NEUROPSYCHOLOGICAL REHABILITATION
https://doi.org/10.1080/09602011.2020.1773278

Effect of Mandarin Verb Network Strengthening Treatment (VNeST) in Mandarin-English bilinguals with aphasia: A single-case experimental design

Ran Li[a], Wen Li[b] and Swathi Kiran [a]

[a]Sargent College of Health & Rehabilitation Sciences, Boston University, Boston, MA, USA; [b]Wheelock College of Education and Human Development, Boston University, Boston, MA, USA

ABSTRACT

This study aimed to improve verb retrieval ability in Mandarin-English bilinguals with aphasia by adapting the Verb Network Strengthening Treatment (VNeST) into Mandarin Chinese. Two Mandarin-English bilingual patients with chronic post-stroke aphasia participated in this study via online conferencing system following a multiple-baseline design. Both of them completed a 10-week of Mandarin VNeST treatment, and were probed on verb retrieval ability in a sentence context in both languages. Response accuracy was analysed to investigate the treatment acquisition, within-language generalization, and cross-language generalization effects. Standardized language assessments in both languages were administered pre- and post-treatment to further examine generalization to other linguistic tasks. Error analysis was conducted to investigate the evolution of within- and cross-language errors. Both patients improved after training in Mandarin VNeST, and showed different patterns of within-language and cross-language generalizations. They also improved in a variety of standardized language tasks. Error analysis showed a decline in semantic errors over the course of treatment in both patients, with cross-linguistic errors showing a decrease during Mandarin probes and an increase during English probes in one of the patients. This study contributes to our current understanding of theories of bilingual verb processing, and provides treatment guidance in Mandarin-English bilinguals with aphasia.

ARTICLE HISTORY
Received 12 December 2019
Accepted 18 May 2020

KEYWORDS
Bilingual aphasia; Verb retrieval; VNeST; Mandarin-English; Cross-language generalization.

Introduction

Bilingual language processing

The common notion according to current psycholinguistic models of bilingual lexical access is that bilingual individuals have a shared semantic system with

CONTACT Ran Li rli92@bu.edu Department of Speech, Language & Hearing Sciences, College of Health & Rehabilitation Sciences: Sargent College, 635 Commonwealth Avenue, Boston University, Boston, MA 02215

This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/09602011.2020.1782586)

separate lexical representations of each language (de Groot, 1992; Francis, 2005; Kroll & Stewart, 1994). The same notion of a shared semantic system initially described for nouns in these psycholinguistic models should also apply to bilingual language processing of verbs, even though the structure of verbs differs from nouns. In many languages, verbs impose greater syntactic processing demands than nouns as verbs require a subject, can assign thematic roles, and are morphologically richer than nouns (Vigliocco et al., 2011).

Models of lexical representations of verbs in monolinguals (Levelt et al., 1999; Pickering & Branigan, 1998; Roelofs, 1992) have illustrated that syntactic information for verbs is represented at the lemma level, with each piece of syntactic information (i.e., syntactic category, tense, person, number, mood) is represented by a separate node activated whenever a verb is used in a particular context. Studies have implemented a syntactic priming paradigm to examine syntactic representations of verbs (Salamoura & Williams, 2007). Findings have suggested a shared lexical representation for verbs in healthy bilinguals, although the extent of overlap may depend on L2 language proficiency (Bernolet et al., 2013). These findings corroborate a shared conceptual system for verb processing in healthy bilinguals. However, it remains unclear whether such an account still holds in bilingual individuals with language impairment (i.e., aphasia).

Bilingual aphasia and treatment

More than a third of the population in the United States speaks a language other than English at home. Chinese is the third most-spoken language in the United States (Ryan, 2013), and is mostly spoken within Chinese-American populations (Lai, 2004). With the rise in the number of Chinese-speaking people in the U.S. also coinciding with an overall increase in the number of older people at risk for stroke and dementia (Hoeffel et al., 2012), bilingual Chinese-English elderly individuals are a growing minority with unique health care needs.

Bi/multilingual aphasia causes the loss of functions in one or more languages and the goal of speech language therapy is to help patients improve communicative skills in those languages. An important aspect of treatment effectiveness for bilingual aphasia is a generalization effect from a treated to an untreated language, which is also known as *cross-language generalization* (Kohnert, 2009). One finding related to cross-language generalization is that low-proficiency bilinguals are more likely to experience this effect than high-proficiency bilinguals after language therapy in the second language (L2) because L2 of low-proficiency bilinguals depends more heavily on the first language (L1) in unbalanced bilinguals (Faroqi-Shah et al., 2010). However, there is a lack of robust evidence for cross-language generalization, which also depends on many other linguistic and stroke-related factors (Kohnert, 2009).

