Semantic Representations in Monolingual and Bilingual Connectionist Networks*

Nicholas Rendell and Eddy J. Davelaar

University of London, UK
{n.rendell, e.davelaar}@bbk.ac.uk

Repeated practice of inhibitory processes in bilinguals leads to an advantage in tasks requiring control processes. This advantage has been postulated to contribute to cognitive reserve as an offset to age and dementia related decline. Two models are presented which have learned the names of a series of pictures belonging to two categories. The first model learned the names of both categories in a single language, the second in two languages. In line with the dopamine hypothesis, change in gain of the log-sigmoidal transfer function was applied to provide a valid age related change. The results demonstrated greater separation of representations for monolinguals than for bilinguals. This occurred both for individual representations within a category and between categories. Furthermore, an interaction between brain reserve capacity, a biological category of cognitive reserve, and whether or not the model was bilingual or monolingual was observed for measures of separation. The results are discussed in terms of retrieval induced inhibition which suggest that the closer representations are to each other, the greater the recruitment of inhibitory processes.

Keywords: Ageing, Connectionist Model, Language, Bilingualism, Cognitive Reserve

* This article is an improved version from the Eurocog-ICCS proceedings.

Date submitted: 08/31/15  Date reviewed: 09/03/15
Date confirmed for publication: 09/30/15
©2015 Institute for Cognitive Science, Seoul National University
1. Introduction

Recent studies of bilingual and multilingual individuals have demonstrated some offsetting of normal cognitive ageing (Kavé, Eyal, Shorek, & Cohen-Mansfield, 2008) and protective effects against the onset of the cognitive symptoms of dementia. For example, Bialystok, Craik, & Freedman (2007) carried out a study in which 184 patients were selected from a set of 228 patients with a diagnosis of probable or possible Alzheimer’s disease (AD) who had been referred to a memory clinic complaining of poor cognitive function. Of the selected patients, 51% were judged to be bilingual. One of the initial results was that on average the bilingual participants were 3.2 years older than the monolinguals when they presented themselves to the clinic with cognitive dysfunction. In terms of the positive effect of bilingualism, it is one of a number of contributing factors known collectively as cognitive reserve (Stern, 2003, 2009), a latent variable representing the poor relationship between an individual’s cognitive intactness in vivo and levels of brain pathology (Mortimer, 1997; Valenzuela & Sachdev, 2006). Two general types of cognitive reserve have been established. The first being brain reserve capacity (BRC), which relates to a quantitative measure of neurons and connections available to an individual was second relates to reserve as the result of continual practice of the neural pathways associated with a particular cognitive function (Stern, 2009). This study examines consequences of bilingualism in respect of its contribution to cognitive reserve by comparing representations of picture categories in monolingual and bilingual networks. Furthermore, this will be carried out over neural network analogues of BRC and age-related dopaminergic decline.

More recent studies of multilingualism and its association with cognitive reserve have attempted to control for education and intelligence. Bak, Nissan, Allerhand, & Deary (2014) utilised the Lothian Birth Cohort, a group of English native speakers of European origin who were initially tested for a level of intelligence at age 11 in 1947. This allowed the authors to control for childhood intelligence, gender and socioeconomic status. The participants, now 73, were tested on fluid intelligence, memory, speed of information processing, reading and verbal fluency. The results
demonstrated a protective effect of bilingualism with no negative effects of having more than one language. Reading verbal fluency and general intelligence were the most affected and general intelligence in particular was related to improvement in executive processes.

