Bilingualism: Language and Cognition 16 (2), 2013, 325–342 © Cambridge University Press 2012 doi:10.1017/S1366728912000533

A computational account of bilingual aphasia rehabilitation*

SWATHI KIRAN
Department of Speech and Hearing Sciences, Boston University & Department of Communication Sciences and Disorders, The University of Texas at Austin
ULI GRASEMANN
Department of Computer Science, The University of Texas at Austin
CHALEECE SANDBERG
Department of Speech and Hearing Sciences, Boston University
RISTO MIIKKULAINEN
Department of Computer Science, The University of Texas at Austin

(Received: November 1, 2011; final revision received: June 25, 2012; accepted: July 12, 2012; first published online 22 October 2012)

Current research on bilingual aphasia highlights the paucity in recommendations for optimal rehabilitation for bilingual aphasic patients (Edmonds & Kiran, 2006; Roberts & Kiran, 2007). In this paper, we have developed a computational model to simulate an English–Spanish bilingual language system in which language representations can vary by age of acquisition (AoA) and relative proficiency in the two languages to model individual participants. This model is subsequently lesioned by varying connection strengths between the semantic and phonological networks and retrained based on individual patient demographic information to evaluate whether or not the model's prediction of rehabilitation matches the actual treatment outcome. In most cases the model comes close to the target performance subsequent to language therapy in the language trained, indicating the validity of this model in simulating rehabilitation of naming impairment in bilingual aphasia. Additionally, the amount of cross-language transfer is limited both in the patient performance and in the model's predictions and is dependent on that specific patient's AoA, language exposure and language impairment. It also suggests how well alternative treatment scenarios would have fared, including some cases where the alternative would have done better. Overall, the study suggests how computational modeling could be used in the future to design customized treatment recipes that result in better recovery than is currently possible.

Keywords: computational modeling, rehabilitation, bilingual aphasia, generalization

1. Introduction

Current research has only begun to inform us about the nature of language impairment in bilingual aphasia. Due to the variety of complicating issues associated with bilingual aphasia, there has been insufficient research into this topic (Lorenzen & Murray, 2008) although attention to the nature of language impairment in bilingual aphasia is increasing due to the practical demands of serving this clinical population (Abutalebi, Rosa, Tettamanti, Green & Cappa, 2009; Green, Grogan, Crinion, Ali,Sutton & Price, 2010; Green, Ruffle, Grogan, Ali, Ramsden, Schofield, Leff, Crinion & Price, 2011; Kohnert, 2004; Laganaro, Di Pietro & Schnider, 2006; Meinzer, Obleser, Flaisch, Eulitz & Rockstroh, 2007; Moretti, Bava, Torre, Antonello, Zorzon, Zivadinov & Cazzato, 2001; Tschirren, Laganaro, Michel, Martory, Di Pietro, Abutalebi & Annoni, 2011). In a recent meta-analysis of treatment studies examining rehabilitation of bilingual aphasia, Faroqi-Shah, Frymark, Mullen & Wang (2010) reviewed 14 treatment studies in which language therapy was provided to bilingual individuals with aphasia. Of these, only one study (Junque, Vendrell, Vendrell-Brucet & Tobena, 1989) had a large subject pool (30 participants); all others were single or multiple case studies. Nonetheless, Faroqi-Shah and colleagues found that training patients in L2 resulted in improved treatment outcomes in the treated language, and further, more than half the participants across the studies showed cross-language generalization. Faroqi-Shah and colleagues also found that age of acquisition (AoA) and time post-stroke did not specifically influence the outcomes of treatment. However, the authors noted

* The computational component and portion of the treatment component of this research was supported by NIDCD # R21DC009446 to the first and last author. Also, a Clinical Research Grant from American Speech Language Hearing Foundation to the first author supported another component of the treatment project. The authors would like to thank the reviewers for their valuable comments and Danielle Tsibulskly, Anne Alvarez and Ellen Kester for their assistance in data collection.

Address for correspondence:
Swathi Kiran, Speech Language and Hearing Sciences, Boston University Sargent College, 635 Commonwealth Ave., Boston, MA 02215, USA
kirans@bu.edu

several other confounding variables across the studies that limit the conclusions that can be drawn regarding the effectiveness of rehabilitation of language recovery bilingual aphasia. One aspect of this review is that some of the studies reviewed are primarily naming therapies, others aimed at improving sentence production and still others more globally directed at improving communication abilities, hence introducing another level of variability across studies. A more recent study by Croft, Marshall, Pring & Hardwick (2011), not included in this meta-analysis, examined five Bengali–English individuals with aphasia who received treatment in the two languages consecutively. Four of the five patients benefited from the naming therapy and three patients showed cross-language generalization from the dominant to the less dominant language.

Although several studies report positive treatment outcomes, the factors that influence treatment outcomes are not well understood. Static factors, such as pre-stroke language state, the etiology of aphasia, and level of impairment between the two languages as well as dynamic factors, such as treatment methodology, and current language exposure, add to the complicated portrait of bilingual aphasia rehabilitation.

Our own work in this area of bilingual aphasia rehabilitation has shed light on the complex interaction between language exposure, level of aphasia severity, and language therapy outcomes. Specifically, in the first study systematically examining cross-language generalization using a theoretically motivated framework, Edmonds and Kiran (2006) conducted a semantically-based treatment to improve naming in three English–Spanish bilingual individuals with aphasia. They examined acquisition of trained items and generalization to untrained semantically related items in the trained language as well as translations of the trained and untrained items in the untrained language. Results showed within- and across-language effects on generalization that were related to pre-stroke language proficiencies. One participant, who claimed equal proficiency in English and Spanish, exhibited within-language generalization in the trained language (Spanish) and some cross-language generalization to the untrained language (English). The other two patients, who reported that English was their stronger language pre-stroke, showed cross-language generalization from the trained language (Spanish) to the untrained language (English) but no within-language generalization (to related words in Spanish).

In a follow-up study, Kiran and Roberts (2010) administered the same semantic treatment in four English–Spanish and English–French individuals with aphasia and measured within-language generalization and cross-language generalization. One Spanish–English patient was pre-morbidly proficient in both languages but severely impaired in naming in both languages. The second Spanish–English patient was pre-morbidly less proficient in Spanish and post-morbidly more impaired in Spanish. One French–English patient was pre-morbidly more proficient in French but equally impaired in both languages whereas the other was pre-morbidly proficient and post-morbidly impaired in two languages. All patients improved their naming of the trained items in the trained language; two of the four patients showed within-language generalization to semantically related items. Cross-language generalization, however, occurred for only one French–English patient. While these results suggested that performance in all four patients was influenced by factors including each patient's pre-stroke language proficiency, AoA, post-stroke level of language impairment, and type and severity of aphasia (Kiran & Roberts, 2010); no clear trends emerged regarding the potential interaction between these factors. Importantly, this replication of the Edmonds and Kiran (2006) study used the same stimuli and treatment approach, but resulted in different rehabilitation outcomes. These two studies highlight the inherent problem of variability in bilingual aphasia research, and this is exactly the kind of situation where a computational model can advance the field.

One obvious approach to testing the applicability of language rehabilitation in bilingual aphasia is to conduct a large scale examination of the relationship between language impairment, languages being tested, the relative proficiency of each language, and the etiology of language impairment in an otherwise homogenous population of bilingual aphasics. However, there are several obstacles to this approach because pre-stroke language proficiency cannot be empirically assessed in patients with aphasia; it is usually collected as an informal survey after the stroke has occurred. Additionally, the presumed relationship between lesion, deficit, and behavioral outcomes is contentious even in the field of monolingual aphasia. That is, it is still unclear whether a specific lesion in a certain brain region in the left hemisphere results in a specific linguistic deficit that can be observed using behavioral measures (Caplan, 2004), thereby severely limiting the predictions one can make about treatment outcomes. More importantly, our limited understanding of language impairment and recovery in bilingual aphasia is complicated by the imperfect, but highly connected relationship between AoA, pre-stroke proficiency/use and impairment (Kiran & Iakupova, 2011; Kiran & Roberts, 2012; Roberts & Kiran, 2007).

An innovative approach to addressing rehabilitation in bilingual aphasia is to use computational modeling to complement our understanding of how language recovery occurs in a bilingual individual with aphasia. Although the physiological structure and location of the lexicon in the brain are still open to some debate, converging evidence from imaging, psycholinguistic, computational, and lesion studies suggests that the lexicon is laid out

as one or several topographic maps, where concepts are organized according to some measure of similarity (Caramazza, Hillis, Leek & Miozzo, 1994; Farah & Wallace, 1992; Spitzer, Kischka, Guckel, Bellemann, Kammer, Seyyedi, Weisbrod, Schwartz & Brix, 1998). Over the last twenty years, connectionist and dynamical systems approaches have made remarkable contributions to our understanding of the mechanisms of impairment of language and cognition (Baron, Hanley, Dell & Kay, 2008; Dell, Schwartz, Martin, Saffran & Gagnon, 2000; Foygel & Dell, 2000; Plaut, 1996; Schwartz & Brecher, 2000; Welbourne & Lambon Ralph, 2005). Computational modeling has also been useful in understanding the effect of the nature of the lesion on brain plasticity (Keidel, Welbourne & Lambon Ralph, 2010; Reggia, Gittens & Chhabra, 2000; Welbourne & Lambon Ralph, 2007). However, apart from one study (Plaut, 1996), computational modeling has not been exploited in predicting naming treatment outcomes for stroke patients. In Plaut's (1996) study, a computer network was trained to recognize a set of artificial typical and atypical words (set of binary values). Once the training was completed, the network was lesioned and retrained on either the typical examples or the atypical examples. Plaut found that retraining atypical examples resulted in improvements in recognition of typical items as well. Training typical items, however, only improved the performance of trained items while performance of atypical words deteriorated during treatment. This computational simulation work was instrumental in catalyzing a series of studies that reported the beneficial effect of training atypical examples versus typical examples within categories in individuals with aphasia with naming deficits (Kiran, 2008; Kiran & Johnson, 2008; Kiran, Sandberg & Sebastian, 2011; Kiran & Thompson, 2003) and the development of the Complexity Account of Treatment Efficacy for treatment of naming deficits in aphasia (Kiran, 2007).