Most prior studies in bilingual aphasia and bilingual aphasia rehabilitation have focused on Indo-European languages (Edmonds & Kiran, 2006), so it is challenging for clinicians to provide efficient speech-language therapy to patients who speak Asian languages, such as Mandarin Chinese. Prior studies in Chinese aphasia have mainly focused on examining language deficits and aphasia treatment in monolingual Chinese patients (Chen & Bates, 1998; Kong & Law, 2009; Law & Or, 2001; Wang & Thompson, 2016). These studies have demonstrated an overall similar pattern of language impairment as in aphasia studies in other language populations. However, there is a lack of clinical guidance for aphasia rehabilitation in bilingual patients who speak Chinese. Hence, the current work is important both from a clinical and scientific perspective (Kiran & Gray, 2018).

Clinical evidence for bilingual aphasia rehabilitation has mainly been derived from treatment of nouns. It is also important to examine the effect of treatment targeting verb retrieval as naming deficits of both grammatical categories are commonly observed in individuals with aphasia (Faroqi-Shah & Waked, 2010), and verbs in many languages carry rich morphological and syntactic information in sentence production critical to conveying one's message (Vigliocco et al., 2011). More work in bilingual aphasia treatment of verbs will contribute to clinical guidance for bilingual aphasia rehabilitation for improved patient outcomes.

Verb Network Strengthening Treatment

Verb Network Strengthening Treatment (VNeST) aims to improve verb retrieval in sentence context for individuals with aphasia (Edmonds et al., 2009). Theoretically speaking, improvements in lexical retrieval of verbs and their thematic roles are based on the notion that they are coactivated (Edmonds & Mizrahi, 2011). Additionally, generalization to lexical retrieval of an untrained verb and its thematic roles is expected because activating semantic representation of a trained verb engages the semantic representations of other verbs that share similar features with that verb (Rösler et al., 2001).

Results from previous VNeST studies (Edmonds et al., 2014; Edmonds & Babb, 2011) have indicated treatment acquisition effects in trained verbs and generalization effects to untrained verbs in monolingual English-speaking individuals with aphasia. However, it is unknown whether such generalization effects from treated verbs to untreated verbs are a consequence of semantic relatedness, as the flow of activation from the semantic system is target-language nonspecific during lexical retrieval (Costa et al., 1999). Some of these VNeST studies also have investigated the effect of treatment on connected speech, a process that requires an integration of cognitive, micro-linguistic, and macro-linguistic skills (Edmonds et al., 2009; Edmonds et al., 2014). Results have suggested a mixed pattern of improvement both within and across participants. A few other studies have adapted VNeST into other languages, which found successful treatment acquisition and within-language generalization in monolingual Korean patients with aphasia (Kwag et al., 2014), and cross-language generalization in some Hebrew-English bilinguals with aphasia (Lerman, 2020; Lerman et al., 2017). Hence, the current study aimed to build on the VNeST literature to investigate the treatment effect of Mandarin VNeST, and its within-language and between-language generalization effects in Mandarin-English bilinguals with aphasia.

Error analysis

Analysis of speech errors can provide additional insights into the underlying mechanism of treatment. Studies that have examined evolution of errors (Edmonds & Kiran, 2006; Kiran & Roberts, 2010) have found a decrease in semantic errors concurrent with increases in response accuracy. However, most of these studies focused on Indo-European languages, and it is not known if evolution of errors in non-Indo-European languages follows the same trajectory as errors that have been examined in the past. Thus, we conducted an error analysis in the current study to examine the effect of Mandarin VNeST on speech production in both trained and untrained languages.

The current study

The overarching goal of the current study was to adapt VNeST into Mandarin Chinese and to investigate its feasibility of improving verb-retrieval ability in Mandarin-English bilinguals with aphasia. Specifically, we addressed the following research questions:

Does training in Mandarin VNeST:

(1) Improve lexical retrieval of *trained* verbs in Mandarin Chinese and their thematic roles in sentence context, from baseline to post-treatment and maintenance?
(2) Generalize to (a) *semantically related untrained* verbs and their thematic roles in Mandarin Chinese, (b) the *untrained language* (English), and (c) lexical retrieval in *standardized language tasks* in both languages?
(3) Show an evolution of speech errors (i.e., decrease in semantic errors)?

