

Short communication

Structural neural predictors of Farsi-English bilingualism

Visar Berisha^{a,b}, Davis Gilton^b, Leslie C. Baxter^c, Steven R. Corman^d, Chris Blais^e, Gene Brewer^e, Scott Ruston^d, B. Hunter Ball^e, Kimberly M. Wingert^e, Beate Peter^{a,f}, Corianne Rogalsky^{a,*}

^a Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287, USA

^b School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA

^c Barrow Neurological Institute and St. Joseph's Medical Center and Hospital, Phoenix, AZ 85013, USA

^d The Hugh Downs School of Human Communication, Arizona State University, Tempe, AZ 85281, USA

^e Department of Psychology, Arizona State University, Tempe, AZ 85287, USA

^f Department of Communication Sciences and Disorders, Saint Louis University, Saint Louis, MO 63101, USA

ARTICLE INFO

Keywords:

Bilingualism

Structural MRI

Decision tree classification

ABSTRACT

The neurobiology of bilingualism is hotly debated. The present study examines whether normalized cortical measurements can be used to reliably classify monolinguals versus bilinguals in a structural MRI dataset of Farsi-English bilinguals and English monolinguals. A decision tree classifier classified bilinguals with an average correct classification rate of 85%, and monolinguals with a rate of 71.4%. The most relevant regions for classification were the right supramarginal gyrus, left inferior temporal gyrus and left inferior frontal gyrus. Larger studies with carefully matched monolingual and bilingual samples are needed to confirm that features of these regions can reliably categorize monolingual and bilingual brains. Nonetheless, the present findings suggest that a single structural MRI scan, analyzed with measures readily available using default procedures in a free open-access software (Freesurfer), can be used to reliably predict an individual's language experience using a decision tree classifier, and that Farsi-English bilingualism implicates regions identified in previous group-level studies of bilingualism in other languages.

1. Introduction

The impact of bilingualism on the brain has been studied for over a century, yet continues to be debated (Paap, Johnson, & Sawi, 2015; Sebastian, Laird, & Kiran et al., 2001). A better understanding of the neural correlates of bilingualism would not only provide insights regarding language acquisition and neural plasticity, but also could provide neural and cognitive targets for enhancing second language learning. The focus of most existing second-language (L2) neuroimaging research has been to identify structural or functional differences in the brains of bilinguals versus monolinguals, particularly as a function of L2 acquisition age or L2 fluency (e.g. Ge et al., 2015; Garcia-Penton, Perez Fernandez, Iturria-Medina, Gillion-Dowens, & Carreiras, 2014; Kovelman, Baker, & Petitto, 2008; Klein, Mok, Chen, & Watkins et al., 2006; Klein et al., 2014; Kim, Relkin, Lee, & Hirsch, 1997; Mahendra, Plante, Magliore, Milman, & Trouard, 2003; Dehaene et al., 1997; Marian, Spivey & Hirsch, 2003). In one of the first neuroimaging studies to examine bilingualism, Kim et al. (1997) compared functional MRI results within left inferior frontal and left posterior superior temporal regions of interest during listening to L1 versus L2 in ten subjects, half

late and half early bilinguals representing several L1 and L2 languages, including English, German, Spanish, French, Turkish, Korean, Chinese and Hebrew. These first fMRI case studies of bilingualism demonstrated that classic frontal and temporal left hemisphere L1 language regions were also engaged by L2, with the overlap between L1 and L2's activations greater for early bilinguals than late bilinguals. Subsequent group studies corroborated Kim et al.'s overall conclusions that regardless of the exact two languages spoken or age of L2 acquisition, L2 generally engages frontal and temporal regions activated by L1, with the degree of spatial separation or amplitude difference in L1 vs. L2 activations correlating with age of L2 acquisition and/or L2 proficiency (Perani & Abutalebi, 2005; Perani et al., 1998; Sakai, Miura, Narafu, & Muraishi, 2004; Wartenburger et al., 2003). However, findings from whole brain analyses also suggest that bilingualism versus monolingualism is associated with structural and functional differences in subcortical regions (Burgaleta, Sanjuan, Ventura-Campos, Sebastian-Galles, & Avila, 2016; Stocco & Prat, 2014), supplemental and pre-supplemental motor areas (Luk, Green, Abutalebi, & Grady, 2011; Rodriguez-Fornells et al., 2005), frontal-parietal cortex (Mechelli et al., 2004; Reiterer et al., 2011; Yokoyama et al., 2006), the cerebellum

* Corresponding author at: Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, AZ 85287-0102, USA.
E-mail address: Corianne.rogalsky@asu.edu (C. Rogalsky).

(Klein et al., 2006; Halsband, 2006), and right hemisphere regions (Mayima, Richards, Coe, Eichler, & Kuhl, 2016; Reiterer et al., 2011; Schlegel, Rudelson, & Tse, 2012). A wealth of electrophysiological work has also identified neural markers of bilingualism and L2 proficiency, including more native-like P600 and N400 responses to L2 morpho-syntactic and grammatical violations as a function of L2 proficiency (Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013).

It has been suggested that the structural and functional neural differences between monolinguals and bilinguals may be related to differences in cognition as a result of bilingualism (Stocco & Prat, 2014), but a causative link between bilingualism and cognitive differences remains controversial (Bialystok, Craik, & Luk, 2012; cf. Paap & Greenberg, 2013). Nonetheless, there are several studies implicating the basal ganglia, particularly the caudate, with bilingualism and particularly with language switching and control (Klein, Milner, Zatorre, Meyer, & Evans, 1995; Luk et al., 2011; Price, Green, & Von Studnitz, 1999; Wang, Wang, Jiang, Wang, & Wu, 2013). Basal ganglia differences between monolinguals and bilinguals (and between low and high proficiency bilinguals) are present during language tasks (Grogan, Green, Ali, Crinion, & Price, 2009; Zou, Ding, Abutalebi, Shu, & Peng, 2012) as well as during some cognitive tasks requiring attentional control (Stocco & Prat, 2014; Grundy, Anderson, & Bialystok, 2017). Voluntary language switching in bilinguals also implicates the supplementary motor area (SMA) and pre-SMA (de Bruin et al., 2014; Luk et al., 2011), both of which also are involved in L2 fluency (Grogan et al., 2009) as well as speech production and motor planning for speech more broadly (Price, 2012; Segawa, Tourville, Beal, & Guenther, 2015).