The purpose of the present study was to develop and examine a computational account, based on self-organizing maps, of bilingual naming deficits in aphasia and the subsequent rehabilitation outcome patterns in individuals with naming deficits. Self-organizing maps (SOMs; Kohonen, 2001) are topographical structures, and are therefore a natural tool to build simulations of the lexicon. SOM-based models have been used recently to understand how ambiguity is processed in the lexicon (Miikkulainen, 1993), how lexical processing breaks down in dyslexia (Miikkulainen, 1997) and how the lexicon is acquired during development (Li, Zhao & MacWhinney, 2007). SOMs form the basic building blocks of the bilingual model used in this work, called DISLEX (Miikkulainen, 1993, 1997). DISLEX consists of separate self-organizing maps for semantics and for orthographic and phonological lexica in different languages, linked with associative connections. Because words are laid out in a map, the model can account for aphasic and dyslexic impairments where categories of words are lost or impaired. Notably, Ping Li and colleagues have implemented a bilingual developmental version of this model, called DEVLEX, which is trained with gradually more words across the two languages and thereby accounts for a range of phenomena in lexical acquisition, including effects of lexical categories such as representation of nouns/verbs, word frequency, word length, and word density (Li, 2009; Li, Farkas & MacWhinney, 2004; Li & Green, 2007; Li et al., 2007). DEVLEX is primarily a computational developmental model of normal language performance, whereas the model we describe here, the bilingual DISLEX model, is a model of adult language that can be used to study language breakdown and recovery. It can be lesioned by deleting units and connections and by adding noise to the connections. It therefore forms a promising starting point for modeling bilingual lexical processing, impairment, and recovery, as we outline here.

We have previously described the conceptual architecture of the computational model DISLEX (Grasemann, Sandberg, Kiran & Miikkulainen, 2011; Miikkulainen & Kiran, 2009), but we highlight the key points here. The computational model follows the logic of the revised hierarchical model proposed by Kroll and Stewart (1994) for the two languages (L1 and L2) because it can account for both AoA and varying levels of proficiency through the differential connection strengths. The organization of the three maps and the associations between them are learned simultaneously. Input symbols are presented to two of the maps at the same time, resulting in activations on both maps. Each individual map adapts to the new input using standard SOM training with a Gaussian neighborhood. Additionally, associative connections between the maps are adapted based on Hebbian learning, i.e., by strengthening those connections that link active units, and normalizing all connections of each unit. As a result of this learning process, when a word is presented to the semantic map, the resulting activation is propagated via the associative connections to the phonetic maps, and vice versa. In this way, DISLEX can model both comprehension and production in both languages (Miikkulainen & Kiran, 2009). As described in this paper, one of the features of the model is the ability to model second language acquisition at different ages and differences in proficiency between L1 and L2 in order to represent the wide range of combinations of language dominance and proficiency in individual bilingual individuals.

As a next step, we examined the validity in simulating individual language performance by using AoA and exposure as training parameters with the goal of matching DISLEX's performance with that of human pre-stroke English and Spanish performance (Grasemann et al.,

Table 1. *Demographic information for seventeen participants with aphasia who were enrolled in the therapy experiment. Details are explained in the text. ND indicates no data was available for that patient; eval = evaluation; mod = moderate.*

		AoA		Pre-stroke language exposure		Post-stroke language exposure		Education history		Self-rating	
Participant	Age at eval	English	Spanish	English	Spanish	English	Spanish	English	Spanish	English	Spanish
UTBA07	56	0 native	0 native	ND high	ND high	ND mod	ND high	1.00	0.00	1.00	0.69
UTBA09	56	5 early	0 native	0.62 high	0.37 low	0.57 mod	0.42 mod	0.77	0.22	1.00	0.81
UTBA16	53	0 native	0 native	0.61 high	0.38 low	0.61 high	0.38 low	0.66	0.33	0.94	0.74
BUBA07	65	45 v. late	0 native	0.09 low	0.90 high	0.01 low	0.98 high	0.00	1.00	0.32	1.00
UTBA01	53	0 native	0 native	0.75 high	0.25 low	0.93 high	0.06 low	1.00	0.00	1.00	0.40
UTBA11	87	11 late	0 native	ND mod	ND high	ND high	ND high	ND	ND	0.97	1.00
UTBA17	52	6 early	0 native	0.65 high	0.34 low	0.55 mod	0.44 mod	0.58	0.41	1.00	1.00
UTBA19	75	27 late	0 native	0.16 low	0.83 high	0.14 low	0.85 high	0.00	1.00	0.20	1.00
UTBA22	41	18 late	0 native	0.09 low	0.90 high	0.37 low	0.62 high	0.00	1.00	0.34	0.94
UTBA23	41	9 late	0 native	0.32 low	0.67 high	0.28 low	0.71 high	0.22	0.77	0.65	0.94
BUBA01	44	19 late	0 native	0.27 low	0.72 high	0.21 low	0.78 high	0.00	1.00	0.88	0.88
BUBA04	37	9 late	0 native	0.73 high	0.26 low	0.66 high	0.33 low	1.00	0.00	1.00	0.48
BUBA12	33	12 late	0 native	0.28 low	0.71 high	0.45 mod	0.54 mod	0.27	0.72	0.80	1.00
UTBA02	54	21 late	0 native	0.30 low	0.69 high	0.50 mod	0.50 mod	0.33	0.66	0.89	1.00
UTBA18	74	17 late	0 native	0.40 mod	0.59 mod	0.00 low	1.00 high	0.25	0.75	1.00	1.00
UTBA20	85	69 v. late	0 native	0.05 low	0.94 high	0.11 low	0.88 high	ND	ND	0.00	0.00
UTBA21	88	5 early	0 native	0.71 high	0.28 low	0.98 high	0.01 low	1.00	0.00	ND	ND

2011) for 18 bilingual individuals with aphasia. Pre-morbid levels of naming performance (i.e., self-rated pre-morbid proficiency), AoA, and relative exposure to Spanish vs. English were collected from all patients, and were used to determine the way in which each patient model was trained. Then, the model was trained with the same relative exposure as patients to both languages and AoAs were simulated by variably delaying L2 training. Results showed that in most cases (80%), the model is able to match the pre-morbid language performance (in addition to AoA and relative exposure) of patients well.

This preliminary work sets the foundation for the work that will be described here. Briefly, in the present paper, we damage the model in order to simulate naming impairments and then retrain the model in one language in order to understand the nature of cross-language generalization. The starting point for rehabilitation is a model that has been fit to an individual patient by adjusting AoA, pre-stroke proficiency, and post-stroke impairment of L1 and L2. That is, each individual patient is represented by a separate, individual instance of the model; all of these models have the same architecture but are initialized to fit the particular patient's language history and impairment profile. The human behavioral treatment study and the computational model share identical input stimuli (i.e., words used for treatment) as well as similar outcome measures (correct naming percentage in each language). The stimuli employed in this study have been used in previous treatment studies (Edmonds & Kiran, 2006; Kiran & Roberts, 2010). In rehabilitation training, the model is presented with selected word-meaning pairs from L1 or L2, and it will continue self-organizing using the same mechanisms as during initial training. Even though only one language is trained, the result of the rehabilitation is measured in the final performance of the model in both L1 and in L2. We then evaluate the fit of the model to actual patient data as well as estimate the predictive power of the model in terms of the optimal language to use for rehabilitation.

2. Methods

2.1 Part I: Behavioral treatment study

Participant demographic information

Seventeen bilingual English–Spanish aphasic patients presenting with naming deficits and concurrent lexical/semantic impairment secondary to a single left hemisphere stroke were involved in treatment (see Table 1 for demographic information). Five of these individuals have been described previously (Edmonds & Kiran, 2006; Kiran & Roberts, 2010) whereas the remaining

12 patients were enrolled in treatment concurrent with the computational simulation portion.[1] All participants (except one individual) experienced a single, unilateral cerebral vascular event (stroke) in the distribution of the left middle cerebral artery or posterior cerebral artery at least 6 months prior to initiation of the experiment. BUBA04 suffered a gunshot wound to the left hemisphere. For all participants, L1 was Spanish and L2 was English, although many of them were English dominant, resulting from higher exposure to English.

Assessment of language proficiency

For all participants, we obtained measures of language AoA, use, and proficiency by administering a questionnaire that obtained the following information: the age of acquisition for each language; a proportion of language exposure during the entire lifetime prior to the stroke; educational history in terms of both the language of instruction and language used by peers; the time spent conversing in each language after the stroke (post-stroke exposure), and a self-rating of pre-stroke proficiency in each language. For pre-stroke language exposure, a weighted average of the proportion of exposure across the lifespan in hearing, speaking, and reading domains was obtained for each language. Likewise, a weighted average of the exposure in each language calculated hour by hour during a typical weekday and typical weekend score reflected the proportion of post-stroke language exposure in each language. Finally, an average proportion score in terms of the participant's ability to speak and understand the language in formal and informal situations in each language reflected participants' perception of their own language proficiency. Details regarding the participants' language backgrounds are listed in Table 1.

Assessment of language impairment in English and Spanish

The Boston Naming Test was administered in both English and Spanish to indicate eligibility for treatment (Kaplan, Goodglass & Weintraub, 2001; Kohnert, Hernandez & Bates, 1998). While all participants performed below 75% accuracy in both their languages there was a range of naming impairments that provided an indication regarding the level of severity of impairment for each patient. Performance on the BNT in English and Spanish are listed for each participant in Table 2.

