemispheric involvement in comprehension tasks (e.g. Kovelman et al., 2008; Horowitz-Kraus et al., 2015). However, it is uncertain whether this increased right hemispheric involvement merely modulates the magnitude of left-lateralisation, or whether it is significant enough to constitute a change in hemispheric dominance. Further, are differences in lateralisation between monolinguals and bilinguals due to developmental differences or is hemispheric dominance in fact plastic even into adulthood? Few neuroimaging studies have looked into ecologically-valid adult language learning, but findings indicate that language learning in adults involves structural changes in cortical thickness and connectivity that could indeed support shifts in lateralisation (Mårtensson et al., 2012; Schlegel et al., 2012; Xiang et al., 2015), suggesting that lateralisation, at least for comprehension, may be susceptible to learning-dependent changes.

We conducted two fMRI experiments, one cross-sectional and one longitudinal, with immersed late language learners, and examined lateralisation of reading, speech comprehension, and verbal production in their native (L1) and non-native (Ln) languages, and how this changed with increasing Ln proficiency. To test both replicability and generalisability of findings, the two experiments were contrasted on several factors such as early language experience of the participants (monolingual vs bilingual) and the language currently being learnt, and the L1-Ln pairs in the two experiments had contrasting degrees of overlap in language families, phonology, and orthography. We hypothesised that (i) lateralisation of comprehension would be more
variable across individuals but production would be left-lateralised, and (ii) with increasing
language proficiency, comprehension may display changes in hemispheric dominance, while
production would remain left-lateralised. We further expected that L1-Ln associations would
change with increasing Ln proficiency, and that the pattern of changes would differ across the
language systems.

Materials and Methods

Participants

Experiment I: basic vs advanced level language learners (cross-sectional)

The final experiment sample consisted of 29 right-handed native Spanish adults (mean age=43.7 ± 9.7 years; 15 female) studying Basque in the same language school at either the basic (A2 level, n=14) or advanced level (C1 level, n=15). The proficiency levels correspond to those specified by the Common European Framework of Reference for Languages (CEFR). Participants were from the Basque Country, Spain; they grew up primarily exposed to Spanish at home and in school, with little early Basque exposure, and had limited knowledge of English or other languages (no difference between groups, p=0.83). The two groups of learners were matched on age, gender, IQ and Spanish proficiency (Table 1). Data from 5 other participants was discarded due to excessive head motion during MRI scanning and these were not counted in the final sample.

Experiment II: intermediate language learners (longitudinal)

The final experimental group consisted of 19 right-handed native Spanish adolescents (mean age=17.2 ± 0.6 years; 16 female) taking part in a 3-month English immersion-style after-school programme for B1 level students. Participants were from the Basque Country, Spain; they were native speakers of Spanish and acquired Basque in school (AoA=2.6 ± 2.06 years).
medium of instruction in school was Spanish/Basque; English was learnt as a foreign language, with little exposure outside of classes. The students had intermediate English proficiency (Table 1). Data from 5 other participants was discarded due to excessive head motion during MRI scanning and these were not counted in the final sample.

Experiments I & II

In both experiments, language proficiency was assessed using picture-naming tasks — an adaptation of the Boston Naming Test (Kaplan et al., 1983) controlled for cognates across Spanish, Basque and English. Participant groups in Experiment I differed significantly in their Basque proficiency, and participants in Experiment II exhibited significant increase in English proficiency after language training (Table 1). All participants had normal or corrected-to-normal vision, and no history of neurological or psychiatric disorders. In compliance with the ethical regulations established by the BCBL Ethics Committee and the guidelines of the Helsinki Declaration, all participants gave written informed consent prior to taking part in the experiment, and received monetary compensation for their participation.

fMRI Task

Inside the MRI scanner, participants performed two tasks: a comprehension and a production task. The order of tasks was counter-balanced across participants.

Language Comprehension Task

The participants performed semantic animacy judgement (living/non-living) with single-word text and speech stimuli in each of their languages. Participants were instructed to fixate on a white cross in the middle of a black screen, and on presentation of stimuli, to indicate their responses as quickly and as accurately as possible via button presses (counter-balanced across
participants) using their dominant (right) hand. Stimuli were high frequency, concrete, imageable nouns (e.g. house, dog, table) with an even split between living and non-living items. Visual stimuli were presented in white letters on a black screen and were 5-8 letters long. Auditory stimuli were presented through headphones and lasted an average of 565 ms (s.d.=86 ms). Each run had 48 stimuli with inter-mixed reading and listening trials. The fMRI design was event-related with six/four runs (Experiment I: 2 languages x 3 runs; Experiment II: 2 languages x 2 runs). To avoid language-switching, the languages were separated and their order was counterbalanced across participants.