Most previous studies of the neurobiology of bilingualism present group-level averaged results or case studies (see above). It is well-established, however, that group-level neuroimaging results alone often do not represent an individual's neural response to language; there is substantial individual variability regarding the location and extent of neural responses to language within known language areas such as Broca's area (Nieto-Castanon & Fedorenko, 2012; Rogalsky, Almeida, Sprouse, & Hickok, 2015). Thus it remains unclear what neural features would predict bilingualism in any given individual. Dehaene et al. (1997) noted that "late second language acquisition is not necessarily associated with a reproducible biological substrate." For example, Dehaene et al.'s group-level fMRI results in French-English bilinguals identified significant activations of the right hemisphere to L2, but also demonstrated that individual subjects show a great deal of variability in right-lateralization of responses to L2, ranging from none to right hemisphere dominance.

In the present study we examined an existing structural MRI dataset of Farsi-English bilinguals and English monolinguals to determine whether volumetric and curvature differences in various brain regions could be used to reliably classify monolinguals versus bilinguals. We hypothesized that a decision tree (DT) classifier will be able to discriminate between monolinguals and bilinguals using structural MRI measurements. We restrict our analysis to 32 anatomically-defined brain regions of interest that have been shown to be reliably involved in language processes by large meta-analyses (e.g. Price, 2010, 2012). DTs have previously been used to classify brain disease states (Aguilar et al., 2013); the present study is the first to use this methodology to predict cognitive-linguistic abilities from structural MRI data. We have included both gray and white matter measurements because the relative degree of gray versus white matter differences in the bilingual brain remains unclear (Garcia-Penton, Fernandez Garcia, Costello, Andoni Dunabeitia, & Carreiras, 2016), and gray and white matter measurements in a given region are not always correlated (Li, Legault, & Litcofsky, 2014). The present study also is the first to our knowledge to investigate the neurobiology of Farsi-English bilingualism. While most of the previous research suggests that the neurobiology of bilingualism is independent of the particular languages acquired (Ueno et al., 2014;

Table 1

Frequency of Freesurfer features identified during the cross-validation procedure (48 iterations). Only features that were identified at least once are listed. L = left hemisphere, R = right hemisphere, GM = gray matter, WM = white matter.

Feature	Frequency
L parstriangularis mean curvature	48
L inferior temporal GM volume	48
R supramarginal GM volume	48
R superior temporal GM volume	41
R supramarginal WM volume	20
R inferior temporal GM volume	4
L inferior temporal mean curvature	3
L inferior parietal GM volume	3
L middle temporal mean curvature	2
R middle temporal mean curvature	2
R superior temporal mean curvature	2
L inferior parietal WM volume	1
L inferior temporal WM volume	1
L superior temporal WM volume	1
R superior temporal WM volume	1
L parstriangularis GM volume	1
R bankssts GM volume	1
L Hippocampus volume	1
LThalamus-Proper volume	1
R Hippocampus volume	1
R Pallidum volume	1
R Thalamus-Proper volume	1
R Cerebellum-Cortex volume	1
R superior frontal mean curvature	1
L superior frontal GM volume	1

Kim et al., 1997), there is recent evidence that the connectivity between language processing regions may vary across drastically different languages (e.g. tonal vs. non-tonal) (Ge et al., 2015); however, none of these studies have focused on Indo-Persian languages, and Farsi is not a tonal language. To that end, we extend the previous literature in two important ways: (1) we add to the limited existing literature regarding structural neural predictors (versus correlates or group-level features) of bilingualism; and (2) we investigate the neurobiology of bilingualism in Farsi bilinguals.

2. Results & discussion

Leave-one-out cross validation is used to evaluate the feature selection and classification methods. For feature selection, we use pairwise *t*-test comparisons of individual features between the monolingual and the bilingual groups on the training data only ($p < 0.1$). We correct for multiple testing by controlling the false discovery rate per the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). The cross-validation procedure results in a different set of features and a different decision tree for each cross-validation fold; however certain features stand out (see Table 1). In the table, we show the list of the selected features, sorted by their selection frequency. In the right hemisphere, three features were selected in most cross-validation folds: the cortical volume of the supramarginal gyrus (selected 100% of the time), the white matter volume of the supramarginal gyrus (selected 41.7% of the time), and the cortical volume of the superior temporal gyrus (selected 85.4% of the time). In the left hemisphere, two features were selected for most cross-validation folds: the cortical volume of the inferior temporal gyrus (selected 100% of the time); and the mean curvature of the pars triangularis (selected 100% of the time). All other features were selected less than 4% of the time.

For the five features identified above, we also report the corresponding statistics on all the data for completeness. As before, alpha levels were adjusted according to the Benjamini-Hochberg method. In the right hemisphere, the three areas were identified as significantly larger in the bilingual group than the monolingual group: the cortical

volume of the supramarginal gyrus, $t(46) = 3.50, p = 0.001$; the white matter volume of the supramarginal gyrus, $t(46) = 2.44, p = 0.007$; and the cortical volume of the superior temporal gyrus, $t(46) = 2.75, p = 0.004$. The two left hemisphere areas also were found to be significantly larger in the bilingual group than the monolingual group: the cortical volume of the inferior temporal gyrus, $t(46) = 3.01, p = 0.001$; and the mean curvature of the pars triangularis, $t(46) = 2.90, p = 0.01$. There were no regions in which the monolingual group exhibited larger measurements, coinciding with previous literature and the assumption that bilingual versus monolingual, not differing L1s between the two groups, are driving these findings (there are no previous findings to suggest structural brain differences between different native languages). The two groups are approximately balanced for gender (57% female in monolingual group; 45% female in bilingual group); however, to ensure that the observed differences between groups are related to language and not gender, we conduct the same pairwise comparison for the features by gender. There was no statistically significant difference between the male and female groups for any of the features at a significance level of 0.1 after correcting for multiple testing by controlling the false discovery rate as above.

For each cross-validation fold, we trained a DT classifier to classify between monolinguals and bilinguals and evaluated its performance on the remaining held-out sample. Only the features selected during that fold were used to train the classifier. This is done in an effort to reduce the complexity of the classification scheme and to prevent overfitting (Liu & Motoda, 2007). By iterating over all data samples, we calculate the confusion matrix that describes the performance of the DT classification algorithm (see Table 2). As the table shows, the algorithm reliably classifies bilinguals with a true positive rate of 85.0% and a false positive rate of 28.6%.