Treatment stimuli

For all patients, target treatment items were selected from a corpus of 300 nouns gathered from our previous treatment studies for word finding in aphasia in both monolingual and bilingual populations (Edmonds & Kiran, 2006; Kiran, 2007, 2008; Kiran & Bassetto, 2008; Kiran & Johnson, 2008; Kiran & Thompson, 2003). Rather than proceeding with a pre-chosen set, target items for each participant were chosen based on a confrontation naming pre-test and hence, stimuli trained during treatment differed for each participant. Thus, for each participant, six individualized stimulus sets were created: English set 1 (e.g., *apple*), Spanish set 1 (e.g., *manzana*), English set 2 (e.g., *orange*), Spanish set 2 (e.g., *naranja*), English unrelated/control set (e.g., *boat*), Spanish unrelated/control set (e.g., *vaca* (*cow*)). Only data for English set 1 and Spanish set 1 are reported in this study as the data collected for the semantically related/control sets in each of the languages dilutes the focus of this paper. The stimuli used in treatment are described in detail elsewhere (Edmonds & Kiran, 2004) and again in Part II. All word pairs were category coordinates (e.g., *horse* and *sheep*). Cognates (e.g., *elephant* and *elefante*) and words with at least 50% phonetic similarity (e.g., *cat* and *gato*) were eliminated. The lists were balanced for average frequency (Bates, D'Amico, Jacobsen, Szekely, Andonova, Devescovi, Herron, Lu, Pechmann, Pleh, Wicha, Federmeier, Gerdjikova, Gutierrez, Hung, Hsu, Iyer, Kohnert, Mehotcheva, Orozco-Figueroa, Tzeng & Tzeng, 2003; Frances & Kucera, 1982) and number of syllables.

Additionally, 261 binary semantic features (e.g., *can fly*, *is a container*, *can be used as a weapon*) were assembled for items across categories.[2] These features were then translated into Spanish by a Spanish–English bilingual research assistant and encoded for each of the 300 words, assigning the value 1 when it was a feature of that word and a value of 0 when it was not a feature of that word. Every picture, therefore, had at least 15 features that were applicable to that particular example.

General treatment procedures

To facilitate access to naming of trained items, a semantic-feature–based treatment (Edmonds & Kiran, 2006) was implemented. All participants received treatment two times per week. Participants performed five treatment steps that emphasize semantic feature attributes of a particular example. First, they were required to label/name the example. Then, the picture was placed in front of the patient, who was asked to choose five features (from a field of 10) that belong to that word and were (a) a superordinate label (e.g., belongs to), (b) a function (e.g., is used for), (c)

[1] UTBA01, UTBA02, UTBA07 were reported in Edmonds and Kiran (2006) and UTBA09 and UTBA11 were reported in Kiran and Roberts (2010).

[2] These features were developed as part of a broader project (Edmonds & Kiran, 2006; Kiran, 2003, 2007) to norm semantic features across semantic categories. Across these projects, normal individuals made judgments about the applicability of semantic attributes for each individual item in each category.

Table 2. *Details are provided for each participant's pre-treatment Boston Naming Test (BNT) scores, average naming scores during baseline, and the language trained for each participant. Also reported are the effect sizes (ES) in the trained language and untrained language and the cross-correlation coefficient (r) between the patient time-series data and the model time-series data.*

	Pre-treatment BNT		Average baseline						
Participant	English	Spanish	English	Spanish	Language trained	Trained language ES	Untrained language ES	Cross- correlation *r* with model trained language	Cross-correlation *r* with model untrained language
UTBA07	0.16	0.15	0.27	0.15	Spanish	12.41	3.11	0.86	0.76
UTBA09	0.56	0.10	0.28	0.08	Spanish	10.97	2.07	0.79	0.59
UTBA16	0.06	0.05	0.01	0.01	English	6.82	0.82	0.67	0.50
BUBA07	0.00	0.15	0.06	0.06	English	2.88	4.08	0.71	0.53
UTBA01	0.00	0.00	0.03	0.03	English	12.70	0.577	0.94	0.39
UTBA11	0.08	0.05	0.04	0.02	English	14.90	1.14	0.89	0.45
UTBA17	0.52	0.08	0.31	0.05	English	5.31	1.19	0.90	0.36
UTBA19	0.03	0.46	0.01	0.43	English	4.83	1.12	0.52	0.44
UTBA22	0.05	0.46	0.40	0.00	Spanish	12.72	1.88	0.90	0.49
UTBA23	0.00	0.02	0.03	0.05	Spanish	13.83	10.68	0.69	0.42
BUBA01	0.36	0.43	0.36	0.52	English	4.91	1.41	0.91	–0.47
BUBA04	0.58	0.11	0.12	0.07	Spanish	16.5	2.61	0.94	–0.51
BUBA12	0.00	0.00	0.00	0.00	English	8.16	0	0.92	0.23
UTBA02	0.43	0.40	0.05	0.05	Spanish	11.07	4.94	0.28	0.50
UTBA18	0.28	0.32	0.01	0.21	Spanish	24.82	1.73	0.43	0.55
UTBA20	0.00	0.00	0.00	0.00	Spanish	0.00	0.00	ND	ND
UTBA21	0.01	0.00	0.00	0.00	English	0.00	0.00	ND	ND

a characteristic (e.g., has/is), (d) a physical attribute (e.g., is made of/appears), and (e) a location (e.g., is found). After these were chosen, the patient was asked to generate an association and a non-association (e.g., reminds me of/doesn't remind me of). Following this, the participant was asked yes/no questions about these features and was required to accept or reject these and other features as being applicable to the target example. Finally, the picture was presented again, and the participant was required to name the target examples.

Baseline and treatment probe task

Picture naming was used to examine lexical retrieval. Prior to treatment, three to five naming probes were given to establish a baseline; the specific number of baseline probes was varied across participants. No feedback was provided during the probes. During treatment, naming probes with the same stimuli as those presented during baseline were given weekly. Responses to naming probes, coded in the same way as in baseline, served as the primary dependent measure. Treatment was discontinued when naming accuracy reached 8/10 items on the trained items in two consecutive sessions or when 20 sessions were completed.

Data analysis

The extent to which changes from baseline to treatment phases are statistically reliable was determined by calculating effect sizes (ES). Effect sizes are calculated comparing the mean of all data points in the post-treatment phase relative to the baseline mean divided by the standard deviation of baseline (Beeson & Robey, 2006).

Treatment results

Results for the seventeen patients are summarized in Table 2 and illustrate the language that was trained for each patient and the effect sizes for the trained language set and the untrained language set. Based on comparable naming treatment studies in monolingual aphasia, an ES of 4.0 was considered small, 7.0 was considered medium and 10.1 was considered large (Beeson & Robey, 2006) for the trained language set.

2.2 Part II: Computational modeling

Architecture of the model

Figure 1 shows the basic architecture of the bilingual DISLEX model. Its three main components are self-organizing maps: one for word meanings and one

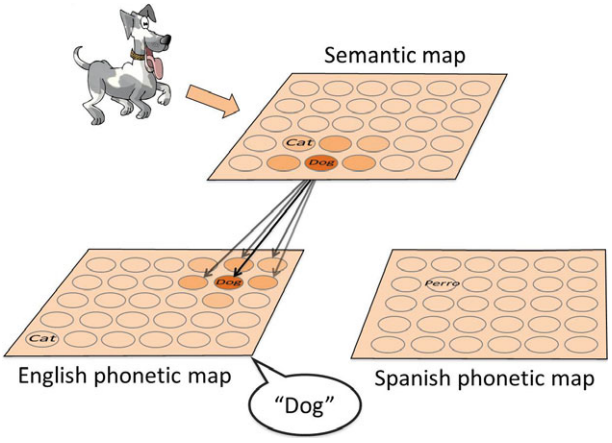

Figure 1. (Colour online) Schematic representation of the architecture of the bilingual DISLEX model adapted from Kroll & Stewart's (1994) theoretical model.

each for the corresponding phonetic symbols in L1 and L2. Each pair of maps is linked by directional associative connections that enable network activation to flow between maps, allowing the model to translate between alternative semantic and phonetic representations of a word.

Each self-organizing map is a two-dimensional grid of computational units trained to represent a set of symbols. The symbols are vectors of real numbers between 0 and 1, representing either semantic or phonetic features of a word. During training, such vectors are presented to the map one at a time, and each unit computes the Euclidean distance d between its weight vector w and the symbol representation v:

$$d\sqrt{\sum_k (w_k - v_k)^2}. \tag{1}$$

The unit with the smallest distance (unit (r,s)) is then found, and the weights of that unit and those in its neighborhood (units (i,j)) are adapted towards the input vector:

$$w'_{k,ij} = w_{k,ij} + \alpha(v_k - w_{k,ij})h_{rs,ij}, \tag{2}$$

where α is the learning rate and $h_{rs,ij}$ is a function defining the neighborhood (usually a Gaussian). This process has two effects: the weight vectors become representations of the symbol vectors, and the neighboring weight vectors become similar. Over successive presentations of each symbol, the array of units then learns to represent the space of symbols in the language, creating a two-dimensional layout of that space where units that are close to each other on the map represent words that are similar either semantically (in the semantic map) or phonetically (in the phonetic maps for L1 and L2).

Associations between maps are learned at the same time as the maps are organized. Two corresponding representations for the same word (e.g., semantic and L1) are presented at the same time, resulting in activations on both maps. Associative connections between the maps are then adapted based on Hebbian learning, i.e., by strengthening those connections that link active units and normalizing all connections of each unit:

$$a'_{ij,mn} = \frac{a_{ij,mn} + a\eta_{ij}\eta_{mn}}{\sum_{uv}(a_{ij,uv} + a\eta_{ij}\eta_{mn})}, \tag{3}$$

where $a_{ij,mn}$ is the weight on the associative connection from unit (i,j) in one map to unit (m,n) in the other map and η_{ij} is the activation of the unit.