Figure 1 shows a schema of the hypothesized relationship between the verb-thematic network of two semantically related verbs in both Mandarin and English. According to previous VNeST studies (Edmonds et al., 2009; Edmonds et al., 2014; Edmonds & Babb, 2011), we hypothesized that Mandarin-English bilinguals with aphasia would improve lexical retrieval of the trained items and semantically related untrained items within the same language. Based on our previous work in bilingual aphasia treatment (Edmonds & Kiran, 2006; Kiran & Roberts, 2010), we expected to observe cross-language generalization effects. However, those studies have focused on noun treatments whereas the current study aimed to apply a verb-focused treatment in a novel study

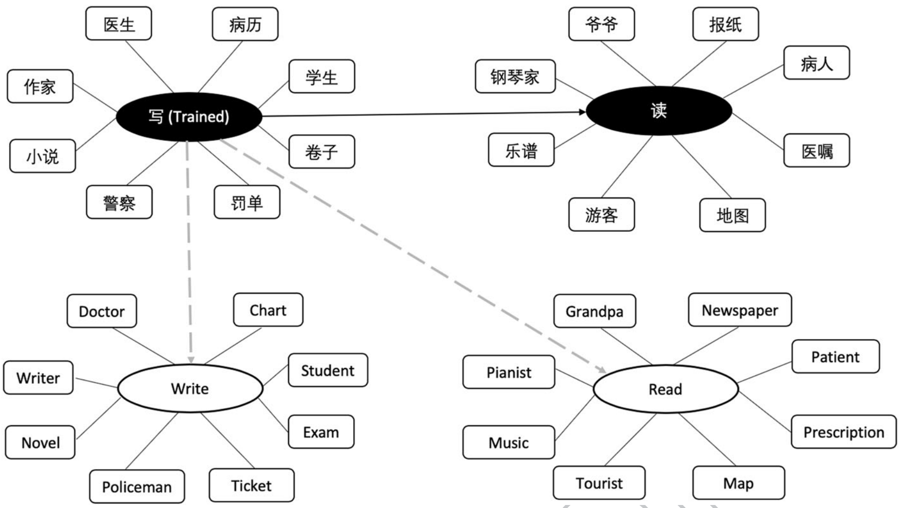

Figure 1. Schema of the hypothesized relationship between the verb-thematic network of *write* and *read* in Mandarin-English bilinguals with aphasia. (1) Training a target verb (e.g., *write*) in Mandarin Chinese should facilitate generation of all possible schemes of concepts (agent and patient), and the connections between the neural substrate for each agent-patient pair and the target verb (surrounding word pairs). (2) Training a target verb will generalize to semantically related verbs (e.g., *read*) as well as the connections to the neural substrates for their thematic pairs within the same language (solid arrow). (3) Training a target verb will further generalize to that verb's direct translation as well as its semantically related verbs in the untrained language (dashed arrows).

population (i.e., Mandarin-English). We also hypothesized that patients would improve in other standardized language tasks due to strengthening of the semantic network (Edmonds et al., 2009; Kiran & Roberts, 2010). Moreover, we expected a decreased pattern of semantic errors in both languages.

Methods

Experimental design

A multiple baseline approach across participants (Connell & Thompson, 1986) was implemented in this study. The design included four phases: (1) baseline ($n = 3$ sessions), (2) treatment ($n = 10$ sessions), (3) post-treatment ($n = 3$ sessions), and (4) 1-month maintenance ($n = 1$ session). A control task was administered at the same frequency as the treatment probe tasks (see details below). Additionally, language outcomes were measured via standardized assessments before and after treatment (see details below).

Participants

The inclusion criteria for this study were (1) those who spoke Mandarin and English fluently before the stroke, (2) diagnosis of aphasia based on the

Western Aphasia Battery-Revised (WAB-R, Kertesz, 2007) for English and the Aphasia Battery in Chinese (ABC; Gao, 1993, 1996) for Mandarin Chinese, (2) age between 18 and 85 years old, (3) with normal/near-normal or corrected-to-normal hearing and vision, (4) right-handed prior to stroke, and (5) no other neurological (i.e., dementia) or learning disorder. Two participants (P1 = 73yrs [F], P2 = 71yrs [M]) with chronic aphasia who met these criteria were recruited from San Francisco, California (months post stroke: P1 = 86, P2 = 140). Neither of the patients had participated in previous VNeST treatment. Table 1 provides patients' demographic information. Both participants gave consent according to BU IRB protocol.

Prior to the study, P1 was diagnosed with Broca's aphasia according to the WAB (AQ = 52.6; CQ = 58.5; LQ = 61) and the ABC (AQ = 38.2; CQ = 37.6; LQ = 46.8), and exhibited with moderate apraxia of speech. P2 was diagnosed with Anomic aphasia based on the WAB (AQ = 89.9; CQ = 81.1; LQ = 78.7) and the ABC (AQ = 80.8; CQ = 69.4; LQ = 77.3).

Individual differences in lexical processing can be influenced by developmental and contextual factors such as L2 age of acquisition (AoA), the degree of lifetime exposure to each language, and the frequency of use of each language (Kastenbaum et al., 2018). Therefore, information about second language acquisition and their language use of each language was collected via the Language Use Questionnaire (LUQ; Kastenbaum et al., 2018), and is shown in Table 2. Both participants were born in China and spoke Mandarin as their native language. They both acquired English as their second language in a school setting (mean age of acquisition = 13 years old) before they moved to the U.S. to receive higher education. P2 also acquired Cantonese before he moved to the U.S. Both P1 and P2's families were from China, and the participants exhibited more use in English in all daily activities and contexts, since they received higher education in English and used English at work, at home, and in social situations. Both participants self-reported 100% fluent in both Mandarin and English before the stroke. Therefore, they were considered highly balanced bilinguals based on provision of their language proficiency information prior to their stroke.