**Language Production Task**

The participants performed a paced form of the semantic verbal fluency task in each language. Participants were instructed to fixate on a white cross in the middle of a black screen and respond overtly to semantic category words (e.g. fruits, animals, clothes) presented on the screen. Each word was displayed eight times, each requiring a novel response, or failing this, an overt response saying “pass” in the relevant language. Fluency was scored as the percentage of valid answers out of eight possible responses for each category. Repetitions, inflections of the same word and erroneous responses were removed, and responses were scored only for correctness and not accent or pronunciation. In the control condition, participants repeated the word presented on the screen. The task had a block design with two runs per language, each run containing eight semantic categories. To avoid language-switching, the languages were separated and their order was counterbalanced across participants.

**MRI Data Acquisition**

Whole-brain MRI data was acquired using a 3-T Siemens Magnetom Trio whole-body MRI scanner and a 32-channel head coil at the Basque Center on Cognition, Brain and Language...
Padded headphones were used to dampen background scanner noise and enable clear transmission of the auditory stimuli. Participants viewed the print stimuli on a screen via a mirror mounted on the head coil. To limit head movement, the head coil was padded with foam and participants were asked to remain as still as possible. Structural T1-weighted images were acquired with a MPRAGE sequence with TR=2530 ms, TE=2.97 ms, inversion time=1100 ms, FA=7°, FoV=256 x 256 mm, 176 slices and voxel size=1 mm³.

Language Comprehension Task

Functional MRI was acquired in the course of six/four separate runs using a gradient-echo echo-planar pulse sequence with the following parameters: TR 2000 ms, TE 30 ms, 32 axial slices with a 3.4 x 3.4 x 4 mm voxel resolution, 0% inter-slice gap, flip angle (FA)=80°, field of view (FoV)=220 mm, 64 x 64 matrix. 186 volumes were collected for each of the functional runs. Prior to each scan, four volumes were discarded to allow for T1-equilibration effects. To improve estimation of the resting baseline in functional analyses, functional runs contained three silent fixation periods of 20 s each. Within each functional run, the order of the trials (reading and listening conditions) and the inter-trial intervals of variable duration corresponding to the baseline MR frames (30% of total collected functional volumes) were determined by an algorithm designed to maximise the efficiency of the recovery of the blood oxygen level dependent response (optseq2, Dale 1999).

Language Production Task

Functional MRI was acquired in the course of four separate runs using a gradient-echo echo-planar pulse sequence with the following parameters: TR 3000 ms, TE 25 ms, 43 axial slices with a 3.0 x 3.0 x 3.0 mm voxel resolution, 10% inter-slice gap, flip angle (FA)=90°, field of view...
(FoV)=192 mm, 64 x 64 matrix. 240 volumes were collected for each of the functional runs. Prior to each scan, four volumes were discarded to allow for T1-equilibration effects.

MRI Data Analysis

Preprocessing

Standard SPM8 (Penny et al., 2011) preprocessing routines and analysis methods were employed. Images were first corrected for differences in timing of slice acquisition and then realigned to the first volume using rigid-body registration. Each subject’s functional volumes were spatially smoothed with a 4-mm full-width half-maximum (FWHM) Gaussian kernel. Next, motion parameters obtained from realignment were used to inform a volume repair procedure (ArtRepair, Mazaika et al., 2009) that identified bad volumes on the basis of scan-to-scan movement (>1 mm) and signal fluctuations in global intensity (>1.3%), and corrected bad volumes via interpolation between the nearest non-repaired scans. Data from subjects requiring more than 20% of volumes to be repaired were discarded. The number of corrected volumes was similar between groups (Experiment I: comprehension task p=0.34, production task p=0.63) and scans (Experiment II: comprehension task p=0.75, production task p=0.46). After volume repair, functional volumes were co-registered to the T1 images using 12-parameter affine transformation and spatially normalised to the MNI space by applying non-linear transforms estimated by deforming the MNI template to each individual’s structural volume (Ashburner and Friston, 2005). During normalisation, the volumes were sampled to 3-mm cubic voxels. The resulting volumes were then spatially smoothed with a 7-mm FWHM Gaussian kernel. Finally, time series were temporally filtered to eliminate contamination from slow frequency drift (high-pass filter with cut-off period of 128 s).
Subject level analyses