One of the benefits of the DT classifier is that it parses the selected features into an easily-interpretable decision rule. In Fig. 1, we show the resulting DT when using all data (48 samples) and the top features for training. We note that this is not the tree used to generate the results in Table 2 since different DTs are generated for each cross-validation fold to generate those results. We simply calculate this DT for visualization. Features closer to the root of this tree are deemed to be most important for correctly classifying between the two groups. The analysis identifies the right supramarginal cortical volume, the left pars triangularis curvature and the left inferior temporal cortical volume as being the three most important and non-redundant regions for classifying between the monolingual and bilingual subjects. Fig. 2 contains pairwise scatter plots of the top three features identified by the decision tree: RH supramarginal cortical volume, the LH pars triangularis curvature, and the LH inferiortemporal cortical volume. As the figure shows, there is a visible difference between the two classes, with bilinguals predominately exhibiting larger cortical volume and increased curvature. For interpretability, it is worth jointly interpreting Figs. 1 and 2. The decision tree indicates that any individual with normalized RH Supramarginal Volume ≤ 0.00574 is automatically classified as monolingual, whereas any individual with a normalized RH Supramarginal Volume > 0.00574 and LH Parstriangularis Curvature > 0.128 is automatically classified as bilingual. These regions roughly

Table 2

Confusion matrix for DT trained using leave-one-out cross validation. This table shows the number of correctly classified instances (in bold along the diagonal) and the number of incorrectly classified instances (in the off-diagonal components) for predicting an individual to be monolingual or bilingual based on cortical volume, white matter volume, and mean curvature. Results were obtained using leave-one-out cross validation.

	Monolingual predicted	Bilingual predicted
Monolingual	20 (71.4%)	8 (28.6%)
Bilingual	3 (15.0%)	17 (85.0%)

correspond to the bottom and top quadrant of Fig. 2(a). It is visually clear from the plot that the decision boundaries automatically inferred by the decision tree provide an effective rule for separating between the two classes.

The location of the most relevant regions of interest (ROIs) identified in the present study also align with previous findings in bilingualism and the neurobiology of language more generally. The gray matter volume of the right supramarginal gyrus (SMG) ROI is the top predictive feature identified here. The SMG has been implicated in phonological and semantic processing (McDermott, Petersen, Watson, & Ojemann, 2003) and the right SMG has been shown to contribute to phonological processing independent of the left SMG's contributions (Hartwigsen et al., 2010). The right SMG has previously been identified in a group-comparison study: Grogan et al.'s (2012) voxel based morphometry study of bilingual and multi-lingual (speakers of 3+ languages) found greater gray matter density in the posterior portion of the right SMG in multilinguals compared to bilinguals. It is important to note that the Freesurfer software ROI protocol also groups a small portion of the posterior superior temporal gyrus (pSTG) into the supramarginal gyrus ROI, so it is possible that the STG contributions are driving the involvement of the supramarginal gyrus ROI; pSTG regions are strongly implicated in phonological and lexical processing of both L1 and L2s, language acquisition, and auditory-motor integration of speech (Hickok & Poeppel, 2007).

The second most relevant feature identified, the cortical volume of the left inferior temporal gyrus (ITG), is a large functionally diverse area of cortex, with subregions implicated in numerous aspects of language including lexical processing and conceptual semantic knowledge (Buckner, Koutstall, Schacter, & Rosen, 2000), while the more posterior aspects are known to be engaged in complex visual perception (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). Abutalebi et al. (2014)'s voxel-based morphometry study reports greater anterior ITG volume in their bilingual participants (Cantonese-English and Cantonese-Mandarin speakers) compared to Italian monolinguals. The ITG also is known to be strongly functionally connected to classic speech perception and comprehension areas in the STG, middle temporal gyri and the inferior frontal gyrus (Bookheimer, Zeffiro, Blaxton, Gaillard, & Theodore, 1995). Thus our finding of left ITG involvement in bilingualism suggests it may be informative for future studies to explore the relative contributions of these semantic and visual processes in ITG to bilingualism.

The third most relevant feature identified was the mean curvature of the pars triangularis, i.e. the middle third of the left inferior frontal gyrus (IFG) and the anterior half of Broca's area. Left IFG, and Broca's area more specifically, has long been implicated in speech production and verbal working memory, as well as syntactic processing, semantic integration and cognitive control (Rogalsky & Hickok, 2011). IFG also has been implicated in numerous group-level studies of bilingualism as a region that has greater volume in bilinguals than monolinguals (Klein et al., 2014; see Stein, Winkler, Kaiser, & Dierks, 2014 for a review). IFG also exhibits different spatial and temporal activation patterns in response to L2 as a function of L2 proficiency (Chai et al., 2016; Kim et al., 1997; Kovelman et al., 2008; Saur et al., 2009; Stein et al., 2006). In fact, Kovelman et al., based on their group-averaged fMRI analyses, suggest that the response of the left IFG to L2 may be a potential "neural signature" of bilingualism. However, the present study suggests that the IFG is perhaps not the most reliable feature of bilingualism, although it clearly contributes as it is the third most relevant feature. One possible reason for the differences in the impact of IFG between the present study and previous group-averaged studies is that Broca's area is known to have substantial individual variability. For example, significant fMRI activations in Broca's area in response to a language task identified in a group-averaged analysis do not well represent a majority of single subjects' activation patterns (Nieto-Castanon & Fedorenko, 2012; Rogalsky et al., 2015). Thus additional predictive studies, as opposed to group-averaged comparisons, are needed to determine the relative

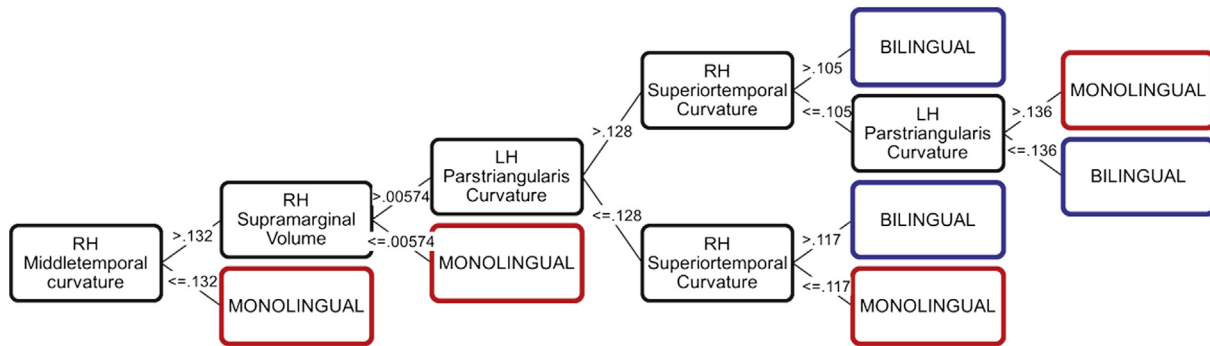


Fig. 1. Example DT using all 48 samples of data and the top features for training. This DT is depicted for visualization purposes, but note that different DTs were generated for each cross-validation fold.

predictive strength of left IFG measurements. Future studies also are needed to determine the role of structural variations among populations from different geographic regions of origin, as this may be contributing to cortical structure individual variability (Isamah et al., 2010).