As a result of this learning process, when a concept is presented to the semantic map, its associated phonetic representations in the L1 and L2 maps are activated. Conversely, when the L1 or L2 map is presented with the phonetic representation of a word, the resulting activation is propagated to the semantic map. DISLEX therefore models both comprehension and production in the lexicon. Note that in the latter case, activation can also be propagated between L1 and L2, since the L1 and L2 maps have direct connections between them as well. Also, the connections between L1 and L2 create a possible alternative path for the flow of activation between the semantic map (S) and either phonetic map. For instance, activation may flow from S → L1 directly, but also from S → L2 → L1. Such indirect flow of activation between maps can potentially simulate and explain how treatment in one language can benefit the other. For example, if the lexicon is presented with input symbols for S and L1, those maps and the connections between them can be adapted using the method described above. However, in addition, the L2 map is activated indirectly, and that activation can be used to train its associative connections as well. How beneficial this "indirect training" is for L2 may depend on several factors, including the strength and quality of the connections between L1 and L2. This model of cross-language transfer forms the basis of the treatment simulations described below.

Stimuli

The input data used for the computational experiments consisted of a corpus of 300 nouns that were taken from the treatment experiment described above. Representations for all words in terms of semantic and phonetic feature vectors were developed and used to train the model. Semantic representations were vectors of the 261 binary features described in the treatment stimuli section above. The numerical representations of the features were used directly to train the semantic map of the lexicon. In order to develop the phonetic representations, both English and Spanish words were transcribed phonetically by a

Spanish–English bilingual research assistant using the International Phonetic Alphabet (IPA). The transcriptions were then translated into numerical input patterns using a method based on previous work by Miikkulainen (1997), and similar to the one used in PatPho (Li & McWhinney, 2002), a pattern generator designed to create feature-based phonological representations. Following Ladefoged (1982), each individual phoneme was first encoded using four features: height, tongue position, length, and rounded-ness for vowels, and place of articulation, manner of articulation, voiced-ness, and lateralization for consonants. These features are not complete descriptions of each possible phoneme, but were more than sufficient to distinguish the English and Spanish input words and to capture phonetic similarities between them. Words were split manually into spoken syllables, and the phonemes of each syllable were inserted into a template vector consisting of two consonants followed by two vowels, followed again by two consonants (CCVVCC), resulting in a feature vector of length 24 (six phonemes with four phonetic features each). If a syllable had only one vowel, that vowel was inserted twice. Single consonants in the onset or coda of a syllable were also repeated; syllables with no consonants in the beginning or end were encoded by inserting zeros. Triple consonants were encoded by averaging the features for the second and third. In contrast to PatPho, representations of an entire word were not left- or right-justified concatenations of syllables. Instead, all words were padded from the left such that the syllable that carried the primary stress lined up for all words. This encoding method solves several problems (e.g., capturing the similarity between the same word with or without a prefix), but leads to relatively long overall word representations. The final length of feature vectors was 120 for English words (5 syllables $\times$ 24 features per syllable), and 168 (7×24) for their Spanish translations.

Training the model

The semantic and phonetic input data were then used to train Spanish–English bilingual DISLEX models. The semantic and phonetic maps were a grid of 30×40 neurons, and were fully connected to each other by directional associative connections. Learning rates, both for maps and associations, were set to 0.25 during initial training. The number of training epochs varied depending on the target age being modeled. Four age levels were simulated in order to reflect the ages of the patient population: young ($<$ 45 years; 800 epochs), middle-aged ($<$ 60 years; 1100 epochs), old ($\leq$ 80 years; 1400 epochs), and very old ($>$ 80 years; 1700 epochs). The number of randomly selected English and Spanish words trained during each epoch was controlled by two exposure parameters (e.g., an English exposure of .4 meant that every word had a 40% chance of being used in training the English language part of the model during each epoch). The variance of the Gaussian neighborhood was initially 3, and decreased exponentially with a half-life of 155 training epochs. These parameter settings were determined empirically.

The effects of second-language AoA were simulated in the model by delaying L2 training. Since all patients were native Spanish speakers, initially only the Spanish phonetic map, the semantic map, and the connections between the two were trained. After an appropriate delay, all three maps and all sets of associative connections were then trained for the remaining epochs. The appropriate training delays for simulating second-language AoA effects were determined based on a few constraints. First, given enough exposure, humans are often able to achieve near-native performance even for relatively late AoA (age 18+ years), so the same needed to be possible in the model. Initial experiments showed that naming performance in the model begins to decline considerably, even for high exposure, when training is delayed until epoch 260 or later. This point in training was therefore used as the equivalent of AoA = 20 years. A second constraint is that, if a second language is acquired early, performance should not differ significantly from a native language. In the model, this is the case if training is delayed until no later than epoch 175, which was therefore used as the equivalent of AoA = 10 years. These two fixed training delays were used to calibrate the simulation of AoA effects in the model. For the present study, four AoA levels were simulated: Native (training epoch 0), early (epoch 100), late (epoch 180), and very late (epoch 300), corresponding to AoAs of 0, 8, 18, and 30 years.

Current language dominance was simulated by varying the exposure parameters for each language. Exposure parameters were set such that the naming performance approximated the performance observed in the control population of the human aphasia study described earlier. The model's English performance thus varied between 75% and 98%; Spanish performance varied between 70% and 97%. Lower average Spanish performance made it necessary to use lower Spanish than English exposure in general. Four levels of relative exposure (Spanish vs. English) were simulated: low Spanish/high English, medium Spanish/moderately high English, moderately high Spanish/medium English, and high Spanish/low English. The corresponding low-to-high exposure parameters were .04, .125, .2, .32 for Spanish, and .2, .25, .325, .45 for English. These values reflect the fact that English is generally the more dominant (though not native) language in the patient group. Based on the possible combinations of four levels of English AoA, four levels of current language dominance, and four levels of patient age, a range of 64 (4^3) "hypothetical patient models" were then created to simulate the range of possible pre-stroke scenarios. For each scenario, five instances of the DISLEX model were trained using

different randomized initial network connections. These models were then used as starting points to simulate stroke damage and to investigate the influence of treatment language and pre-stroke conditions on treatment outcome. Note that while it would be straightforward to map AoA to a continuously varying training delay, the patients' AoA estimates are too noisy for such a mapping to be useful. The fixed points make sense because differences between the broad categories of early and late AoA categories are generally reliable in the patient data leading to reliable conclusions about the two groups.

Lesioning the model

In order to simulate pre-treatment starting points of the patients in the human aphasia study, each individual patient was first matched to one of the 64 scenarios (described in the previous section) according to age, self-reported AoA, and current exposure to English and Spanish. The DISLEX models corresponding to each patient were then damaged ("lesioned") to varying degrees to reflect the patient's known naming impairment. Our general approach to lesioning the model (i.e., to simulate stroke damage and the following naming impairment) was guided by published work in monolingual aphasia (Baron et al., 2008; Dell et al., 2000; Foygel & Dell, 2000; Howard, Nickels, Coltheart & Cole-Virtue, 2006; Miikkulainen, 1993, 1997; Plaut, 1996; Schwartz & Brecher, 2000; Welbourne & Lambon Ralph, 2005). Pertinent to the present study, one set of computational models that describe naming deficits that can arise either from incorrect/incomplete activation of semantic nodes or phonological nodes (Dell, Schwartz, Martin, Saffran & Gagnon, 1997; Foygel & Dell, 2000) or from a failure in the bi-directional link between them (Dell et al., 1997). Specifically, Foygel and Dell (2000) have suggested that introducing noise in the connection weights between the semantic-lexical (s-lesions) and lexical-phonological (p-lesions) levels in the interactive activation model simulates naming deficits in aphasia.

However, lesioning a bilingual model is slightly more complex as the precise location and effect of lesion can have consequences for whether the impairment is parallel or differential between the two languages. An appropriate way to damage the DISLEX model was chosen based on the following constraints. First, since patients can show differential language impairment, any lesion that affects only the semantic map was ruled out, since it is shared between both languages. Second, naming deficits in all patients were not limited to any specific category of words. The lesion must therefore be applied to one or more entire maps, or to one or more entire sets of connections between maps. Third, while naming deficits exist to some degree in the entire group of patients, comprehension was often relatively spared. Assuming that the same phonetic map is used for both production and comprehension of words, this rules out direct damage to one or both phonetic maps.

Given these constraints, the only remaining way to model the naming impairment of the patient group with a common type of lesion was damage to the associative connections between phonetic and semantic maps. Damage was applied by adding Gaussian noise with $\mu = 0$ to all connections from the semantic map to the phonetic maps. The amount of damage (the "lesion strength") in each case was adjusted by changing the σ of the noise. The lesion strength was adapted independently for each patient and each language such that the resulting impaired naming performance matched the pre-treatment aphasic performance of each patient. The patients' average performance on the pre-treatment naming baseline (see Table 2) was used to determine the target naming performance for the model. Damage was increased gradually and independently for each language until the average naming performance of the lesioned models matched the patient data within 1%.