Statistical analyses were performed on individual subject data using the general linear model (GLM). fMRI time series data were modelled by a series of impulses convolved with a canonical haemodynamic response function. Six motion parameters for translation (x, y, z) and rotation (yaw, pitch, roll) were included as covariates of non-interest in the GLM. In the event-related-design comprehension task, each trial was modelled as an event, time-locked to the onset of the presentation of each stimulus, and error responses were modelled separately. In the block-design production task, each trial was modelled as an epoch of 31 s each, time-locked to the beginning of the presentation of each block. The remaining functions were used as covariates in the GLM, along with a basic set of cosine functions that high-pass filtered the data, and a covariate for session effects. The least-squares parameter estimates of the height of the best-fitting canonical HRF for each experimental condition were used in pairwise contrasts.

Laterality analyses

For every subject, lateralisation of activation in the classical language network regions was calculated for each task x language. Laterality is typically quantified as a normalised ratio of left and right hemisphere contributions, ranging between +1 (fully left-lateralised activation) and -1 (fully right-lateralised activation). Each subject’s whole-brain t-maps were masked with anatomical language regions from the AAL atlas (Tzourio-Mazoyer et al., 2002) — six bilateral regions from classical language models (Friederici 2012; Hagoort 2013): inferior frontal gyrus (IFG) pars orbitalis, IFG pars triangularis, IFG pars opercularis, superior temporal gyrus (STG), middle temporal gyrus (MTG), and inferior parietal lobule (IPL). Since laterality indices are highly threshold-dependent, in line with the latest recommendations (Bradshaw et al., 2017), a threshold-independent bootstrapping method was used to calculate the laterality index using the LI-toolbox (Wilke and Lidzba, 2007), in which 10,000 indices were iteratively calculated at
different thresholds, yielding a robust mean laterality index. Three analyses were carried out to examine proficiency-dependent plasticity of (i) L1-Ln correlation, (ii) hemispheric dominance, and (iii) modality clustering. L1-Ln correlations were calculated for each group x task, and Cohen’s $q$ was used to quantify the difference in L1-Ln correlation between basic/advanced proficiency and before/after training in each modality. To examine hemispheric dominance, a lateralised dissociation index was calculated such that:

$$\text{Lateralised Dissociation Index} = |L_{L1} - L_{Ln}| \times \text{hem}$$

$$\text{hem} = \begin{cases} 
1 & \text{if opposite laterisation} \\
-1 & \text{if same laterisation}
\end{cases}$$

i.e. the absolute difference between laterality indices for each language and a factor $\text{hem}$ to code whether the two languages were lateralised to the same or opposite hemispheres. Positive values indicated that languages were lateralised to opposite hemispheres, while negative values indicated that the languages were lateralised to the same hemisphere. Cohen’s $d$ was used to measure the magnitude of proficiency-dependent change in each modality: difference between medians in cross-sectional Experiment I, and difference in repeated measures in longitudinal Experiment II. To examine the modality-wise clustering of the joint L1-Ln distribution, 85% data ellipses were plotted for each modality, and the joint distribution difference (JDD) between any two modalities was calculated as:

$$\text{Joint Distribution Difference} = \frac{\text{distance between centroids}}{\text{maximum distance}} \times \frac{\text{angle between major axes}}{\text{maximum angle}}$$

i.e. standardised difference between the bivariate L1-Ln group means and difference between joint spread of the data. This index lies between 0 and 1, with higher values indicating greater
difference between modalities. The maximum Euclidean distance between centroids was considered to be 1 for laterality data, and maximum angle between the axes is 90°. Proficiency-group differences were measured in terms of percentage difference in the difference index.