The top five features include three cortical volume measures, one white matter volume measure and two mean curvature measures. Any explanations we could include here regarding why these particular features of these regions were the top features would mostly be post hoc

conjecture, so we resist the urge to do so. However, the fact that the top five features include cortical volume, white matter volume and mean curvature measures suggests that the inclusion of multiple structural measures may be advantageous for identifying and understanding the differences between the monolingual and bilingual brain. As mentioned in the introduction, previous neuroimaging studies of bilingualism have reported group-level differences in both gray and white matter volume measurements, as well as cortical thickness and surface area, for several

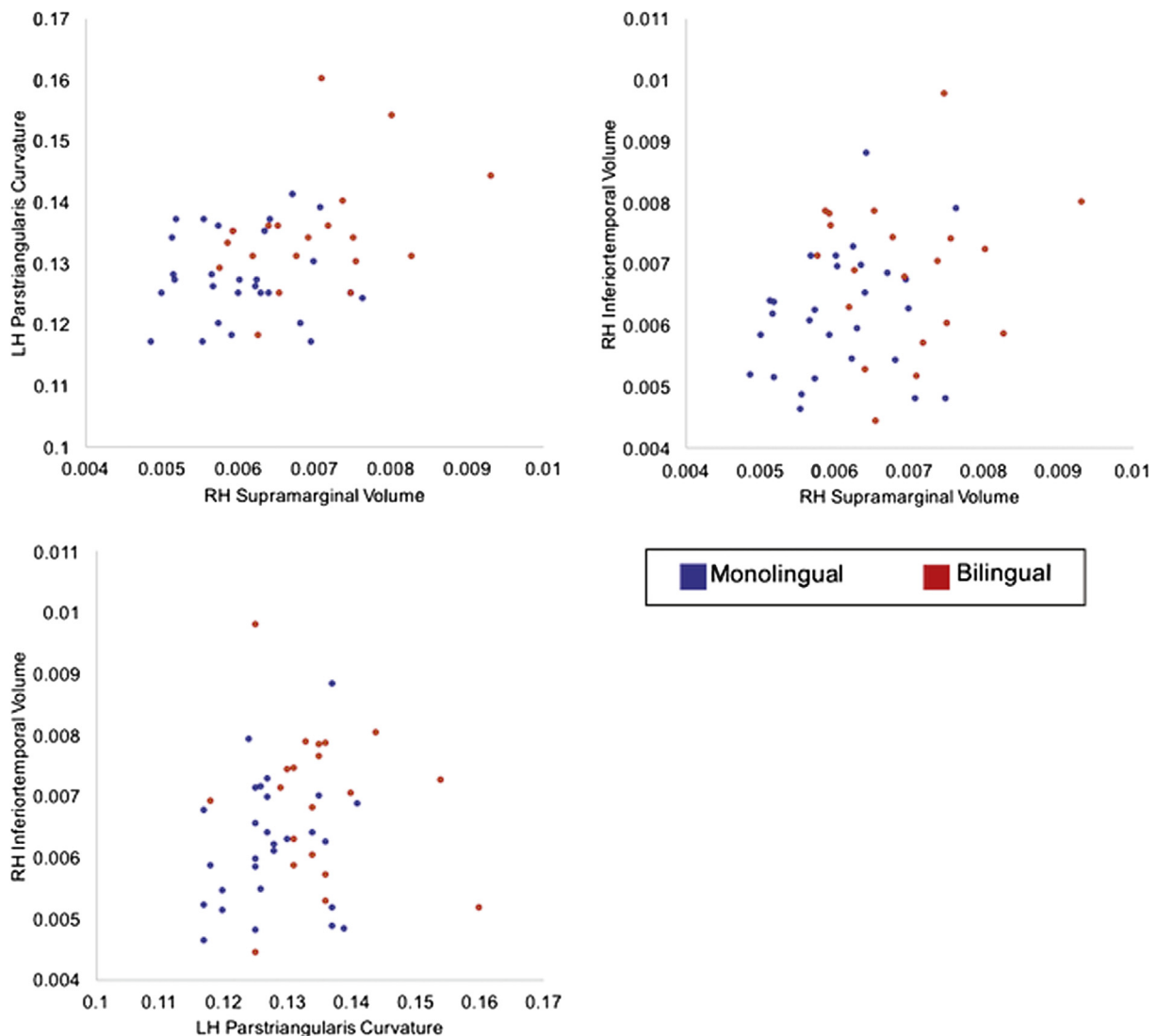


Fig. 2. Pairwise scatter plots for the top three features identified by the DT depicted in Fig. 1.

ROIs including the ones we identified in SMG, STG, Broca's area. Measures of local gyrification such as mean curvature have been investigated less frequently, but our results suggest that a gyrification measure also may be advantageous in understanding brain differences between monolinguals and bilinguals. Gyrification measures have previously been used to distinguish between controls and several neurological diseases (e.g. Prins et al., 2016; Wallace et al., 2013). There is some debate regarding the optimal way to quantify gyrification (Schaer et al., 2008; Van Essen & Drury, 1997; Zilles, Armstrong, Schleicher, & Kretschmann, 1988), so future work is needed to better understand the optimal way to measure it for comparisons between neurotypical groups such as monolinguals and bilinguals.