To understand why bilingualism confers an advantage to cognitive ageing, the cognitive mechanisms involved in speaking more than one language must be unpacked. Studies in metalinguistic capabilities have uncovered the mechanisms behind the cognitive advantage in bilingual individuals. For example, Bialystok (1988) found that bilingual children demonstrated an advantage in tasks requiring cognitive control. Further, in error checking and explanation of ungrammatical sentences, Galambos & Goldin-Meadow (1990) found that bilingual children performed better in the trials which required a change in the focus of attention. However, both monolinguals and bilinguals performed equally on the actual explanation of the errors. Nonlinguistic studies have also demonstrated benefits for bilinguals in executive control tasks using perceptual stimuli (Costa, Hernández, & Sebastián-Gallés, 2008). This suggests that bilingualism provides a holistic strengthening of executive control processes. This assertion has been supported with neuroimaging studies which demonstrate stronger resting-state connectivity in the frontal lobe for bilingual rather than monolingual individuals (e.g. Grady, Luk, Craik, & Bialystok, 2015). Furthermore, Gold (2014) asserts that increased activity within frontal regions as a result of bilingualism serves to protect against age-related decline in those circuits related to executive processing.

Whilst this paper does not suggest that two completely distinct conceptual stores exist for bilinguals, it does assume that differences exist between bilinguals and monolinguals in conceptual or semantic representational storage. Even in compound bilinguals, while concrete nouns may overlap in conceptual storage, concepts relating to abstract terms may differ. For example, the word ‘love’ in English has very different conceptual connotations to the word ‘amour’ in French. As such, there will be a significant increase in concepts stored in bilingual individuals compared to monolingual individuals. However, it is worth noting that for coordinate bilinguals who have learnt different languages in different environments, two conceptual systems may exist (Koven, 1998; Pavlenko, 1999).
Whilst the cognitive advantages of bilingualism appear well documented, the linguistic deficits associated with having more than one language are equally well-researched. For example, it is generally accepted that one of the predominant negative effects of bilingualism is the vocabulary size. This is generally smaller compared to monolinguals for both languages spoken (Mahon & Crutchley, 2006; Portocarrero, Burright, & Donovick, 2007). In addition to size of lexicon, bilinguals also appear to have more trouble accessing particular words. Picture naming tasks have shown that bilinguals are slower than their monolingual counterparts (e.g. Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Further, verbal fluency tasks in which participants are asked to name as many words as possible for a given category or categories, have demonstrated a disadvantage for bilinguals (e.g. Rosselli et al., 2000). Deficits such as those described above suggest there exists implications with regards to lexical storage and spacing of representations when it comes to learning more than one language. Reduced vocabulary size relates to a reduced ability to store as many representations. However, poorer performance in picture naming tasks appears to be the result of the increased crowding within the representational space.

In order to distinguish between target representations and their neighbors in a more crowded storage space, greater recruitment of inhibitory processes are required. Therefore, rather than the inhibition of an entire language, the bilingual advantage may also occur due to greater recruitment of inhibitory processes on the level of individual item retrieval. This idea stems from the theory of retrieval induced inhibition (Anderson, Bjork, & Bjork, 1994) which posits that retrieval of one memory can inhibit the recollection of other, similar memories. This is due to the recruitment of inhibitory processes which serve to discriminate between the target representation and similar representations (for a review, see: Storm et al., 2015). Therefore, any differences in spatial separation between semantic representations within languages of monolinguals and bilinguals will be demonstrated by training a neural network with a picture naming task. Similarity of the representations themselves will be controlled for by randomizing the input patterns, it is predicted that spatial constraints will cause increased clustering of representations within languages.

The aim of this study was to investigate the relationship between
semantic representations in monolinguals and bilinguals. Further, the relationship between any differences between the two will be described in terms of cognitive reserve and its development over lifespan. To this end, two neural network models were trained to remember the names of a number of ‘pictures’ in one (monolingual) or two (bilingual) languages. The purpose of the task was to replicate the learning of semantic representations within each language for the monolingual and bilingual models. Ageing of the networks was simulated by adjusting the gain of the transfer function (Li, Lindenberger, & Sikström, 2001; Servan-Schreiber, Printz, & Cohen, 1990).