Table 1. Demographic and lesion information.

Pt	Sex	Age (years)	Education (years)	Race	MPO	Aphasia Type	Stroke or lesion information
1	F	73	18	Asian	86	Broca	Left MCA CVA secondary to atrial fibrillation. Involved regions including left frontoparietal region and anterior temporal lobe.
2	M	71	20	Asian	140	Anomic	Left MCA CVA secondary to right-sided carotid endarterectomy. Involved regions in left basal ganglia. Three months prior, another ischemic stroke resulted in a mild left hemiparesis with no other symptoms. Stenosis in the right internal carotid artery and left MCA were found at that time.

Pt = Patient; F = Female; M = Male; MPO = months post onset; MCA = middle cerebral artery; CVA = cerebrovascular accident.

Table 2. Language profiles for P1 and P2.

			Usage Before stroke (%)		Usage After stroke (%)		Fluency before stroke (%)		Fluency after stroke (%)	
Pt	Sex	AoA (years)	CH	EN	CH	EN	CH	EN	CH	EN
1	F	16	4	96	5	95	100	100	29	29
2	M	10	33	67	49	51	100	100	66	66

Pt = Patient; AoA = age of acquisition; F = Female; M = Male; CH = Mandarin Chinese; EN = English.

Materials

During each probe session, a sentence probe task and a control task in each language were administered. A complete list of treatment stimuli and target responses for probe stimuli are shown in Appendix I. There were six probe conditions in the current study: (1) Mandarin trained, (2) Mandarin untrained, (3) Mandarin control, (4) English trained, (5) English untrained, and (6) English control. The three English conditions were direct translations of each corresponding Mandarin condition.

Sentence probe pictures

Figure 2 shows examples of the picture stimuli. Eighteen pairs (total $n = 36$) of semantically related Mandarin Chinese verbs were selected for this study (e.g., “量” for *measure*, “拉” for *pull*), which were single-character and ditransitive verbs. Thirty-six pictures corresponding to each verb stimulus were hand-drawn (black and white) for sentence elicitation. These pictures were drawn approximately 4″ × 6″ and centred on 8″ × 11″ white paper. All sentences elicited an agent, verb, and patient in Mandarin Chinese (e.g., “木匠量楼梯” for *carpenter measures stairs*). Most of the agents had specific titles in order to promote specific language use instead of general words (e.g., “女人” for *woman*, “男人” for *man*). For those pictures that did not have a specific title, answers other than general words were accepted if they were appropriate to the picture description. The thirty-six verb stimuli were divided into verb set 1 and verb set 2, in which each verb in one set was semantically related to a

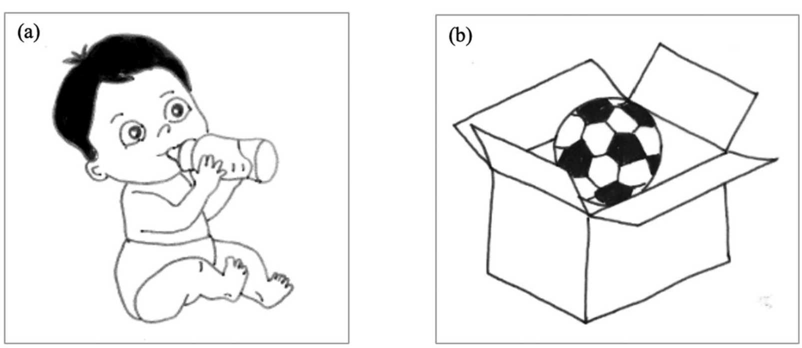

Figure 2. Examples of picture stimuli for sentence (a) and control (b) probe tasks.

verb in the other set (e.g., "量/称" for *measure/weigh*). The English verb sets were directly translated from Mandarin by the first author. Agents and patients for both Mandarin sets and English sets were matched for frequency (Brysbaert & New, 2009; Cai & Brysbaert, 2010). Based on two-tailed *t*-tests, there was no significant difference between agent and patient in set 1 for Mandarin Chinese (CH, $t(22) = .76$, $p = .45$) and their English translations (EN, $t(27) = .78$, $p = .44$), or set 2 (CH: $t(20) = -1.56$, $p = .13$; EN: $t(22) = -1.42$, $p = .17$). There was also no significant difference of frequency between agents (CH: $t(17) = 1.61$, $p = .13$; EN: $t(18) = 1.70$, $p = .11$) and patients (CH: $t(33) = .54$, $p = .60$; EN: $t(34) = -.34$, $p = .74$) for each verb set.