The present study is meant to be an initial attempt at using decision tree classification on neuroimaging data to predict bilingualism, and as such there are several limitations to this study that should be considered when interpreting our findings and planning future studies. It is likely that our monolingual and bilingual groups differ on more relevant factors than just language(s) acquired; for example, there was a significant difference between groups regarding number of years living in the United States. More detailed participant information, including parents' education level and reason for migration to the U.S. would be helpful to ensure that the features identified are in fact predictive feature of bilingualism *per se*. However, we reduced these possibilities by limiting our analyses to ROIs previously reliably implicated in language processing, and our findings do largely coincide with neural correlates of bilingualism determined by previous group-averaged studies of bilingualism (Klein et al., 2014; Parker Jones et al., 2012; Sebastian et al., 2001; Wei et al., 2015). Furthermore, only bilingual > monolingual differences were detected, coinciding with previous structural MRI findings (Abutalebi et al., 2014; Mechelli et al., 2004; Pliatsikas, Johnstone & Marinis, 2014). If differences in ethnicity or environmental factors were driving the results, one also would expect some of the features to be larger in one group while other regions' measurements to be larger in the other. For example, a recent study comparing structural brain measures of Caucasians and Han Chinese participants found the Chinese group to have relatively larger temporal and cingulate structures, but smaller frontal and parietal cortices than the Caucasian group (Tang et al., 2018). Previous work also has found significant positive correlations between level of education and structural brain measures (Steffener et al., 2016), but there was no significant difference between our groups' education levels. It also is likely that the bilingual group contains substantial intra-group variability. For example, the bilingual group was quite varied regarding self-reported English proficiency, and subjective self-assessments of linguistic and cognitive abilities can be unreliable (Kruger & Dunning, 1999). But variance in L2 proficiency amongst the bilinguals does not diminish the present findings because it would only add noise to our dataset, thereby making accurate classification more difficult. Future studies may find greater classification accuracy if level of proficiency is controlled for and measured by objective assessments.

Despite the limitations of the present study, our findings suggest that measurements extracted from a T1 structural MRI scan be used to detect brain differences between monolinguals and bilinguals. We focused on three measures that are automatically computed by Freesurfer's standard pipeline, namely cortical volume, white matter volume, and mean curvature of the gray/white boundary. These measures sometimes co-vary as a function of age or disease state (Bajaj, Alkozei, Dailey, & Killgore, 2017; Lemaitre et al., 2012; cf. Li et al., 2014), but the calculations for each are independent of one another (unlike, for example, surface area and volume). Recent neuroimaging work has expanded upon the typical volume and surface measurements to quantify neural differences regarding functional and structural connectivity, white matter density, and even cytoarchitectonic profiles (e.g. Del Gaizo et al., 2017; Ohtani et al., 2017; Assaf, 2018). These informative measures certainly can be included in decision tree classifier algorithms and would likely further improve classification

accuracy, and, perhaps more importantly, provide valuable insights into the underlying mechanisms that are driving the gray and white matter differences in the ROIs we have identified. For example within our data set, one could ask if the increased white matter volume in SMG for bilinguals is related to increased connectivity to additional brain regions compared to monolinguals, or due to increased density of white matter connections within the same pathways? Or, what layer(s) of SMG are mostly contributing to the increased volume? Nonetheless, there also is merit in being able to reliably identify features of a cognitive state (including bilingualism) from a single, commonly-employed MRI protocol (T1, acquisition time less than ten minutes) using measures easily accessible via a free, readily available, automatic processing stream such as Freesurfer.

In summary, there is great consistency in the brain regions that are identified as being important to distinguish between monolinguals and bilinguals and those in previous group-studies comparing bilinguals and monolinguals (i.e. IFG, STG, and SMG). This is particularly important because different L1 and L2 languages were investigated across studies, suggesting language-invariant predictors of L2 experience are possible. This coincides with much previous work indicating that the same brain regions are engaged across many different languages (Ge et al., 2015; Ueno et al., 2014; Kuhl et al., 2016), although the functional connectivity of the brain regions may vary across languages (Ge et al., 2015). Thus, the present study is a strong step towards identifying reliable cortical structural predictors of bilingualism, with 85% accuracy in a cross-sectional data set. Future studies with multiple language combinations are needed to determine the accuracy of our DT algorithm with structural MRI in predicting bilingualism and to validate the hypothesis that there exist language-invariant neurobiological biomarkers of bilingualism.

3. Methods

MRI data was collected as part of an unrelated research project at the Keller Center for Imaging Innovation at Barrow Neurological Institute (BNI) in Phoenix, AZ, on a research dedicated 3T Philips Ingenia scanner. The resulting MRI database consisted of 28 English-speaking monolinguals (16 female, all right-handed, ages 18–36, education range = high school–master's degree, median education = 14 years of education (i.e., years living in the United States range = 3 months – 33 years with a mean of 18.7 years) and 20 Farsi-English bilinguals (9 female, L1 = Farsi, L2 = English, all right-handed, ages 18–36, education range = high school–master's degree, median education = 16 years of education, years living in the United States range = 5 months – 20 years with a mean of 6 years). There was a significant difference between the monolingual and bilingual groups regarding years living in the United States ($\alpha < .05$): the monolingual group had spent significantly more years in the United States than the bilingual group, $t(44) = 5.1, p < .001$. Education level was collected by participants selecting one of seven categories ranging from "some high school" to "doctoral degree." There was no significant difference ($\alpha < .05$) between the education levels of the monolingual and bilingual groups (Mann-Whitney $U = 257, p = .40$).

All participants were asked to rate their fluency in English in three areas (speaking, reading, and writing) on a scale of 1–7 with 1 indicating "not at all fluent" and 7 indicating "very fluent." On all three rankings, the monolingual group's responses ranged from 6 to 7, while the bilingual group's responses ranged from 4 to 7. There was a significant difference ($\alpha < .05$) between the monolingual and bilingual groups' self-rankings for English speaking fluency (Mann-Whitney $U = 192.5, p = .008$), but no significant difference regarding their self-rankings of English reading fluency (Mann-Whitney $U = 255, p = .22$) or English writing fluency (Mann-Whitney $U = 240, p = .11$).

Each subject underwent a T1-weighted gradient-echo structural MRI protocol with the following parameters: TR = 6.75 ms, TE = 3.11 ms, flip angle = 90 degrees, FOV = 256 × 256, 170 sagittal slices, in plane

resolution = 1 mm × 1 mm, slice thickness = 1.2 mm.