Modeling rehabilitation

The lesioned models (five DISLEX models for each of the 17 individual patients that were treated for aphasia in the patient study) were then used to assess the model's ability to model and predict rehabilitation effects. Naming treatment was simulated by selecting a treatment language and retraining each model on the original input words in that language. Associative connections of the untreated language were also trained, using indirect activation in the way described above in the "Stimuli" section. In this manner, part of the model is always trained directly and deliberately, and another part indirectly and spontaneously. For each patient model, treatment was simulated once in English and once in Spanish, independent of the language that was actually used to treat the patient. Associative connections of the untreated language were also trained, using indirect activation in the way described above in the "Stimuli" section. The model thus provides both a prediction of treatment outcome in the patient (for the treatment language that was actually used) and a prediction of what would have happened if the other language had been chosen. One epoch of retraining was performed for each treatment session conducted with the actual patient, i.e., if a patient (e.g., BUBA01) received 10 treatment sessions, the model was accordingly trained for10 epochs. For all models, the learning rate was reduced uniformly to .01 to reflect the average effect of a single treatment session in the patient data. Exposure parameters for treatment were set to 1 for the treated and 0 for the untreated language.

After each simulated treatment session, the naming performance of the model was tested by presenting all 300 words to the semantic map, propagating the resulting activation to the phonetic maps, and examining

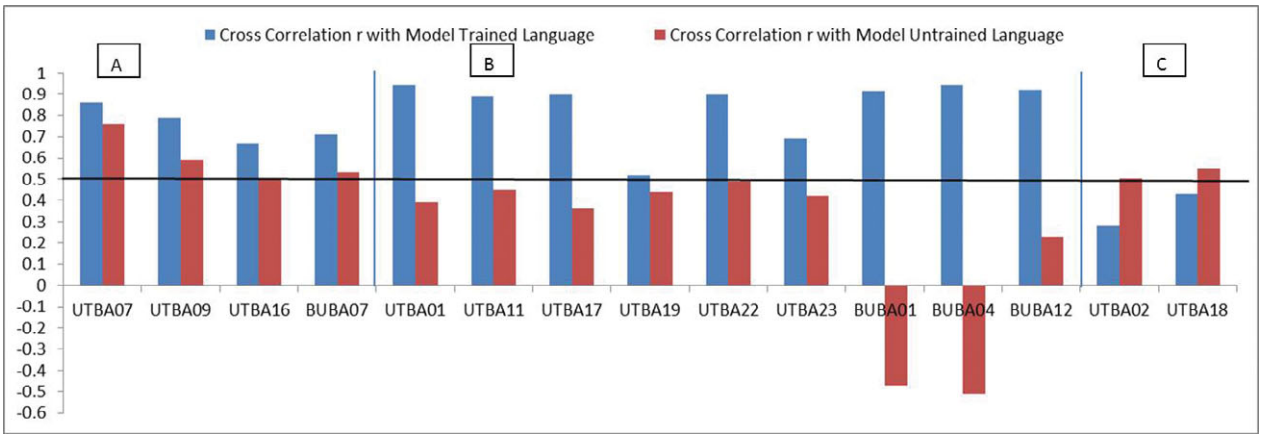

Figure 2. (Colour online) Summary representation of cross-correlation functions between patient performance and model performance for trained language (in blue) and untrained language (in red). Three distinct groups of participants emerged, in the first group (A), model matched patient performance for both the trained and untrained language, in the second group (B), model matched patient performance for the trained language only, and in the third group (C), model matched patient performance for untrained language better than the trained language.

the resulting activation. A word was scored as correctly named if the most active phonetic unit was the one representing the correct word. For each language, the percentage of semantic symbols correctly translated to their phonetic representation was reported as the model's naming performance. In this way, similar outcome measures (percentage of correctly named words in each language) were used to compare the treatment response of each model to that of the corresponding patient.

Comparison of patient and model performance

In order to evaluate the model's output relative to the patient data, the patient's actual behavioral treatment performance was compared to that of the lesioned DISLEX models. The extent of the match between the rehabilitation outcomes of patients and models was examined first by visually inspecting the slopes, and then by calculating cross-correlation coefficients using the autoregressive integrated moving average (ARIMA) procedure. Cross-correlation analysis models identify the relationship between two time series by examining coinciding changes over time (Box, Jenkins & Reinsel, 1994). Therefore, for each patient, we correlated the time series of language performance for each language with the model's corresponding language.

3. Results

Of the 17 patients, 14 showed improvements as a function of therapy in the trained language and three showed improvements in the untrained language, indicating cross-language transfer to the translations of the items trained. The patient and model performance are described for patients in Figure 2 and detailed results including cross-correlation functions are provided in Table 2. In each subsequent figure (Figures 3 and 4), the patient performance (white background) is compared with model performance (black background). For the patients, the pre-treatment baselines are reported before the first vertical line. For the model, the first data is computed as the average of the pre-treatment baselines for that patient. For each patient, the probes conducted during treatment are reported between the two vertical lines and the post-treatment probes (when administered) are reported after the second vertical line. The post-treatment averages are only relevant for the computation of the effect sizes. For the purposes of comparison with the computational model, only the data between the two vertical lines (i.e., training phase) is important. In what follows, we discuss the results in four subgroups of cases, one where the model accurately simulates both the trained and untrained language, a second group where the model accurately simulates only the trained language, a third group where the model simulates the untrained language only and a fourth group where the model is unable to match the patient performance sufficiently.

In the first group, for four patients (UTBA07, UTBA09, UTBA16, BUBA07) the model accurately simulates performance in both the trained and the untrained language. For instance, UTBA07 learned both languages as a native and exposure to both languages was high before stroke, indicating equivalence between the two languages. Both languages were moderately impaired and English exposure lessened after the stroke. Although UTBA07 was trained in Spanish, s/he shows greater performance in English during baseline and this performance steadily increases over time. Performance on the trained Spanish increases towards the end of the treatment. Interestingly,

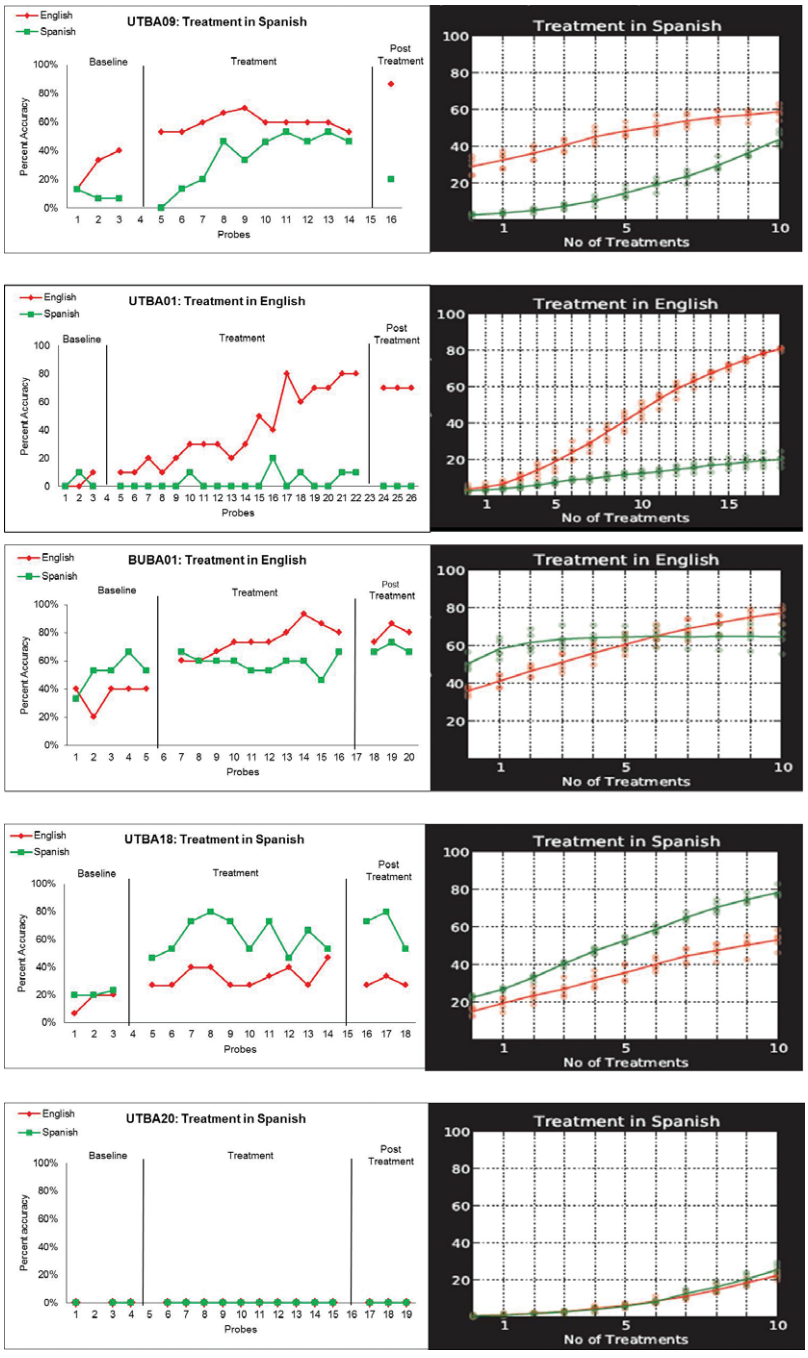

Figure 3. (Colour online) Representative samples of patient performance (white background) compared with model performance (black background). For the patients, the pre-treatment baselines are reported before the first vertical line, for the model the first data is computed as the average of the pre-treatment baselines for that patient. For each patient, the probes conducted during treatment are reported between the two vertical lines and the post-treatment probes (when administered) are reported after the second vertical line. (A) One representative patient where model output matches the trained language and the untrained language. (B) Two representative patients where model output matches the trained language. No changes are noted in the untrained language. (C) Sample of a patient and model performance where model matches performance for untrained language better than for the trained language. (D) Sample of a patient where model and patient performance do not change as a function of treatment.

the model adequately captures the changes in both languages as evidenced by the relatively high correlation between the model's performance and the patient's performance in the two languages (Spanish $r = .86$; English $r = .76$).