In both of the verb sets, only semantically "heavy" verbs (e.g., *measure, drink*) with specific semantic meanings were used, rather than semantically "light" verbs (e.g., *do, make*). Verb set 1 and verb set 2 were matched for verb frequency (CH: $t(23) = .98$, $p = .34$; EN: $t(29) = .53$, $p = .60$), imageability (CH: $t(31) = .38$, $p = .71$; EN: $t(15) = .33$, $p = .75$), familiarity (CH: $t(27) = -.10$, $p = .92$; EN: $t(16) = 1.31$, $p = .21$), and number of syllables for English ($t(31) = -.59$, $p = .56$). Frequencies for the Mandarin stimuli were retrieved from the *Chinese Single-character Word Database* (Liu et al., 2007), and word frequencies for their English translations were accessed from the MRC Psycholinguistic Database (Coltheart, 1981).

Control task

There were 10 hand-drawn pictures to elicit a target preposition (e.g., *on*), and two object names (e.g., *cake and table*), which were developed to generate a three-word sentence phrase as in the sentence probe task. It was hypothesized that patients would not improve on this task, because preposition words and features of objects were not explicitly trained.

Standardized language assessment

A variety of standardized language assessments were administered at pre- and post-treatment time points. These tests were administered to determine language abilities for both Mandarin Chinese and English at baseline, and to serve as outcome measures after treatment to answer research question 2 (c). Table 3 shows individual scores for each measure from pre- to post-treatment. Both participants completed the WAB-R (Kertesz, 2007) and the ABC (Gao, 1993, 1996). The ABC is a Chinese-adapted version of the WAB. Reliability and validity assessment outcomes from 199 post-stroke patients with aphasia and 165 post-stroke patients without aphasia indicated this test is applicable to the Chinese population (Gao, 1993; 1996). Single-word lexical retrieval in English was evaluated with the Boston Naming Test Long Form (BNT; Goodglass et al., 2001), which contains pictures of 60 common objects. A 30-item version for the Chinese BNT (Chen et al., 2014; Cheung et al., 2004) was administered to assess single noun retrieval ability in Mandarin Chinese. A cut-off score of 24 in spontaneous naming generated a sensitivity of 73.1% and specificity of

Table 3. Probe response accuracies at each phase, and standardized language measures at pre- and post-treatment.

	P1			P2		
Tests	Pre	Post	Main	Pre	Post	Main
Sentence probes (%)						
Trained – CH	1.9	37.0***	27.8***	53.7	94.4**	94.4**
Untrained – CH	0	11.1***	11.1***	72.2	75.9	66.7
CLT (trained)	1.9	7.4**	0	55.6	75.9**	61.1
CLT (untrained)	0	22.2***	0	59.3	79.6*	72.2
Control probes (%)						
CH	0	0	0	46.7	60**	40
EN	0	0	0	43.3	66.7***	60**
Aphasia Severity						
WAB-AQ	52.6	54	N/A	89.9	92.4	N/A
Information	5	7	N/A	8	9	N/A
Fluency	4	4	N/A	10	9	N/A
Comprehension	7.8	6.5	N/A	9.8	9.4	N/A
Repetition	4.9	4.3	N/A	8.9	9.4	N/A
Naming	4.6	5.4	N/A	8.3	9.4	N/A
ABC-AQ	38.2	46.8	N/A	80.8	83.7	N/A
Information	3	3	N/A	4	5	N/A
Fluency	13	16	N/A	24	24	N/A
Comprehension	5.5	6.7	N/A	9.0	9.2	N/A
Repetition	2.8	5.1	N/A	8.3	9.6	N/A
Naming	2.0	1.6	N/A	5	4.9	N/A
Lexical Retrieval (%)						
BNT – EN	13.3	23.3*	N/A	78.3	78.3	N/A
BNT – CH	6.7	6.7	N/A	70	73.3	N/A
NAVS – VNT Total	13.6	50*	N/A	77.3	77.3	N/A
1-place	40	80	N/A	60	100	N/A
2-place	10	40	N/A	90	70	N/A
3-place	0	42.9	N/A	71.4	71.4	N/A
NAVS – ASPT Total	25	18.8	N/A	84.4	90.6	N/A
1-place	40	0	N/A	60	100	N/A
2-place	13.3	33.3	N/A	86.7	86.7	N/A
3-place	33.3	8.3	N/A	91.7	91.7	N/A
Cognitive functions						
CLQT Composite Severity	2.6	3.2	N/A	3.2	3.8	N/A
Attention	2	3	N/A	2	4	N/A
Memory	2	3	N/A	3	3	N/A
Executive functions	4	4	N/A	4	4	N/A
Language	1	2	N/A	4	4	N/A
Visuospatial skills	4	4	N/A	3	4	N/A
Connected Speech						
WAB Number of words	22	31	N/A	32	43	N/A
WAB MLU	2.27	2.71	N/A	7.4	9.4	N/A
WAB Number of utterances	11	14	N/A	5	5	N/A
WAB Complete Sentence (%)	9.1	35.7	N/A	100	100	N/A
WAB Number of CIUs	9	14	N/A	30	37	N/A
WAB % CIUs	40.9	45.2	N/A	93.8	86	N/A
WAB CIUs per min	4.5	3.5	N/A	60	37	N/A
ABC Number of words	15	13	N/A	54	48	N/A
ABC MLU	2.67	1.38	N/A	9	8	N/A
ABC Number of utterances	3	8	N/A	6	6	N/A
ABC Complete Sentence (%)	0	0	N/A	83	67	N/A
ABC Number of CIUs	9	6	N/A	48	46	N/A
ABC % CIUs	60	46.2	N/A	88.9	95.8	N/A
ABC CIUs per min	4.5	6	N/A	48	46	N/A