3.1. Data processing

The raw MRI data was pre-processed, reconstructed and segmented using standard procedures provided by Freesurfer software (<http://surfer.nmr.mgh.harvard.edu/>; Dale, Fischl, & Sereno, 1999; Fischl & Dale, 2000). Pre-processing included removal of non-brain tissue (Segonne et al., 2004), automated Talairach transformation, segmentation of the subcortical white matter and deep gray matter structures (2004; Fischl et al., 2002), intensity normalization (Sled, Zijdenbos, & Evans, 1998), tessellation of the gray matter/white matter boundary, and automated topology correction (Fischl, Liu, & Dale, 2001; Segonne, Pacheco, & Fischl, 2007). The cerebral cortex was then automatically parcellated into 34 anatomical regions of interest in each hemisphere and measurements were extracted for each subject for each region using default Freesurfer procedures (Desikan et al., 2006; Fischl et al., 2004). Freesurfer's automated procedure for volumetric measures also was implemented to parcellate non-cerebral cortex brain regions into 40 regions as explained in detail in Fischl et al., 2002. For the present study, we restrict our analysis to a set of 20 Freesurfer cortical regions (10 in each hemisphere), 10 subcortical regions (5 in each hemisphere) and the cerebellum (one ROI for each hemisphere), corresponding to regions that are frequently implicated in language comprehension and/or production (Price, 2010, 2012; Sebastian et al., 2001) and are included in leading models of the neurobiology of language (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). In each hemisphere the same ten cortical regions were selected: (Freesurfer abbreviations are in parentheses): superior temporal gyrus (superiortemporal), middle temporal gyrus (middletemporal), inferior temporal gyrus (inferiortemporal), pars opercularis (parsopercularis), pars triangularis (parstriangularis), supplemental and pre-supplemental motor areas (superiorfrontal), supramarginal gyrus (supramarginal), inferior parietal lobule (inferioparietal), transverse temporal gyrus (transverse-temporal), and the banks of the superior temporal sulcus (bankssts). The five subcortical ROIs in each hemisphere were: thalamus, caudate, putamen, pallidum, and hippocampus.

For each cortical ROI, we used Freesurfer to estimate the mean curvature, cortical volume, and white matter volume; for the cerebellum and subcortical ROIs, total volume was estimated using Freesurfer (Fischl & Dale, 2000). Mean curvature is computed as the integral of $(k_1 + k_2)/2$ where k_1 and k_2 represent, respectively, the maximum and minimum curvatures in each vertex of the white matter surface constructed by Freesurfer (Van Essen & Drury, 1997). Freesurfer computes cortical volume as the product of the gray matter surface area and cortical thickness (cortical thickness is defined as the shortest Euclidean distance between the white matter and pial surfaces at each vertex in the ROI) (Fischl & Dale, 2000). White matter volume is computed from Freesurfer's subcortical volume segmentation procedures (Fischl et al., 2002), assigning white matter to the nearest cortical ROI label. The subcortical ROIs have a measure of gray matter volume; they do not have mean curvature as this is a measure of cortical gyrfication. Subcortical ROIs also do not have white matter volume estimates as the subcortical ROIs are defined such that they mostly contain gray matter. We also included only a total volume for the cerebellum. Future studies, utilizing specialized segmentation procedures for the cerebellum are optimal to compute additional measures (e.g. Romero et al., 2017; Yang et al., 2016) and are outside the scope of this study. The volume measures, but not mean curvature, were normalized by the total volume of each individual's total cranial volume because previous work has found a significant correlation between volumetric measurements of many brain structures with overall head size (Sanfilippo, Benedict, Zivadinov, & Bakshi, 2004), and there is no consensus regarding if or how mean curvature scales with head size (Schwarz et al., 2016; Toro et al., 2008). This results in a set of 72 candidate features (20 cortical ROIs, each having mean curvature, cortical volume, and

white matter volume measurements; 12 cerebellum and subcortical ROIs each with a volume measurement).

3.2. Feature selection

We use leave-one-out cross-validation to select the features and to train the binary decision tree classifier (Kohavi, 1995). For each cross-validation fold we select 47 of the 48 samples for training and leave one out for testing. Using only data from the training set, we downselect the number of features and train a decision-tree classifier on the 47 samples. The remaining sample is used to evaluate the fully-trained decision-tree. This procedure is repeated 48 times by leaving out a different sample each time until we generate a model prediction for each sample in our data.

For each of the 72 candidate features, we used an independent samples *t*-test on the training data to evaluate the hypothesis that there is no difference between the means of the monolingual and bilingual groups for that feature. We corrected for multiple tests by controlling for the false discovery rate per the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). We identified the subset of regions that exhibit a statistically significant difference between the two groups at a two-tailed threshold of $p < 0.1$.

3.3. Classification algorithm

The statistical analysis resulted in a subset of features (from the original set of 54) that we use to construct a binary decision tree classifier. A classifier was trained for each cross-validation fold using only the training data. The DT classifier was trained using the Classification and Regression Tree (CART) algorithm in Matlab (Breiman, Friedman, Olshen, & Stone, 1984). We report the decision tree (DT) and the resulting confusion matrix on the test data only. Neither the feature selection nor the classification procedures made use of any of the held-out test data for each cross-validation fold.

Statement of significance

There is no consensus regarding how the bilingual brain differs from a monolingual brain. This manuscript presents an approach to reliably identify brain regions affected by bilingualism. These brain regions in turn can be investigated to better understand the differences in the neural mechanisms associated with bilingualism.

Acknowledgments

Funding provided by DARPA-12-03-N2-FP-004 (Corman), the ASU Fulton Undergraduate Research Initiative (Gilton), and grant 1R21DC013812 from the National Institutes of Health (Berisha). The authors thank Arianna LaCroix for her assistance with the demographic and behavioral data analysis.

References

- Abutalebi, J., Canini, M., Della Rosa, R. A., Sheung, L. P., Green, D. W., & Weekes, B. S. (2014). Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiology of Aging*, 35(9), 2126–2133.
- Aguilar, C., Westman, E., Muehlboeck, J. S., Mecocci, P., Vellas, B., Tzolaki, M., ... Wahlund, L.-O. (2013). Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research*, 212(2), 89–98.
- Assaf, Y. (2018). Imaging laminar structures in the gray matter with diffusion MRI. *Neuroimage*, epub ahead of print: doi: 10.1016/j.neuroimage.2017.12.096.
- Bajaj, S., Alkozei, A., Dailey, N. S., & Killgore, W. D. S. (2017). Brain aging: Uncovering cortical characteristics of healthy aging in young adults. *Frontiers in Aging Neuroscience*, 9, 412.
- Benjamini, Yoav, & Hochberg, Yosef (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, 289–300.
- Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and