Similar to UTBA07, UTBA09 learned Spanish as a native and English early in life. Early exposure to English was high and exposure to Spanish was low; however, exposure after stroke was moderate in both languages. Impairment in English was moderate-severe (29% accuracy) but impairment in Spanish was severe. UTBA09 was trained in Spanish, which was the weaker language and performance in that language improved during the course of treatment; while performance on the stronger language improved initially and then remained stable. The model captures the performance in both languages very accurately (Spanish $r = .79$; English $r = .59$) (see Figure 3A). A third patient UTBA16 learned both languages as a native. Pre-stroke early exposure to English was high, but exposure to Spanish was low, and this pattern of exposure continued post-stroke. Impairment was equally severe in both languages. UTBA16 was trained in English, the stronger language. This patient shows improvement in the trained language but no changes in the untrained weaker language (Spanish). The model also predicts modest improvement in the trained language which matches the patient performance (English $r = .67$), and for improvements in the untrained language (Spanish $r = .50$). BUBA07 learned Spanish as a native and English late in life. Pre-stroke exposure to English was low, but exposure to Spanish was high, and this pattern of exposure continued post-stroke. Impairment was severe in both languages. BUBA07 was trained in English, the weaker language, and showed marginal improvements in both the trained English language and the untrained Spanish language. The model, however, over-predicts performance in the two languages for the same duration of treatment although the trends for the model and patient generally correlate well (English $r = .71$; Spanish $r = .53$). It appears that given more time in treatment, the patient very well could have achieved the same level of treatment outcome as the model predicts but this is a speculation at this point.

In the second subgroup of patients (UTBA01, UTBA17, BUBA04, UTBA11, UTBA22, BUBA01, BUBA12, UTBA19, UTBA23), the model accurately matches the patient output in the trained language but not for the untrained language. In several cases that follow, lower correlation coefficients for the untrained language are due to minimal change in both the model and the patient data that is indicative of no generalization to the untrained language (UTBA01, BUBA01, UTBA22, BUBA04). In other cases, the model predicts generalization to the untrained language whereas the patient does not show generalization to the untrained language (UTBA11, UTBA23). The cases that follow, however, are grouped by proficiency and language use. For instance, UTBA01 learned both languages as a native. Pre-stroke exposure to English was high, but exposure to Spanish was low, and this pattern of exposure continued post-stroke. Both languages were severely impaired (high impairment). UTBA01was trained in English and improved in English, with no improvement noted in the untrained Spanish (see Figure 3B). The model predicts the same outcome for English ($r = .94$), but predicts very small improvements in Spanish that were not observed ($r = .39$). Likewise, UTBA17 learned Spanish as a native and English early in life. Early exposure to English was high, but exposure to Spanish was low. Exposure after stroke was moderate in both languages. Impairment was severe in Spanish but moderate in English. UTBA17 was trained in English, which was the stronger language, and performance on this language improved remarkably during the course of treatment. Again, cross-language transfer was limited. The model's performance captures this differential re-learning and matches the performance for English well ($r = .90$), but predicts slightly more improvements in Spanish than is actually observed ($r = .36$). Like UTBA01 and UTBA17, BUBA04 learned Spanish as a native and English later in life. Pre-stroke exposure to English was high, but exposure to Spanish was low, and this pattern of exposure continued post-stroke. Impairment was severe in both languages. BUBA04 was trained in the weaker language, Spanish, which improved as a function of treatment. English, the stronger language, did not show much improvement by the end of treatment. Again, the model's performance closely mirrors the performance for the trained Spanish language ($r = .94$). Performance in English, however, shows slight improvements overall which is different from the initial increase followed by a decrease in performance observed for the patient ($r = -.51$).

The next five patients all learned English late in life and were stronger in Spanish prior to their stroke. Hence, UTBA11 learned Spanish as a native and English late in life. Early exposure to English was moderate and exposure to Spanish was high; however, exposure after stroke was high in both languages. Impairment was equally severe in both languages. UTBA11 was trained in English and performance in this language improved significantly; however, no changes were observed in the untrained Spanish language. The model, however, predicts that training in English should improve both the trained language as well the untrained language (Spanish). Therefore, while the model accurately captures changes in the trained language (Spanish $r = .89$), the match in the untrained language is not adequate (English $r = .45$). UTBA22 learned Spanish as a native and English late in life. Pre-stroke exposure to English was low, but exposure to Spanish was high, and this pattern of exposure

continued post-stroke. Impairment was moderate in Spanish but severe in English. UTBA22 was trained in Spanish and showed a dramatic improvement as a function of treatment, however no remarkable changes were observed in the untrained language (English). The model's performance matches the patient's output very closely for Spanish ($r = .90$); however, it does not perfectly match performance in English ($r = .49$). BUBA01 learned Spanish as a native and English late in life. Pre-stroke exposure to English was low, but exposure to Spanish was high, and this pattern of exposure continued post-stroke. Impairment was moderate in both languages. BUBA01 was trained in English, which was this patient's weaker language. This patient shows a remarkable increase in the trained language which surpasses the untrained but stronger Spanish language performance (see Figure 3B). Interestingly, the model's performance closely mirrors this trend; with performance in English improving over time and performance in Spanish remaining stable (English $r = .91$; Spanish $r = -.47$). BUBA12 learned Spanish as a native and English late in life. Early exposure to English was low and exposure to Spanish was high; however, exposure after stroke was moderate (i.e., equal) in both languages. Impairment was severe in both languages. BUBA12 was trained in English, the weaker language, and showed only improvement in the trained language with little to no change in the untrained but stronger language. The model's performance closely matches the patient's performance in the trained language (English $r = .92$), however, it slightly over-predicts performance in the stronger untrained Spanish language (Spanish $r = .23$)

UTBA23 learned Spanish as a native and English later in life. Pre-stroke exposure to English was low, but exposure to Spanish was high, and this pattern of exposure continued post-stroke. Impairment was severe in both languages. UTBA23 was trained in Spanish and performance in the trained language for this patient improved whereas performance in the untrained English language only improved towards the end of treatment. The model's performance matches the patient's performance in the trained Spanish language and untrained English; however, since the patient's performance is more variable than the model's, the correlation coefficients are not as strong as expected (Spanish $r = .69$; English $r = .42$).

In a third subgroup of patients (UTBA02, UTBA18), the model's performance accurately matches patients' performance on the untrained language but the matches for the trained language are not perfect. For instance, UTBA02 was a native speaker of Spanish but learned English late. Before stroke, exposure to both languages was moderate (i.e., equal) and both languages were severely impaired after stroke. UTBA02 was trained in Spanish and while performance in Spanish improves immediately once treatment commences, performance in the untrained English improves, but the data is variable. The model's performance is more consistent, and based on the AoA, exposure and lesion parameters, predicts greater generalization to the untrained English language than what is actually observed (Spanish $r = .28$; English $r = .50$). For UTBA18, the model does not do as good a job matching the patient output for the trained language but adequately matches the untrained language. UTBA18 learned Spanish as a native and English late in life. Early exposure to both English and Spanish was moderate (i.e., equal), however, exposure after stroke was high in Spanish and low in English. Impairment was high in both languages although English was less impaired than Spanish. UTBA18 was trained in English. Although performance on the trained language improved, performance was highly variable. The untrained language (Spanish) also showed some marginal improvements towards the end of treatment (see Figure 3C). The model's performance is similar to the performance of the patient, but because improvements in the model are more stable and consistent in the two languages than the actual patient data, the correlation between model and patient performance is not strong (Spanish $r = .43$; English $r = .55$).

In the last subgroup of two patients (UTBA20, UTBA21) cross-correlation coefficients could not be calculated due to the lack of variance in the data as there are no improvements in the patient data. UTBA20 learned Spanish as a native and English late in life. Pre-stroke exposure to English was low, but exposure to Spanish was high, and this pattern of exposure continued post-stroke. Impairment was severe in both languages. UTBA20 was trained in Spanish, which would have been considered the stronger language, but output was very limited in both languages. This patient did not improve on either the trained language (Spanish) or the untrained language (English). However, the model predicts some improvements in outcome which are not observed in the patient's performance (see Figure 3D). Like UTBA20, UTBA21 learned Spanish as a native speaker but English early in life. Unlike UTBA20, pre-stroke exposure to English was high, but exposure to Spanish was low for this patient, and this pattern of exposure continued post-stroke. Impairment was severe in both languages. UTBA21 was trained in English, which was the stronger language, but training in English did not improve performance in English or Spanish. Like the previous patient, the model predicts slight improvements in outcome which are not observed in the patient's performance.