Main = maintenance; CLT = cross-language generalization; EN = English; CH = Mandarin Chinese; MLU = mean length of utterance; WAB-AQ = Wester Aphasia Battery Aphasia Quotient; ABC-AQ = Aphasia Battery in Chinese Aphasia Quotient; BNT = Boston Naming Test; NAVS = Northwestern Assessment of Verbs and Sentences; VNT = Verb Naming Test; ASPT = Argument Structure Production Test; CIU = content information unit; min = minute; CLQT = Cognitive Linguistic Quick Test; Probes: * = small effect size, ** = medium effect size, *** = large effect size; Lexical retrieval: * = $p < .05$.

75.3% in differentiating normal from participants with brain injury (Cheung et al., 2004), suggesting the culturally adapted 30-item version of BNT is applicable to the Chinese-speaking population. Additionally, verb retrieval in English was evaluated at the single-lexical level using the Verb Naming Test (VNT) from the Northwestern Assessment of Verbs and Sentences (NAVS, Thompson, 2012), and verb retrieval in sentences was evaluated with the Argument Structure Production Test (ASPT) from the NAVS (Thompson, 2012). Following the same procedures in Edmonds et al. (2009), we did not show or read the verb to the participant in the ASPT. We in addition collected connected speech samples using the picnic scene from the WAB-R (Kertesz, 2007) and the ABC (Gao, 1993, 1996). Both patients also completed the Cognitive Linguistic Quick Test (CLQT; Helm-Estabrooks, 2001) for assessing their attention, memory, executive function, and visuospatial skills.

Procedures

Baseline

Three baseline sessions were administered for each participant in both Mandarin Chinese and English separately. During each baseline session, all probe stimuli (36 sentence probes and 10 control probes) were presented in a random order for sentence elicitation at the beginning of each session and were audio- and video-recorded. No semantically related verbs (e.g., *eat, drink*) were presented consecutively. For each sentence probe picture, participants were asked to "make a sentence that includes him/her, the action, and this" in either Mandarin or English, while the experimenter pointed to the agent, verb, and patient (Edmonds et al., 2009, 2014). Verbal instructions were conveyed in the same language as the target language for the probe. For each control probe stimulus, participants were asked to "make a sentence that includes this object, the location, and this object" in either language, while the experimenter pointed to object 1, preposition, and object 2 correspondingly. No prompts or feedback were given to the patients at this phase, unless the patient used a general agent/patient (e.g., *woman, man*), or a general verb (e.g., *cut* instead of *chop*).

Treatment and post-treatment probes

Sentence and control probes ($n = 46$) were administered in every treatment session, alternating between Mandarin and English (e.g., Mandarin probes in probe session 1, English probes in probe session 2, Mandarin probes in probe session 3, etc.). Therefore, there were a total of 20 treatment probes administered, from which there were 10 for each language. Treatment was terminated after each participant completed all 20 sessions of treatment. Three post-treatment probes for each language were administered subsequent to the treatment phase following the same procedures above. In order to assess if any treatment gains were maintained, participants completed one more probe session in both languages one-month after the treatment.

Treatment materials and administration

Eighteen treatment stimuli (verb set 1 or set 2 in Mandarin Chinese) were selected for treatment, and they were counterbalanced across participants (P1 received verb set 1, P2 received verb set 2). Because Mandarin was their L1 and the less-dominant language, both participants received treatment only in Mandarin and cross-language generalization was examined in English. For each treatment item in the verb set, stimuli consisted of: (a) one PowerPoint slide each for all the verb stimuli, (b) 6–8 textboxes for each verb containing 3–4 agents and 3–4 patients that together form 3–4 thematic pairs, (c) 5 textboxes containing *wh*-question words: *who, what, where, when, why*, and (d) 12–16 sentences containing the target verb from the following four categories were orally presented to each participant for semantic judgment: (i) correct sentences; (ii) inappropriate agent; (iii) inappropriate patient; (iv) thematic reversal.