- brain. *Trends in Cognitive Sciences*, 16(4), 240–250.
- Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, 3(2), 93–106.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Boca Raton, FL: Chapman & Hall.
- Buckner, R. L., Koutstall, W., Schacter, D. L., & Rosen, B. R. (2000). Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain*, 123(3), 620–640.
- Burgaleta, M., Sanjuan, A., Ventura-Campos, N., Sebastian-Galles, N., & Avila, C. (2016). Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage*, 125, 437–445.
- Chai, X. J., Berken, J. A., Barbeau, E. B., Soles, J., Chen, J. K., & Klein, D. (2016). Intrinsic functional connectivity in the adult brain and success in second-language learning. *Journal of Neuroscience*, 36(3), 755–761.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9, 179–194.
- De Bruin, A., Roelofs, A., Dijkstra, T., & Fitzpatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: fMRI evidence from trilingual speakers.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., ... Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *NeuroReport*, 8, 3809–3815.
- Del Gaizo, J., Fridriksson, J., Yourganov, G., Hillis, A.E., Hickok, G., Mísic, B. ... Leonardo Bonilha (2017). Mapping language networks using the structural and dynamic brain connectomes. *eNeuro* 4(5). Pii: ENEURO.0204-17.2017.
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31, 968–980.
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences U S A*, 97, 11050–11055.
- Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Transactions on Medical Imaging*, 20, 70–80.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., ... Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3), 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., ... Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14, 11–22.
- Frenc-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, 128(3), 528–536.
- García-Penton, L., Fernández García, Y., Costello, B., Andoni Dunabeitia, J., & Carreiras, M. (2016). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, 31(3), 303–327.
- García-Penton, L., Pérez Fernández, A., Iturría-Medina, Y., Gillion-Dowens, M., & Carreiras, M. (2014). Anatomical connectivity changes in the bilingual brain. *NeuroImage*, 84(1), 495–504.
- Ge, J., Peng, G., Lyu, B., Wang, Y., Zhuo, Y., Niu, Zhendong, ... Gao, J. H. (2015). Cross-language differences in the brain network subserving intelligible speech. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10), 2972–2977.
- Grogan, A., Green, D. W., Ali, N., Crinion, J. T., & Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex*, 19(11), 2690–2698.
- Grogan, A., Parker Jones, O., Ali, N., Crinion, J., Orabona, S., Mechias, M. L., ... Price, C. J. (2012). Structural correlates for lexical efficiency and number of languages in non-native speakers of English. *Neuropsychologia*, 50(7), 1347–1352.
- Grundy, J. G., Anderson, J. A. E., & Bialystok, E. (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Annals of the New York Academy of Sciences*, 1396(1), 183–201.
- Halsband, U. (2006). Bilingual and multilingual processing. *Journal of Physiology, Paris*, 99(4–6), 355–369.
- Hartwigsen, G., Baumgaertner, A., Price, C. J., Koehnke, M., Ulmer, S., & Siebner, H. R. (2010). Phonological decisions require both the left and right supramarginal gyri. *Proceedings of the National Academy of Sciences*, 107(38), 16494–16499.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Isamah, N., Faison, W., Payne, M. E., MacFall, J., Steffens, D. C., Beyer, J. L., ... Taylor, W. D. (2010). Variability in frontotemporal brain structure: The importance of recruitment of African Americans in neuroscience research. *PLoS One*, 5(10), e13642.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, 96(16), 9379–9384.
- Kim, K. H., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388(6638), 171–174.
- Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, 92(7), 2899–2903.
- Klein, D., Mok, K., Chen, J. K., & Watkins, K. E. (2014). Age of learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, 131, 20–24.
- Klein, D., Zatorre, R. J., Chen, J. K., Milner, B., Crane, J., Belin, P., & Bouffard, M. (2006). Bilingual brain organization: A functional magnetic resonance adaptation study. *NeuroImage*, 31(1), 366–375.
- Kohavi, Ron, (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*. 14(2).
- Kovelman, I., Baker, S. A., & Petitto, L. (2008). Bilingual and monolingual brains compared: A functional magnetic resonance imaging investigation of syntactic processing and a possible “neural signature” of bilingualism. *Journal of Cognitive Neuroscience*, 20(1), 153–169.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
- Kuhl, P. K., Stevenson, J., Corrigan, N. M., van den Boesch, J. J. F., Deniz Can, D., & Richards, T. (2016). Neuroimaging of the bilingual brain: Structural brain correlates of listening and speaking in a second language. *Brain and Language*, 162, 1–9.
- Lemaitre, H., Goldman, A.L., Sambataro, F., Verchinski, B.A., Meyer-Lindenberg, A., Weinberger, D.R. & Mattay, V.S. (2012). Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and grey matter volume? *Neurobiology Aging* 33(3), 617e.1-617e.9.
- Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, 58, 301–324.
- Liu, H., & Motoda, H. (2007). *Computational methods of feature selection*. CRC Press.
- Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2011). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes*, 27(10), 1479–1488.
- Mahendra, N., Plante, E., Magliore, J., Milman, L., & Trouard, T. P. (2003). fMRI variability and the localization of languages in the bilingual brain. *NeuroReport*, 14, 1225–1228.
- Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, 86, 70–82.
- Mayima, P. C., Richards, T. L., Coe, B. P., Eichler, E. E., & Kuhl, P. K. (2016). Brain white matter structure and COMT gene are linked to second-language learning in adults. *Proceedings of the National Academy of Sciences of the USA*, 113(26), 7249–7254.
- McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia*, 41(3), 293–303.
- Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R., & Price, C. J. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature*, 431(7010), 757.
- Nieto-Castanon, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, 63(3), 1646–1669.
- Ohtani, T., Nestor, P. G., Bouix, S., Newell, D., Melonakos, E. D., & McCarley, R. W. (2017). Exploring the neural substrates of attentional control and human intelligence: Diffusion tensor imaging of prefrontal white matter tractography in healthy cognition. *Neuroscience*, 341, 52–60.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258.
- Paap, K. R., Johnson, H. A., & Sawi, H. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265–278.
- Parker Jones, O., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., ... Price, C. J. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex*, 22(4), 892–902.
- Perani, D., & Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, 15, 202–206.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., ... Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.
- Pliatsikas, C., Johnstone, T., & Marinis, T. (2014). Grey matter volume in the cerebellum is related to the processing of grammatical rules in a second language: A structural voxel-based morphometry study. *Cerebellum*, 13(1), 55–63.
- Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191, 62–88.
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816–847.
- Price, C. J., Green, D. W., & Von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain*, 122(Pt12), 2221–2235.
- Prins, D., Plank, T., Baseler, H. A., Gouws, A. D., Beer, A., Morland, A. B., ... Cornelissen, F. W. (2016). Surface-based analyses of anatomical properties of the visual cortex in macular degeneration. *PLoS One*, 11(1), e0146684.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
- Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., ... Ackermann, H. (2011). Individual differences in audio-vocal speech imitation aptitude in late bilinguals: Functional neuro-imaging and brain morphology. *Frontiers in Psychology*, 2, 271.
- Rodriguez-Fornells, A., Van der Lugt, A., Rotte, M., Britti, B., Heinze, J. J., & Munte, T. F. (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, 17, 422–433.
- Rogalsky, C., Almeida, D., Sprouse, J., & Hickok, G. (2015). Sentence processing selectivity in Broca's area: Evident for structure but not syntactic movement. *Language, Cognitive Neuroscience*, 30(10), 1326–1338.
- Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension.