2. Method

2.1 Architecture

The models used in this study were simple three layer, feedforward back propagating neural networks. Two versions were used, a monolingual version and a bilingual version. The input layer was 26 nodes and the output was 40 nodes. For each of the two networks, 50 simulants were trained for hidden layer sizes of 5, 10, 15, and 20 nodes. The purpose of varying the hidden layer size was to represent the variability in BRC (Stern, 2009). The learning rate was 0.5 and the momentum was fixed at 0.1.

2.2 Stimulus Patterns

Given the focus of study for the models was the hidden layer rather than performance, a compromise between an artificial language and a realistic corpus was used for input. The inputs used in both models were patterns of 26 binary digits. The first 20 digits were randomized to control for similarity between representations since categorization due to input features was not the focus of the study. The next three binary digits represented a language tag. This was added to guarantee separation of the two sets of pictures in the bilingual model since the rest of the input consisted of a random pattern. The final three binary digits of each input presentation related to the membership of a semantic category, for example, emotions and characteristics (see Figure 1).
Figure 1. Example of a single binary input vector representing one picture. Tags for language and category at the end of the picture each take two forms to represent membership of one of two languages, English and Greek and two categories of representations.

34 input patterns were used in the monolingual model, the bilingual input set was augmented with a further 34 input patterns for the bilingual model, making 68 in total for the bilingual model. The matching output ‘words’ used for the monolingual network were taken from a dataset of English phonemes which had been converted to a binary input set using a set of 19 features (Thomas & Karmiloff-Smith, 2003). The output set for the monolingual network comprised of 34 English words with a further 34 Greek words produced for the bilingual model. The English words (L1) were used both in the monolingual and bilingual model and the Greek words represented the second language in the bilingual model (L2). In the monolingual model, the first 19 nodes in each output pattern were taken up by a representation of an English ‘word’ whilst the rest of the nodes in each pattern were left at zero. This set of output patterns was the same for the first 34 output patterns in the bilingual model. However, a further 34 output vectors were added for which the first 19 nodes were set to zero and the remaining 21 nodes represented a Greek ‘word’.

2.3 Training

Both networks were initially trained for 800 epochs. The starting weights for each was seeded randomly from a uniform distribution of between 0 and 1. Training for both networks took around 200 epochs for the error to reach an asymptotic state. For comparison, test data was introduced to both monolingual and bilingual networks in the form of both categories of L1 only. Overall, error settled at a slightly higher level in the bilingual network. This can be attributed to the increase in constraints in the bilingual network as it needed to accommodate the same amount of ‘pictures’ as
the monolingual network but in both languages. Given that an asymptotic state was achieved around 200 epochs, it was decided that at 220 epochs the network was considered mature for the purposes of analysis and interventions. Therefore, it was at this point that dopamine decline was initiated by incrementally making the gain of the sigmoidal transfer function closer to zero by steps of 0.0015 at each epoch. Equation 1 shows the transfer function where \( y \) is the output calculated from the activation \( a \), \( h \) is the threshold and \( k \) is the gain.

\[
y = f(a) = \frac{1}{1 + e^{-k(a-h)}}
\]  
(1)

Changing the gain to gradually approach zero reduces the steepness of the sigmoid function and as such makes the nodes in the hidden layer of the neural network in question increasingly less responsive to changes in the input. As such, this manipulation reflects the effects of reduced catecholamine effectiveness over age and the subsequent cognitive effect of decreased ability to detect a signal embedded in noise (Li, Lindenberger, & Sikström, 2001; Servan-Schreiber, Printz, & Cohen, 1990).

2.4 Analysis

Multidimensional analysis was applied to the activation profile of the hidden layer for both monolingual and bilingual networks. For an equal comparison, this was only applied to both categories in L1 for each of the network types. The purpose of this analysis was to observe any differences in spacing between representations in the single language within either of the categories. Furthermore, the development of representational spacing was carried out by calculating the sum of distances between a calculated centroid and all representations within a single category at each epoch. To complete the investigation, differences between categories of representations as a whole were calculated between monolingual and bilingual models. In order to observe this, the calculation of an F-value was derived. To this end, a single centroid was calculated using k-means clustering. From this point, distances were calculated to all representations within L1. Distances to the representations within each category from the overall centroid were also
calculated. This provided a measure of within and overall variance upon which to calculate an F-value. This was calculated at each epoch of training.