3.1 Preliminary results on predicting rehabilitation outcomes

One of the important aspects of the model simulations is that both language treatments can be simulated. Therefore, as described above in section "Modeling rehabilitation", the model was trained in both languages irrespective

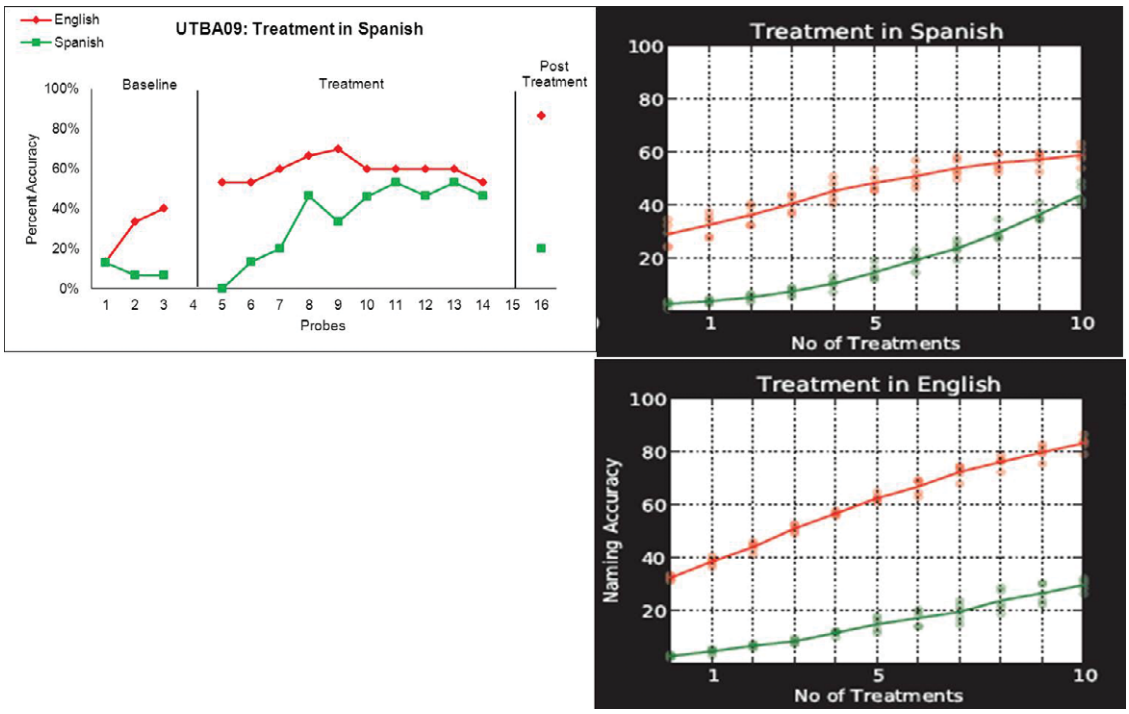

Figure 4. (Colour online) Model's prediction of treatment in both languages. In this scenario, the patient was trained in Spanish and the model accurately predicts performance in both languages. When, however, the model is trained in English, there are greater improvements predicted in the trained language. See text for details.

of which language the patient was trained in. For every patient, the model's output for both languages consequently predicts what the treatment outcome may have been if treatment was provided in the other language. As an example, UTBA09 was trained in Spanish, which was this patient's weaker language (see Figure 4). The model's predictions match the treatment outcome in that the trained Spanish language improved but improvements in English were marginal. If treatment were instead provided in English, which was the stronger language, the model predicts greater improvements in English and lesser improvements in the untrained language, Spanish. While it is purely speculative at this point, the results from the model provide a provocative alternative for the treated language for this patient: It is possible that the overall improvements in treatment would have been greater following training in English than Spanish for this patient and consequently, UTBA09 was trained in the wrong language.

4. Discussion

The present study aimed to address the important issue of language recovery following treatment in bilingual aphasia by comparing performance in treated bilingual aphasic individuals with the performance of a computational model simulating aphasia in a bilingual lexicon. From a clinical standpoint, the establishment of the efficacy of rehabilitation in each of the languages of the individual with bilingual aphasia is important because there are currently no clear recommendations on the best approach for rehabilitation of bilingual aphasia. Preliminary results obtained from the simulations of 17 patients allows a direct comparison of outcomes using two parallel yet complementary scientific approaches and show that the model successfully simulates improvements in the trained language in 13 out of the 17 patients. For a majority of the patients, the model accurately matches improvements in the trained language (see Figure 2). Additionally, when cross-language transfer is observed in patients, the model accurately simulates this cross-language transfer (e.g., UTBA02, BUBA07). Importantly, there are several cases where patients do not show cross-language transfer and the model also does not predict transfer (e.g., UTBA01, UTBA22). There are other cases where the model does not accurately match the patient data, likely because the patient data is much more variable than the model's outcomes. This lack of overlap may be more reflective of the variability observed in the patient treatment outcomes than the model's inability to accurately match performance. Individuals with aphasia often show intra-subject variability, i.e., performance varies from one day to another as well

as inter-subject variability. Neurological, psychological, social/motivational factors are just some of the factors that may influence the heterogeneity. This is precisely why single subject experimental designs are well suited to examine rehabilitation outcomes as individual variability is revealed as a function of time. The model, however, is modest in the number of parameters that determine treatment outcome. Future studies on computational rehabilitation in aphasia need to take into account the additional factors that may influence patient performance in treatment.

Nonetheless, there are several observations regarding the model and patient outcomes that are noteworthy. First, it appears that in individuals with differential proficiency prior to stroke, when the stronger language is trained, only that language improves, no changes are observed in the weaker untrained language. This pattern is observed in several patients (UTBA01, UTBA16, and UTBA17) and the model accurately simulates these outcomes. In one case, when the stronger language was trained, there was generalization to the weaker language (UTBA23), but this was a patient who learned both languages early in life suggesting that perhaps age of acquisition is an important factor in determining treatment outcomes. The model's architecture provides some insights into why this may be the case. For early AoA, training of both maps starts approximately at the same time (and using a similar large neighborhood size) whereas for late second language AoA, only L1 is trained first and training of L2 begins later, when the neighborhood is already smaller and well-defined. Since the global map structure is determined early in training, simulating late AoA can be expected to affect the global organization of the L2 phonetic map but not that of the semantic and L1 maps. Notably, the observation that the timing of L2 acquisition impacts the structural representation of L1 and L2 maps, such that L2 becomes parasitic on L1 has been reported previously in the developmental DEVLEX model (Zhao & Li, 2010).

Second, in individuals with differential proficiency prior to stroke, when the weaker language is trained, there are instances of cross-language generalization to the stronger language (UTBA09, UTBA07, BUBA04, BUBA07). Specifically, UTBA07 and UTBA09 learned English early but BUBA07 learned English late in life. UTBA09 and BUBA07 were less exposed to English than Spanish whereas UTBA07 and BUBA04 were less exposed to Spanish than English. When all these patients were trained on their weaker language, they showed improvements in the trained weaker language as well as cross-language transfer to the stronger language. Except for BUBA04, the model does a reasonable job simulating outcomes (correlation coefficients $> .5$) for the remaining three patients. These findings are consistent with our original work suggesting that training the weaker language may be more effective in facilitating generalization to the stronger language (Edmonds & Kiran, 2006) but are strengthened by replication across a more diverse group of participants.[3] In cases where there is no generalization from the trained language to the untrained language, with the exception of two patients, it is usually in patients with severe language impairments (low starting points), indicating that the severity of language impairment may be another factor determining the extent of cross-language generalization.

It should be noted that when the model does not accurately predict outcome, there are several reasonable explanations. For instance, when the total number of treatment sessions for UTBA07 (seven sessions) was precisely matched by the model, the model predicts improvements in both languages but not that performance in Spanish surpasses that in English. However, when the model was extended in time (to total of 10 sessions) the model's prediction matched the patient's performance very well. This finding indicates that the learning rates and treatment duration, which were kept consistent with the patient data in this study, can also be systematically manipulated for each patient in order to make more accurate predictions for treatment.

In another scenario, the model slightly over-predicted outcomes for two patients who showed no improvements as a function of treatment (UTBA20, UTBA21). It is important to note, however, that these patients were very severely limited in their verbal output and likely had concurrent apraxia of speech/phonological planning deficits. Phonological planning deficits are out of the scope of this computational model as they require further subcomponents of the L1/L2 model to be specified, and have been the focus of other computational models of monolingual speech production (Callan, Kent, Guenther & Vorperian, 2000; Guenther, Hampson & Johnson, 1998). Therefore, it is completely reasonable to expect the systematically increasing performance in the model, which does not include a speech planning component or impairment. Future versions of this model should focus on specifying stages involved in lexical selection and phonological planning in the two languages of bilingual individuals (Costa, La Heij & Navarrete, 2006). Third, while the associative connections between L1 and L2 currently modulate during training and rehabilitation, in the future their effect on modulating performance needs to be studied as well. The computational model can be used to gain insight and formulate hypotheses, which can then be verified in human subject experiments.

Finally, ability of the model to predict the optimal treatment language is a potentially revolutionary and powerful outcome of this study. For every patient, the model generated simulation outcomes based on treatment

[3] UTBA07 was one of the patients reported in the Edmonds and Kiran (2006) study.

in each language whereas our existing patient data only addresses treatment provided in one language (since each patient is unique and can be rehabilitated only once). As illustrated by the example patient (UTBA09), where the patient was trained in the weaker language and the model predicted the outcome correctly; the model also predicted that training in the stronger language would have resulted in even better outcome. Clearly, in order to confirm the predictive power of the model we need to systematically train patients based on the model's predictions and then compare the patient and model outcomes. In the interim, the advantage of the model is that it can be used to estimate what the best treatment for each individual patient might be based on individual patient language histories and language impairment, and thereby improve the state of the art in rehabilitation of bilingual aphasia and is a unique and novel approach in this field.

There are several limitations of this study. First, despite the fairly large sample size ($N = 17$ patients with bilingual aphasia); these patients were self-selected and therefore the model and the patient treatment data do not reflect every possible combination of AoA, language exposure and language impairment/lesion. A more complete study, where a large number of patients are treated with alternative treatments and predictions matched with the model, is warranted. Second, since this was a preliminary examination of computational lesioning and retraining; we kept the learning, lesion and retraining parameters consistent across patients, possibly explaining the lack of complete overlap between the patient and model data in some cases. Again, a systematic examination of free parameters that include varying the nature of language exposure, lesion strength and location and learning rates may reveal significant insights and may ultimately change the field of bilingual aphasia rehabilitation specifically and aphasia rehabilitation in general. We are currently pursuing this line of work and have completed an initial examination of the effect of lesion strength on naming impairment in bilingual aphasia (Grasemann, Kiran, Sandberg & Miikkulainen, 2012).

To summarize, the computational model described in this paper is an innovative approach to predict optimal rehabilitation protocols by simulating a bilingual language system in which language representations can vary by age of acquisition and relative proficiency, and are subsequently lesioned and retrained to improve output in order to facilitate the greatest amount of language recovery in bilingual aphasia.