- Journal of Cognitive Neuroscience*, 23(7), 1664–1680.
- Romero, J. E., Coupe, P., Giraud, R., Ta, V., Fonov, V., Park, M. T. M., ... Manjón, J. V. (2017). CERES: A new cerebellum lobule segmentation method. *Neuroimage*, 147, 916–924.
- Sakai, K. L., Miura, K., Narafu, N., & Muraishi, Y. (2004). Correlated functional changes in prefrontal cortex in twins induced by classroom education of second language. *Cerebral Cortex*, 14, 1233–1239.
- Sanfilippo, M. P., Benedict, R. H., Zivadinov, R., & Bakshi, R. (2004). Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: The proportion vs. residual method. *Neuroimage*, 22(4), 1732–1743.
- Saur, D., Baumgaertner, A., Moehring, A., Buchel, C., Bonnesen, M., Rose, M., ... Meisel, J. M. (2009). Word order processing in the bilingual brain. *Neuropsychologia*, 47(1), 158–168.
- Schaer, M., Cuadra, M. B., Tamarit, L., Lazeyras, F., Eliez, S., & Thiran, J. P. (2008). A surface-based approach to quantify local cortical gyrification. *IEEE Transactions on Medical Imaging*, 27(2), 161–170.
- Schlegel, A. A., Rudelson, J. J., & Tse, P. U. (2012). White matter structure changes as adults learn a second language. *Journal of Cognitive Neuroscience*, 24(8), 1664–1670.
- Schwarz, C. G., Gunter, J. L., Wiste, H. J., Przybelski, S. A., Weigand, S. D., Ward, C. P., ... Carrillo, M. C. (2016). A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *Neuroimage Clinical*, 11, 802–812.
- Sebastian, R., Laird, A. R., & Kiran, S. (2001). Meta-analysis of the neural representation of first and second language. *Applied Psycholinguistics* 32. P. 799–819.
- Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience*, 27(4), 819–831.
- Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., ... Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3), 1060–1075.
- Segonne, F., Pacheco, J., & Fischl, B. (2007). Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Transactions on Medical Imaging*, 26(4), 518–529.
- Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17(1), 87–97.
- Steffener, J., Habeck, C., O'Shea, D., Razlighi, Q., Bherer, L., & Stern, Y. (2016). Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiology of Aging*, 40, 138–144.
- Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, M., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *Neuroimage*, 33(2), 774–783.
- Stein, M., Winkler, C., Kaiser, A., & Dierks, T. (2014). Structural brain changes related to bilingualism: Does immersion make a difference? *Frontiers in Psychology*, 5, 1116.
- Stocco, A., & Prat, C. S. (2014). Bilingualism trains specific brain circuits involved in flexible rule selection and application. *Brain and Language*, 137, 50–61.
- Tang, Y., Zhao, L., Lou, Y., Shi, Y., Fang, R., Lin, X., Liu, S., & Toga, A. (2018). Brain structure differences between Chinese and Caucasian cohorts: A comprehensive morphometry study. *Human Brain Mapp*, epub ahead of print.
- Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, 16, 367–382.
- Toro, R., Perron, M., Pike, B., Richer, L., Veillette, S., Pausova, Z., ... Paus, T. (2008). Brain size and folding of the human cerebral cortex. *Cerebral Cortex*, 10, 2352–2357.
- Ueno, T., Saito, S., Saito, A., Tanida, Y., & Patterson, K. (2014). Lambon Ralph M.A., Not lost in translation: generalization of the primary systems hypothesis to Japanese-specific language processes. *Journal of Cognitive Neuroscience*, 26(2), 433–446.
- Van Essen, D. C., & Drury, H. A. (1997). Structural and functional analyses of human cerebral cortex using a surface-based atlas. *Journal of Neuroscience*, 17(18), 7079–7102.
- Wallace, G. L., Robustelli, B., Dankner, N., Kenworthy, L., Giedd, J. N., & Martin, A. (2013). Increased gyrification, but comparable surface area in adolescents with autism spectrum disorders. *Brain*, 136(6), 1956–1967.
- Wang, X., Wang, Y. Y., Jiang, T., Wang, Y. Z., & Wu, C. X. (2013). Direct evidence of the left caudate's role in bilingual control: An intra-operative electrical stimulation study. *Neurocase*, 19(5), 462–469.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37, 159–170.
- Wei, M., Joshi, A. A., Zhang, M., Mei, L., Manis, F. R., He, Q., ... Zhong, L. L. (2015). How age of acquisition influences brain architecture in bilinguals. *Journal of Neurolinguistics*, 36, 35–55.
- Yang, Z., Ye, C., Bogovic, J. A., Carass, A., Jedynek, B. M., Ying, S. H., ... Prince, J. L. (2016). Automated cerebellar lobule segmentation with application to cerebellar structural analysis in cerebellar disease. *Neuroimage*, 127, 435–444.
- Yokoyama, S., Okamoto, H., Miyamoto, T., Yoshimoto, K., Kim, J., Iwata, K., ... Kawashima, R. (2006). Cortical activation in the processing of passive sentences in L1 and L2: An fMRI study. *NeuroImage*, 30(2), 570–579.
- Zilles, K., Armstrong, E., Schleicher, A., & Kretschmann, H. J. (1988). The human pattern of gyrification in the cerebral cortex. *Anatomy and Embryology*, 179, 173–179.
- Zou, L., Ding, G., Abutalebi, J., Shu, H., & Peng, D. (2012). Structural plasticity of the left caudate in bimodal bilinguals. *Cortex*, 48(9), 1197–1206.

Further reading

- Hall, Mark, et al. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11.1, pp. 10–18.