3. Results

The first three dimensions from the results of multidimensional scaling were plotted to illustrate any differences of semantic storage in representational space (see Figures 2 & 3). The overall trend in categorical separation is driven by the feature differences between the two categories. However, spreading of the representations within the categories is demonstrated over the increasing hidden layer sizes for both models. However, overall the effect appears greater with the monolingual model. Conversely, clustering of representations within the hidden layer of the bilingual model is tighter. This was also reflected by the development of representational spacing of a single category from L1 in both monolingual and bilingual networks carried out over the period of training (Figure 3).

Figure 2. Scatterplots representing the distributions of representations of categories A and B within L1 of the monolingual network. Each graph refers to hidden layer sizes of five (A), ten (B), fifteen (C) and twenty (D) nodes. The blue dots relate to category A and the red dots relate to picture representations in category B.
Figure 3. Scatterplots representing the distributions of representations of categories A and B within L1 of the bilingual network. Each graph refers to hidden layer sizes of five (A), ten (B), fifteen (C) and twenty (D) nodes. The blue dots relate to category A and the red dots relate to picture representations in category B.

Higher hidden layer size networks showed the greatest spacing between representations with the two most dispersed categories belonging to the monolingual network. Using data from the developmental projection (Figure 4), a 2*2 ANOVA was carried out on the distances from the individual representations to a centroid of a single representational category with monolingual or bilingual as one factor and hidden layer as the other. The analysis demonstrated main effects for both hidden layer (F (3,392) = 1539, p<.001) and type of network (F (1,392) = 516, p<.001) together with an interaction between the two (F (3,392) = 1953, p<.001). Therefore, it appeared that such an effect may be due to the differences in space cause by the additional constraints of second language storage offset only by a higher storage capacity.

Inline analysis of the separation of categories in L1 demonstrated an increasing separation of up until maturity for all groups (Figure 5). However, characteristics of the different projections differ in response to gain decline. Separation of categories appears almost immediately
Figure 4. Line graph demonstrating the projection of the sum of the distances from calculated centroid in category for monolingual and bilingual models over all hidden layer sizes. Lines represent mean score of 50 simulants. Dotted lines represent bilingual projections, solid lines represent monolingual projections. Different colors relate to the four hidden layer sizes.

Figure 5. Projections of F-values reflecting separation between semantic categories in both models overall all hidden layer sizes. Dotted lines represent bilingual projections, solid lines represent monolingual projections. Different colors relate to the four hidden layer sizes.
Figure 6. Bounded line graph of monolingual (blue) and bilingual (red) F-Value scores for ten unit hidden layer versions only. The shaded area around each represents one standard error of the mean.

for ten and fifteen node monolingual models. However, for the other sizes of hidden layer for monolingual and all bilingual hidden layer sizes, separation was a slower process. For all groups, categorical separation declined with dopaminergic decline over the ageing process. To understand any differences in variability between groups, F-value projections for simulants from the ten hidden layer networks for monolingual and bilingual networks were plotted (see Figure 6).