References

Abutalebi, J., Rosa, P. A., Tettamanti, M., Green, D. W., & Cappa, S. F. (2009). Bilingual aphasia and language control: A follow-up fMRI and intrinsic connectivity study. *Brain and Language, 109 (2–3),* 141–156.

Baron, R., Hanley, J. R., Dell, G. S., & Kay, J. (2008). Testing single- and dual-route computational models of auditory repetition with new data from six aphasic patients. *Aphasiology, 22 (1),* 1–15.

Bates, E., D'Amico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pleh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin Review, 10 (2),* 344–380.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychological Review, 16 (4),* 161–169.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd edn.). Englewood Cliffs, NJ: Prentice Hall.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research, 43 (3),* 721–736.

Caplan, D. (2004). The neuro in cognitive neuropsychology [comment/reply]. *Cognitive Neuropsychology, 21 (1),* 17–20.

Caramazza, A., Hillis, A., Leek, E., & Miozzo, M. (1994). *The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits*. In L. A. Hirschfeld & S. A. Gelman (eds), *Mapping the mind: Domain specificity in cognition and culture*, pp. 68–84. Cambridge: Cambridge University Press.

Costa, A., La Heij, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition, 9 (2),* 137–151.

Croft, S., Marshall, J., Pring, T., & Hardwick, M. (2011). Therapy for naming difficulties in bilingual aphasia: Which language benefits? *International Journal of Language & Communication Disorders, 46 (1),* 48–62.

Dell, G. S., Schwartz, M. F., Martin, N. M., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104,* 801–838.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (2000). The role of computational models in neuropsychological investigations of language: Reply to Ruml and Caramazza (2000). *Psychological Review, 107 (3),* 635–645.

Edmonds, L. [A.], & Kiran, S. (2004). Confrontation naming and semantic relatedness judgements in Spanish/English bilinguals. *Aphasiology, 18 (5–7),* 567–579.

Edmonds, L. A., & Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in bilingual aphasia. *Journal of Speech, Language, and Hearing Research, 49 (4),* 729–748.

Farah, M., & Wallace, M. (1992). Semantically bounded anomia: Implications for the neural implementation of naming. *Neuropsychologia, 30 (21),* 609.

Faroqi-Shah, Y., Frymark, T., Mullen, R., & Wang, B. (2010). Effect of treatment for bilingual individuals with

aphasia: A systematic review of the evidence. *Journal of Neurolinguistics, 23,* 319–341.

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43 (2),* 182–216.

Frances, N., & Kucera, H. (1982). *Frequency analysis of English usage*. Boston, MA: Houghton Mifflin.

Grasemann, U., Kiran, S., Sandberg, C., & Miikkulainen, R. (2012). Computational simulation of lexical-semantic naming deficits in bilingual aphasia. Ms., The University of Texas at Austin.

Grasemann, U., Sandberg, C., Kiran, S., & Miikkulainen, R. (2011). Impairment and rehabilitation in bilingual aphasia: A SOM-based model. Presented at the 8th Workshop on Self-Organizing Maps (WSOM 2011), Espoo, Finland.

Green, D. W., Grogan, A., Crinion, J., Ali, N., Sutton, C., & Price, C. J. (2010). Language control and parallel recovery of language in individuals with aphasia. *Aphasiology, 24 (2),* 188–209.

Green, D. W., Ruffle, L., Grogan, A., Ali, N., Ramsden, S., Schofield, T., Leff, A. P., Crinion, J., & Price, C. J. (2011). Parallel recovery in a trilingual speaker: The use of the Bilingual Aphasia Test as a diagnostic complement to the Comprehensive Aphasia Test. *Clinical Linguistics & Phonetics, 25 (6–7),* 449–512.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105 (4),* 611–633.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition, 100 (3),* 464–482.

Junque, C., Vendrell, P., Vendrell-Brucet, J. M., & Tobena, A. (1989). Differential recovery in naming in bilingual aphasics. *Brain and Language, 36 (1),* 16–22.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd edn.). Philadelphia, PA: Lippincott Williams & Wilkins.

Keidel, J. L., Welbourne, S. R., & Lambon Ralph, M. A. (2010). Solving the paradox of the equipotential and modular brain: A neurocomputational model of stroke vs. slow-growing glioma. *Neuropsychologia, 48 (6),* 1716–1724.

Kiran, S. (2007). Complexity in the treatment of naming deficits. *American Journal of Speech-Language Pathology, 16 (1),* 18–29.

Kiran, S. (2008). Typicality of inanimate category exemplars in aphasia treatment: Further evidence for semantic complexity. *Journal of Speech, Language, and Hearing Research, 51 (6),* 1550–1568.

Kiran, S., & Bassetto, G. (2008). Evaluating the effectiveness of semantic-based treatment for naming deficits in aphasia: What works? *Seminars in Speech and Language, 29 (1),* 71–82.

Kiran, S., & Iakupova, R. (2011). Understanding the relationship between language proficiency, language impairment and rehabilitation. Evidence from a single case study. *Clinical Linguistics & Phonetics, 25 (6–7),* 565–583.

Kiran, S., & Johnson, L. (2008). Semantic complexity in treatment of naming deficits in aphasia: Evidence from well-defined categories. *American Journal of Speech Language Pathology, 17 (4),* 389–400.

Kiran, S., & Roberts, P. M. (2010). Semantic feature analysis treatment in Spanish–English and French–English bilingual aphasia. *Aphasiology, 24 (2),* 231–261.

Kiran, S., & Roberts, P. M. (2012). What do we know about assessing language impairment in bilingual aphasia? In M. R. Gitterman, M. Goral & L. K. Obler (eds.), *Aspects of multilingual aphasia*, pp. 35–51. Clevedon: Multilingual Matters.

Kiran, S., Sandberg, C., & Sebastian, R. (2011). Treatment of category generation and retrieval in aphasia: Effect of typicality of category items. *Journal of Speech, Language, and Hearing Research, 54,* 1101–1117.

Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research, 46 (4),* 773–787.

Kohnert, K. (2004). Cognitive and cognate-based treatments for bilingual aphasia: A case study. *Brain and Language, 91 (3),* 294–302.

Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain and Language, 65 (3),* 422–440.

Kohonen, T. (2001). *Self-organizing maps*. Berlin: Springer.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language, 33 (2),* 149–174.

Ladefoged, P. (1982). *A course in phonetics* (2nd edn.). Fort Worth, TX: Harcourt College Publishers.

Laganaro, M., Di Pietro, M., & Schnider, A. (2006). What does recovery from anomia tell us about the underlying impairment: The case of similar anomic patterns and different recovery. *Neuropsychologia, 44 (4),* 534–545.

Li, P. (2009). Lexical organization and competition in first and second languages: Computational and neural mechanisms. *Cognitive Science, 33 (4),* 629–664.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17 (8–9),* 1345–1362.

Li, P., & Green, D. W. (2007). Neurocognitive approaches to bilingualism: Asian languages. *Bilingualism: Language and Cognition, 10 (2),* 117–119.

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, and Computers, 34,* 408–415.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science: A Multidisciplinary Journal, 31 (4),* 581–612.

Lorenzen, B., & Murray, L. L. (2008). Bilingual aphasia: A theoretilcal and clinical review. *American Journal of Speech-Language Pathology, 17 (3),* 299–317.

Meinzer, M., Obleser, J., Flaisch, T., Eulitz, C., & Rockstroh, B. (2007). Recovery from aphasia as a function of language therapy in an early bilingual patient demonstrated by fMRI. *Neuropsychologia, 45 (6),* 1247–1256.

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge, MA: MIT Press.

Miikkulainen, R. (1997). Dyslexic and category-specific impairments in a self-organizing feature map model of the lexicon. *Brain and Language, 59,* 334–366.

Miikkulainen, R., & Kiran, S. (2009). Modeling the bilingual lexicon of an individual subject. *Lecture Notes in Computer Science, 5629,* 191–199.

Moretti, R., Bava, A., Torre, P., Antonello, R. M., Zorzon, M., Zivadinov, R., & Cazzato, G. (2001). Bilingual aphasia and subcortical-cortical lesions. *Perceptual and Motor Skills, 92 (3.1),* 803–814.

Plaut, D. C. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language, 52 (1),* 25–82.

Reggia, J. A., Gittens, S. D., & Chhabra, J. (2000). Post-lesion lateralisation shifts in a computational model of single-word reading. *Laterality: Asymmetries of Body, Brain and Cognition, 5 (2),* 133–154.

Roberts, P. M., & Kiran, S. (2007). Assessment and treatment of bilingual aphasia and bilingual anomia. In A. A. E. Ramos (ed.), *Speech and language disorders in bilinguals*, pp. 109–131. New York: Nova Science.

Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language, 73 (1),* 62–91.

Spitzer, M., Kischka, U., Guckel, F., Bellemann, M. E., Kammer, T., Seyyedi, S., Weisbrod, M., Schwartz, A., & Brix, G. (1998). Functional magnetic resonance imaging of category-specific cortical activation, evidence for semantic maps. *Cognitive Brain Research, 6 (4),* 309–319.

Tschirren, M., Laganaro, M., Michel, P., Martory, M. D., Di Pietro, M., Abutalebi, J., & Annoni, J. M. (2011). Language and syntactic impairment following stroke in late bilingual aphasics. *Brain and Language, 119 (3),* 238–242.

Welbourne, S. R., & Lambon Ralph, M. A. (2005). Using computational, parallel distributed processing networks to model rehabilitation in patients with acquired dyslexia: An initial investigation. *Aphasiology, 19 (9),* 789–806.

Welbourne, S. R., & Lambon Ralph, M. A. (2007). Using parallel distributed processing models to simulate phonological dyslexia: The key role of plasticity-related recovery. *Journal of Cognitive Neuroscience, 19 (7),* 1125–1139.

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13 (5),* 505–524.