Mandarin VNeST was conducted by the first author, who is a native speaker of Mandarin Chinese with a Master's degree in Speech-Language Pathology. The treatment procedures were adapted and modified based on Edmonds (2014)'s VNeST treatment tutorial. Given that both participants were living in California during the treatment, the study was conducted over an online videoconference system, GoToMeeting (www.gotomeeting.com, Perron & Ruffolo, 2011). The Mandarin VNeST was provided for each patient twice per week for 10 weeks, with 2 h per session (total = 40 h). Because each session included approximately 20 min – 50 min of probes, we provided daily homework to the participants to supplement the treatment (Appendix II). Since P1 and P2 exhibited with different severities of aphasia, P1 received treatment of 3–4 items per session, and P2 received treatment of 6–7 items per session. The number of repetitions for each trained item was balanced within each participant.

Treatment steps

The treatment steps followed closely to the original VNeST (Edmonds, 2014), with slight adjustment for the purpose of online treatment delivery (Appendix III).

Treatment reliability

In order to ensure the treatment protocol was conducted consistently for each participant, the second author who is a native speaker of Mandarin Chinese, conducted fidelity check for 25% of the video-taped sessions. The treatment protocol was followed with a reliability of 97% on average (P1: 94%, P2: 100%).

Scoring

Probe tasks

Audio files for all probe sessions were transcribed for future scoring. Scoring criteria closely followed prior VNeST studies (Edmonds et al., 2009). One point was credited for each correct elicitation of agent, verb, and patient for each picture, with one phonemic error (e.g., *hottle* for *bottle*) allowed per lexical item. Because the goal of treatment was to improve lexical retrieval of target words in a sentence context, grammatical and morphological errors were not penalized. Acceptable alternative words were given credit (*e.g., father* for *uncle*). Incorrect sentence elicitation was scored zero. Non-acceptable responses included words that did not match the picture stimuli, any general words (e.g., *woman, man*) after prompt was provided, or more than one phonemic error per word. The second author also conducted interrater fidelity check for 25% of probe scoring. A point-to-point evaluation showed 96.5% agreement in scoring on average (P1: 99%; P2: 94%).

Pre- and post-treatment language measures

Response accuracy in each standardized language assessment was calculated based on guidelines within each test manual. Participants were given credit in the ASPT of the NAVS if all required lexical items were present in a sentence, regardless of word order or verb inflections (Edmonds et al., 2009). In order to examine generalization effects to these tasks, we compared pre- and post-assessment scores for each language task. It was considered clinically significant if scores improved more than 5 points on the WAB-AQ and 3.30 points on the BNT (Gilmore et al., 2019).

Connected speech

Lexical retrieval in connected speech was measured using the picnic scene from the WAB-R and the ABC for both Mandarin Chinese and English. During each task, participants were asked to describe the picture using sentences. Responses were then transcribed and coded for number of words and CIUs (Nicholas & Brookshire, 1993) using the CLAN programme (MacWhinney, 2014). Percentage of CIUs (% CIUs) and CIUs per minute were calculated to assess potential changes in efficiency, and total number of utterances and mean length of utterance (MLU) were calculated. The percentage of complete sentences was additionally measured. A complete utterance contained an agent and a verb (or patient if necessary) that were relevant to the scene, regardless of grammatical, morphological, and phonemic errors. After scoring, the number of complete utterances was divided by the total number of utterances to obtain the percentage of complete sentences.

Error analysis

We included probe responses from the baseline sessions, the fifth treatment session as a midway point, post-treatment sessions, as well as the maintenance phase for each language. In addition to the inclusion criteria mentioned above for VNeST treatment responses, we included self-corrected responses and dialectal differences as correct responses. All the excluded responses were classified into the following error types: (1) phonological (e.g., *hottle* for bottle), (2) semantic (e.g., *brush* for scrub), (3) morphosyntactic (e.g., *father watch TV* for father watches TV), (4) neologism (defined as more than 50% of the word is unrecognizable, e.g., *hibo*), (5) lexical (lexical omission of target words; e.g., *father TV* for father watches TV), (6) no response (NR or "*I don't know*"), and (7) cross-language (e.g., *student* for 学生). Additionally, we qualitatively rated each utterance based on the severity of speech errors following a 10-point scale (Appendix IV). Specifically, lower numbers in the scale represent more severe speech errors (*e.g.,* lexical), whereas higher numbers represent less severe error types (*e.g.,* morphosyntactic) (Kendall et al., 2013). Speech errors were coded and rated by both first and second authors, and inter-rater discrepancies were resolved to 100% agreement.