Results

In-Scanner Behavioural Performance

Experiment I: basic vs advanced level language learners (cross-sectional)

A series of mixed-model analyses of variance (ANOVAs) were conducted on the behavioural measures of the fMRI tasks: comprehension task accuracy, production task fluency, and comprehension task reaction times (Figure 1A). The comprehension task accuracy ANOVA with between-subjects factor Group (basic, advanced) and within-subject factors Language (L1, Ln) and Modality (reading, speech) showed a significant Group x Language interaction (F(1,26)=16.18, p=0.0004). The production task fluency ANOVA with between-subjects factor Group (basic, advanced) and within-subject factor Language (L1, Ln) also showed a significant Group x Language interaction (F(1,23)=31.36, p=0.00001). Post-hoc simple-effect analyses (two-sample t-tests) of these Group x Language interactions showed that the advanced proficiency group had significantly higher Ln task accuracy than the basic proficiency group in both comprehension (t(18.08)=3.20, p=0.002, one-sided) and production (t(22.28)=5.502, p=0.000008, one-sided), but there was no significant difference between groups in L1 task accuracy (comprehension: t(25.48)=−0.93, p=0.360, two-sided, production: t(21.683)=1.03, p=0.31, two-sided). Finally, the ANOVA for comprehension task reaction times showed a main effect of Language, and both groups were significantly slower in their Ln than their L1 (F(1,26)=40.41, p=0.000001).
Experiment II: intermediate language learners (longitudinal)

A series of repeated-measures ANOVAs were conducted on the behavioural measures of the fMRI tasks: comprehension task accuracy, production task fluency, and comprehension task reaction times (Figure 1B). The comprehension task ANOVAs with 3 within-subject factors Training (before, after), Language (L1, Ln), and Modality (reading, speech), showed main effects of Language (L1 > Ln, $F(1,17)=338.64$, $p=0.000000000001$) and Modality (reading > speech, $F(1,17)=30.05$, $p=0.00004$) on task accuracy. The production task fluency ANOVA with 2 within-subject factors Training (before, after), and Language (L1, Ln) showed a main effect of Language (L1 > Ln, $F(1,15)=146.01$, $p=0.0000000004$). The comprehension task reaction times ANOVA revealed a significant Training x Language interaction ($F(1,17)=5.48$, $p=0.031$). Post-hoc simple-effect analyses (paired t-tests) showed that reaction times decreased significantly after training in Ln ($t(17)=2.83$, $p=0.006$, one-sided), but not in L1 ($t(17)=0.21$, $p=0.836$, two-sided).

Language lateralisation

1. Lateralisation in comprehension and production

Laterality indices were calculated for the language network regions in each task and language using a threshold-free method, with values between +1 (left lateralisation) and -1 (right lateralisation). In both experiments, Wilcoxon signed-rank tests of paired samples revealed significant differences between each of the modalities. Comprehension and production displayed robust differences in lateralisation, with significant differences between both reading and verbal production (Experiment I: $W=421$, $p=0.000000000006$, Experiment II: $W=1099$, $p=0.000000004$) as well as between speech comprehension and verbal production (Experiment I: $W=824$, $p=0.000002$, Experiment II: $W=729.5$, $p=0.0000000000006$). Reading and speech comprehension also differed significantly (Experiment I: $W=1998.5$, $p=0.021$, Experiment II: $W=3840$, $p=0.012$). In reading and speech comprehension, lateralisation was highly variable and indices spanned the full range of possible values between the two languages, while verbal production was clearly left-lateralised.
At the group level, comprehension appeared bilateral and production was left-lateralised. This result was consistent across the cross-sectional and longitudinal experiments (Figure 2).

2. Learning-dependent changes in lateralisation

To examine patterns of learning-dependent changes in lateralisation while accounting for the high inter-individual variability across tasks and languages, L1 lateralisation was used as a baseline for each subject’s Ln lateralisation, and the linear association between L1 and Ln was assessed using Pearson’s $r$. In lower proficiency learners, L1 and Ln lateralised similarly, regardless of left/right lateralisation. However, with increasing proficiency, this pattern reversed for comprehension, and L1 and Ln lateralised to opposite hemispheres. This learning-dependent change was not observed in verbal production (Figure 3). Cohen’s $q$ was used to quantify the proficiency-dependent change in L1-Ln correlation for each task, confirming that, across both studies, learning-dependent change in lateralisation was large in reading comprehension, medium in speech comprehension, and small in verbal production.