Multilevel analysis was used in which individual scores were used as the dependent variable with the epoch and as the first level predictor. The second level was grouped by whether the model was monolingual or bilingual.

\[
Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}
\]

Where Y represents the F-value at Epoch i for group j. A Likelihood ratio test demonstrated a significantly better fit with the inclusion of the random component ‘group’ \((p<.001)\). Therefore, inclusion of whether or not the data was produced by the monolingual or bilingual network provided a better prediction of F-value scores.
4. Discussion

This study represents an initial attempt at exploring the way in which bilingualism influences how representations within languages are stored and develop. This study explored differences between monolingual and bilingual models, each with differing levels of BRC. Both models were trained over a number of epochs, with the introduction of a gain change in the log-sigmoidal transfer function to represent ageing. The results of this study provide commentary on the way in which representational items and the categories that they belong to separate. Furthermore, this study adds validity to the suggestion that recruitment of inhibitory processes occurs at many levels. In this regard, greater clustering of representations adds to the overall complexity for which greater control is required. Furthermore, this interpretation contributes to the literature regarding the bilingual advantage which may provide the genesis for cognitive reserve in ageing individuals (Bialystok et al., 2007).

The main finding the study is that single representations within a semantic representation store of a single language of a bilingual speaker are more clustered than their monolingual counterparts. This difference interacts with differing levels of BRC with high BRC bilinguals demonstrating greater representational spacing than low BRC monolinguals. When the spatial difference between entire semantic categories is taken into account the pattern remains the same. This finding can be interpreted in terms of the retrieval induced inhibition hypothesis so that the bilingual advantage occurs due to repeated and greater recruitment of inhibitory mechanisms at the level of semantic and/or conceptual items.

Over the course of the lifespan of both monolingual and bilingual models in this study the separation of both concepts within categories and the categories themselves changes. For most of the groups, separation appears to continue up until maturity with the exception of the bilingual groups with low levels of BRC. This reduction in separation of representations at the developmental stage represents an increased recruitment of inhibitory processes over this period. This suggests that cognitive reserve will be greater for pure bilinguals for whom exposure to both languages occurs from a very early age. Furthermore, this effect will be increased in those
pure bilinguals with low levels of BRC. Further research might include manipulating the time at which a second language is used from which behavioral predictions can be made.

The analysis carried out in this study suggests two main effects of bilingualism on semantic memory. Firstly, Categories within a single representational space in a bilingual speaker are sensitive to space. Controlling for BRC demonstrates greater contraction and overlap of representations that the monolingual equivalent. Secondly, gain change produces differing effects according to the constraining factors. Ten and fifteen node hidden layer monolingual networks declined in representational separation almost immediately compared to the higher BRC monolingual model and all bilingual versions. Therefore, the highest amount of cognitive reserve would be observed in individuals with bilinguals with low BRC. This suggests a theoretical dissociation between the two types of reserve, a finding for which further investigation for which clinical and behavioral measures would be warranted.

This study was carried out with twice as many representations in the bilingual network as opposed to the monolingual network. However, while this might be the case with coordinate bilinguals there may be large amounts of variability in the size of the conceptual store in bilinguals according to the circumstances under which the second language was learned. The results of the current study support research which only demonstrates an increased amount of cognitive reserve in immigrant bilinguals rather than non-immigrant bilinguals (Chertkow et al., 2010) since they would have learned their first language under entirely different circumstances and as such developed a much larger conceptual store due to cultural differences than their non-immigrant counterparts.

The theory underlying these findings may also be transferred over to the lexical level of monolingual and bilingual speakers. Rather than semantic and/or conceptual storage varying in size according to the circumstances under which the second language was learned, pure bilinguals should require close to double the amount of lexical storage than their monolingual counterparts. This can be carried out computationally using a recurrent neural network carrying out a word naming task in two languages. In doing so, this may confirm the increased recruitment of inhibitory processes at all
levels of bilingual speech production and go some way in explaining the lexical based deficits observed in verbal fluency tasks (Bialystok, 2008).

A novel analysis of the representational space was carried out on simple three layer networks portraying monolingual and bilingual speakers. Decreases in the separation of representations were observed for bilingual speakers with lower levels of BRC. This would in turn increase the need for greater recruitment of inhibitory processes. Prolonged practice of speech in both languages would therefore lead to an increase of cognitive reserve in bilinguals.

References

Semantic Representations in Connectionist Networks

*Cognition*, 33(7), 1220–1234.