To examine whether increasing proficiency involved changes in hemispheric dominance for each modality, lateralised dissociation indices were calculated for each subject such that absolute values indicated the magnitude of L1-Ln difference, and direction (i.e. positive or negative) indicated whether the languages were lateralised to same or opposite hemispheres (positive=opposite hemispheres, negative=same hemisphere). There was a significant proficiency-related increase in absolute dissociation between L1 and Ln lateralisation across modalities (Experiment I: Mann-Whitney U tests: across modalities: $W = 584.5$, $p = 0.013$; reading comprehension: $W = 41$, $p = 0.007$; speech comprehension: $W = 88$, $p = 0.579$; verbal production: $W = 53.5$, $p = 0.022$; Experiment II: Wilcoxon signed-rank tests: across modalities: $V = 374.5$, $p = 0.023$; reading comprehension: $V = 15$, $p = 0.004$; speech comprehension: $V = 53$, $p = 0.142$; verbal production: $V = 78$, $p = 0.330$), and Cohen’s $d$ was used to quantify learning-dependent change in
hemispheric dominance for each modality. In both experiments, the same pattern of changes emerged: large in reading comprehension, medium in speech comprehension, and small in verbal production (Figure 4).

Finally, modality-wise clustering of joint L1-Ln lateralisation was plotted using 85% data ellipses to examine overlap between modalities. One-way MANOVAs and a joint distribution difference index were used to test and quantify the separation between: (i) comprehension (both reading and speech) and production (Figure 5-I), and (ii) reading and speech comprehension (Figure 5-II), and the effects of proficiency were tested using non-parametric two-sample/paired tests of difference/change in cluster separation between modalities (Euclidean distance) and quantified with percent change in the JDD. The one-way MANOVA modelled the joint L1-Ln distribution differences between modalities, and the index quantified this difference by taking into account the difference in both bivariate mean and spread of data, with values between 0 (overlapping distributions) and 1 (no similarities). MANOVAs revealed significant differences between comprehension and production (Experiment I: basic proficiency group: F(1.8,65.2)=11.73, p=0.0005; advanced proficiency group: F(1.9,63.8)=22.96, p=0.00000002; Experiment II: before training: F(1.7,73.2)=21.67, p=0.00000002, after training: F(1.7,70.7)=38.94, p=0.000000000004) and with increasing proficiency, comprehension and production dissociated further (Experiment I: the advanced proficiency group displayed 1042.35% greater comprehension-production dissociation than the basic proficiency group, Mann-Whitney U test of group difference in cluster separation: W = 67398, p = 0.000000000003; Experiment II: participants displayed 47.38% increase in comprehension-production dissociation after training, Wilcoxon signed-rank test of post-training change in cluster separation: V = 101769, p = 0.000000000000002). There were no significant differences in L1-Ln joint distribution between reading and speech comprehension (Experiment I: basic proficiency group: F(1.9,45.5)=1.84, p=0.18; advanced proficiency group: F(1.7,41.4)=0.32, p=0.71; Experiment II: Before Training: F(1.9,71.2)=1.98, p=0.15, After Training: F(1.8, 60.5)=2.09, p=0.13), and reading and speech comprehension converged further with
increasing proficiency (Experiment I: the advanced group displayed 87.27% greater comprehension-production overlap than the basic group, Mann-Whitney U test of group difference in cluster separation: $W = 18073$, $p = 0.177$; Experiment II: participants displayed 27.13% increase in comprehension-production overlap after training, Wilcoxon signed-rank test of post-training change in cluster separation: $V = 39306$, $p = 0.0005$).

Discussion

In the present work, we examined hemispheric specialisation and learning-dependent plasticity of the language network concurrently in three language systems: reading, speech comprehension and verbal production. We conducted cross-sectional and longitudinal fMRI experiments in separate populations of immersed language learners. Both experiment samples had the same L1 (Spanish), but were contrasted in other factors: (i) early language experience: monolingual vs sequential bilingual, (ii) language being learnt: Basque vs English, (iii) phonological similarity with native language: high overlap vs low overlap, (iv) orthographic depth: transparent vs opaque. Across these contrasting experimental designs and participant groups, we found a highly consistent pattern of results in both experiments: (i) across native and non-native languages, lateralisation for language comprehension was variable but language production was strongly left-lateralised, and (ii) with increasing non-native language proficiency, reading and speech comprehension displayed significant changes in hemispheric dominance (reading > speech), while verbal production remained left-lateralised. The converging results from separate experiments provide unique insight into the long-standing debate on hemispheric specialisation of language and the effects of language experience (Gainotti, 1993; Price, 1998, 2012; Jung-Beeman, 2005; Hickok and Poeppel, 2007; Friederici, 2012; Hervé et al., 2013; Tzourio-Mazoyer et al., 2016).
The first result showing variably-lateralised (bilateral at the group-level) comprehension and left-lateralised verbal production across different languages suggested that comprehension is flexible while verbal production is hard-wired to be left-lateralised. Previously, conflicting evidence from studies in monolinguals had led to a range of different conclusions and models of comprehension: from left-lateralised to partly-bilateral, bilateral, or right-lateralised function (Booth et al., 2000; Gaillard et al., 2000; Jung-Beeman, 2005; Hickok and Poeppel, 2007; Lidzba et al. 2011). Few studies have compared different modalities in the same participants, and though lateralisation was seen to be highly modality-dependent in the current study, it did not appear to depend on the exact task used, since lateralisation for the single-word overt tasks in the current study was consistent with results from far more complex discourse-level covert tasks in previous studies (Dehaene et al., 1997; Lidzba et al., 2011; Bhattasali et al., 2019). There were also subtle differences between the two experiments, with similar L1 lateralisation but differing central tendencies for L1 laterality. This pattern is consistent with the literature on the influence of early language experience: meta-analyses have found that early bilinguals (L2 acquired before age 6) typically show bilateral hemispheric involvement, while monolinguals and late bilinguals show greater left hemisphere dominance (Hull & Vaid, 2006, 2007; Bloch et al., 2009; Liu & Cao, 2016).

Thus the convergent results in the present work indicate that inter-individual variability in lateralisation for language comprehension is not an artefact of task or methodology, but that instead, language comprehension is differently lateralised across individuals. Lesion studies in children have found dissociative effects of lesion side on comprehension and production: while lesions in the left hemisphere were associated with more severe delays in production compared to comprehension, comprehension delays were more common — but not universal — in children with right-hemisphere damage (Marchman, Miller and Bates, 1991; Thal et al., 1991). In line with these findings, developmental neuroimaging studies all found left-lateralised language production, but reached conflicting conclusions on comprehension, leading to a suggestion of differing maturational mechanisms for comprehension and production (Hervé et al., 2013).
Clinical studies have recommended that both comprehension and production tasks be used in
determining language lateralisation for clinical purposes (Wilke et al., 2010; Lidzba et al., 2011;
Vilasboas et al., 2017; Woodhead et al., 2018). Modality-dependent lateralisation, i.e. variably-
lateralised comprehension vs left-lateralised production, could explain the long-standing
conflicts among previous studies that used tasks tapping into different modalities, and shed new
light on the question of functional specialisation for language.

Our analytic approach to examining learning-dependent changes in language
lateralisation built on the observed inter-individual variability and used within-subject measures
calculated with each subject’s L1 as a baseline for their Ln. We used three measures — L1-Ln
correlation, L1-Ln distance, and modality clustering — and quantified the change within each
language system. These revealed that (i) L1 and Ln were similarly lateralised in lower-proficiency
language learners and tended to dissociate with increasing Ln proficiency, (ii) the change was
largest in reading, smaller in speech comprehension, and smallest in verbal production, and (iii)
with increasing proficiency, comprehension and production dissociated, while reading and
speech comprehension converged.

Convergence and dissociation of neural activation for different languages and language
systems has been of considerable research interest. Neuroimaging studies of language have by
and large come to the conclusion that all languages do indeed recruit the same language regions,
and that language experience modulates the amount of overlap, leading to the “convergence
hypothesis” (Perani and Abutalebi 2005; Gurunandan et al., 2019). The current study built on this
finding, and characterised lateralisation patterns for L1 and Ln within the common language
network, finding that increasing Ln-proficiency led to increasing dissociation in lateralisation
between the two languages. There has been much debate on whether language control in
bilinguals is language-specific or domain-general, with mixed evidence (Hernández et al., 2013),
and it is possible that, apart from any changes in the involvement of language control regions,
the greater hemispheric separation of languages in more proficient non-monolinguals also
contributes to their improved language control. Future studies looking concurrently at dissociation within the language network and recruitment of language control regions are needed to test this idea. Comprehension and production also dissociated with increasing Ln-proficiency. In lower proficiency learners, there was lower separation between modalities, possibly indicating variable strategies of Ln access and variable activation profiles (Dehaene et al., 1997), but as individuals attained higher proficiency, their activation profiles stabilised and became more uniform. Turning to the question of convergence between language systems, print-speech convergence has been considered a universal signature of native language proficiency (Shankweiler et al., 2008; Rueckl et al., 2015; Preston et al., 2016), but it is less well-studied in multilinguals. In the current study, we found increasing convergence of joint L1-Ln lateralisation for reading and speech comprehension with increasing language learning, suggesting that reading-speech convergence is also sensitive to increasing Ln-proficiency.

The pattern of plasticity differences between the language systems, i.e. plasticity for reading > speech comprehension > verbal production, was strikingly similar to their perceived difficulty in real-world language learning in adults. Two observations support the idea that the differential plasticity of language systems contribute to differential learning. First, learners in the longitudinal study had switched languages from same to opposite hemispheres in reading within a relatively short time-frame, while fewer did so for speech comprehension, and none for production. Further, individuals who had L1 and Ln lateralised in opposite hemispheres maintained this dissociation post-training, and individuals who had L1 and Ln in the same hemisphere tended to dissociate post-training to varying degrees depending on the modality. This suggested that opposite hemispheric dominance of languages could be advantageous for language learning, and further, that shifts in hemispheric dominance are limited by the plasticity of the sensory/motor cortices corresponding to each language system. Neuropsychological evidence from stroke recovery patterns in adults who showed greater (but not complete) recovery in comprehension than in production (Lomas and Kertesz 1978), as well as different
reorganisation patterns for comprehension and production (Musso et al., 1999; Heiss and Thiel 2006) further supports our conclusion. Though the visual, auditory and motor cortices are all bilateral, each of them exhibits hemispheric advantages for processing specific features (Benke and Kertesz, 1989; Deruelle and Fagot, 1997; Flinker et al., 2019; Albouy et al., 2020), and previous studies with monolinguals have found differences in visual lateralisation of different writing scripts (Tzeng et al., 1979; Kuo et al., 2001), asymmetrical sensitivity of the auditory cortices (Friederici and Alter 2004; Boemio et al., 2005), and left-lateralisation of auditory and articulatory motor areas (Morillon et al., 2010), pointing to differential potential for post-critical-period plasticity of these sensory/motor regions that matches the converging pattern of language system plasticity found in the current study. Second, the adolescent learners in the second experiment displayed substantial neural changes after just three months of training, while the adult learners in the first experiment displayed similar neural differences for a much larger proficiency difference between groups. This finding is compatible with age-related decrease in neural plasticity, and sheds further light on the source of the difficulty of late language learning. However, despite the convergence of the neural results in experiments I and II, the modest behavioural effect in Experiment II limited any further interpretation of the neural changes in relation to behavioural outcomes at the individual level in naturalistic language learning. In sum, taken together with previous evidence, the converging findings in the present work point to the sensorimotor cortices playing a large role in both the lateralisation of language as well as the asymmetric decrease in plasticity of the language network.

Methodological studies and reviews of language lateralisation have often warned against over-interpretation of results from a single task, small regions of interest, or non-robust analytical methods (e.g. Bradshaw et al., 2017a, 2017b; Bain et al., 2019). These were avoided in the current study and interpretations were based on robust patterns of results verified by corroborating analyses that were replicated in contrasting experiments. However, the current study used classical single word tasks, and while the lateralisation results were consistent with
the findings from far more complex comprehension tasks (Dehaene et al., 1997; Lidzba et al., 2011; Bhattasali et al., 2019), future studies are needed to establish whether the results presented in the current study would be as or possibly even more pronounced in sentence/discourse processing (Hagoort 2019). Further, a priori power analysis was not conducted nor was a replication sample examined. The two experiments involved ecologically valid language learning, and the lateralisation results were sensitive to participants’ real-world language-learning progress, i.e. CEFR level, rather than their performance or improvement on the in-scanner semantic tasks involving high-frequency stimuli. In fact, while performance on the tasks was relatively uniform across participants, lateralisation exhibited much larger variation in both languages, supporting the idea of multi-factorial modulation of hemispheric specialisation (Tzourio-Mazoyer et al., 2016), since participants were carefully selected to control for language backgrounds as much as possible, but actual experimental control on early or previous language exposure was not possible. The replication of findings in language learners at different levels of proficiency suggested that the learning-dependent neural changes were not temporary, but further studies are necessary to disentangle the effects of learning vs proficiency, and test lateralisation of languages in early balanced bilinguals. Finally, the two experiments featured distinct L1-Ln language pairs that were contrasted on factors such as overlap in language families, phonological, and orthographic properties, but did not involve more sensory differences such as visual differences between writing systems or auditory perception and motor production of tones, as in, say, English-Chinese. Following from our idea that the sensorimotor cortices are the limiting factor in language learning and its associated neural changes, it is possible that late acquisition of a language that requires greater sensorimotor learning would entail smaller proficiency-dependent neural changes in lateralisation and the size of the changes would decrease more sharply with increasing age than in the current study.

In conclusion, our study design with cross-sectional and longitudinal experiments in contrasting samples of real-world language learners, testing of different language systems, and a
multi-pronged analytical approach revealed robust and converging patterns of modality-dependent lateralisation and plasticity of the language network. Our findings suggest that language lateralisation for reading and speech comprehension is plastic well into adulthood, while production shows strong left-hemisphere specialisation. Plasticity for reading was greater than for speech comprehension which was in turn greater than for verbal production. Taken together with previous evidence in the literature, we propose that hemispheric specialisation for language may arise from the sensorimotor cortices, and that the differential plasticity of language systems is tied to the plasticity of the associated sensorimotor systems.

References


Dax G (1863) Observations tendant à prouver la coïncidence constante des dérangements de la parole avec une lésion de l'hémisphère gauche du cerveau. C R Acad Sci III 56:536


Hagoort P (2013) MUC (memory unification control) and beyond. Front Psychol 4:416


Tables and Legends

<table>
<thead>
<tr>
<th>Table 1. Participant demographics and linguistic scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment I</strong></td>
</tr>
<tr>
<td>Basic proficiency group</td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>42.9 (10.1)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>7 female, 7 male</td>
</tr>
<tr>
<td><strong>Proficiency</strong></td>
</tr>
<tr>
<td><strong>L1</strong></td>
</tr>
<tr>
<td>99.35 (1.88)</td>
</tr>
<tr>
<td><strong>Ln</strong></td>
</tr>
<tr>
<td>52.6 (14.66)</td>
</tr>
</tbody>
</table>

Note: Values correspond to the mean with standard deviation in parentheses.

**Figure 1.** Behavioural measures Accuracy and Response Time for in-scanner semantic tasks plotted as a function of Group, Language, and Modality in Experiment I (A), and as a function of Training, Language, and Modality in Experiment II (B). Error bars represent standard deviation and asterisks statistically significant differences (*** $p < 0.001$, ** $p<0.01$, * $p<0.05$).
Figure 2. Laterality indices plotted as a function of Group, Language, and Modality in Experiment I (A), and Training, Language, and Modality in Experiment II (B). Laterality indices were obtained from individual whole brain activation in the neuroanatomical language network, and the respective line graphs display mean and standard deviation of laterality indices across participants in each Modality and Language.

Figure 3. Linear associations between L1 and Ln lateralisation indices (Pearson’s r) as a function of Group and Modality in Experiment I (A) and of Training and Modality in Experiment II (B). Cohen’s q quantified the learning-dependent changes in L1-Ln correlation in each Modality.

Figure 4. Lateralised dissociation indices (LDI) as a function of Group and Modality in Experiment I (A), and Training and Modality in Experiment II (B). Positive values indicate that L1 and Ln lateralised to opposite hemispheres, and negative values indicate that L1 and Ln lateralised to the same hemisphere. Cohen’s d quantified the learning-dependent changes in LDI in each Modality.

Figure 5. Modality-wise clustering (I: comprehension versus production, II: reading versus speech comprehension) in joint distributions of L1-Ln lateralisation indices plotted as a function of Group in Experiment I (A) and as a function of Training in Experiment II (B). A Joint Distribution Difference index with values between 0 and 1 quantified overlap in each group, with higher values indicating larger separation between modalities. Asterisks represent statistically significant differences (p < 0.001).