Data analysis

Probe responses

We conducted logistic mixed-effects models (McCulloch & Neuhaus, 2015) for each language to predict treatment outcomes and generalization effects. We did not include responses from the maintenance phase in the analysis as we expected them to decline at one-month post-treatment, and thus, such change might affect the overall estimation of treatment effect. Statistical analyses were performed in R Studio version 3.5.1 (Rstudio Team, 2016) using package "*lmerTest*" (Kuznetsova et al., 2017). We started each model with the maximum random structure (Barr et al., 2013). The random term that explained the least data variance was removed until the model reached the convergence.

In the model that aimed to predict treatment outcomes and within-language generalization, the dependent variable was the percent accuracy of Mandarin responses (0 = inaccurate, 1 = accurate). Fixed factors included probe sessions (i.e., session 1, session 2, etc.), sentence condition (i.e., Mandarin trained, Mandarin untrained, Mandarin control), and a session-by-condition interaction. We also included a random intercept for items and a by-session random slope. The model predicting cross-language generalization included the same model structure except the dependent variable was the percent accuracy of English responses, and sentence condition included English trained, English untrained and English control, which were direct translations of the Mandarin conditions.

Treatment effect sizes were further obtained using the Cohen's *d* calculation (Cohen, 1988; $d = M_2 - M_1/\sigma$, where *M* is the mean, σ is the standard deviation at pretreatment) for each sentence condition (i.e., Mandarin trained, Mandarin untrained, Mandarin control, English trained, English untrained, English control) from pre- to post-treatment, as well as from pre-treatment to maintenance. If the pre-treatment standard deviation was zero (e.g., P1's data), the pooled standard deviation from the trained and untrained scores were calculated instead. We followed the benchmarks of effect sizes in Edmonds et al. (2014), from which *d* values of 2.3, 3.7, and 5.5 corresponded to small, medium, and large effects respectively for the trained items, and *d* values of 1.2, 1.7, and 3.3 were designated as small, medium, and large effects respectively for the untrained items.

Pre- and post-treatment standardized measures

In order to assess generalization to other standardized language tasks, we conducted McNemar's Chi-squared tests on the Mandarin and English BNTs (Chen et al., 2014; Goodglass et al., 2001), as well as on the VNT and the ASPT from the NAVS (Thompson, 2012).

Error analysis

We conducted two sets of analyses to investigate trends of speech errors: (1) types of speech errors, and (2) qualitative count of speech errors. For each analysis, we started the model with the maximum random structure (Barr et al., 2013) in R Studio version 3.5.1 (Rstudio Team, 2016), and the random term that explained the least variance was removed until we reached the model convergence.

We first conducted Poisson generalized mixed-effects models (McCulloch & Neuhaus, 2015) to predict responses other than NRs for each probe language separately. The count of speech errors was the dependent variable. Within each regression model, predictors included probe session (i.e., baselines, treatment, post-treatment, maintenance), types of errors (i.e., phonological, semantic, morphosyntactic, neologism, lexical, cross-language), sentence condition (i.e., trained, untrained, control), and a session-by-type and a session-by-condition interaction. We also included a random intercept for items.

Next, we conducted linear mixed-effects models to predict the change of rating scores for each language separately. Rating scores other than 10s (correct/grammatical target utterance) were included as the dependent variable. Fixed factors included probe session, sentence condition, and a session-by-condition interaction. We also included a random intercept for items.

Results

Treatment acquisition and generalization

Table 3 and Figure 3 display the response accuracies in Mandarin and English probes. These results addressed research questions 1, 2 (a), and 2 (b), which aimed to investigate the (1) treatment acquisition effect, (2) within-language generalization effect, and (3) cross-language generalization effect.

We conducted a generalized linear model (GLM; Nelder & Wedderburn, 1972) to predict P1's responses in Mandarin Chinese and English due to non-convergence of the logistic mixed-effects models. Results did not reveal significance in both languages. Nevertheless, the Mandarin trained condition showed a large effect size from baseline to both post-treatment and maintenance, suggesting that Mandarin VNeST was effective in improving verb retrieval for P1. There was also a large effect size for semantically related untrained items from baseline to both post-treatment and maintenance, indicating that training in Mandarin VNeST enabled generalization to semantically related items within the same language. Meanwhile, we found a medium effect size in the English trained condition and a large effect size in the English untrained condition from baseline to post-treatment. These findings indicated a cross-language generalization effect. At the maintenance phase, performance in both English conditions fell to the baseline (0%). There was no evidence suggesting improvement in the control task for P1.

Results from the logistic mixed-effects model conducted to predict treatment outcomes in P2's Mandarin responses showed a significant main effect of the trained condition ($\beta = 4.54$, SE = 1.46, $|z| = 3.12$, $p < .01$), the untrained condition ($\beta = 2.89$, SE = 1.40, $|z| = 2.07$, $p < .05$), as well as a significant session-by-condition interaction ($\beta = .53$, SE = .17, $|z| = 3.16$, $p < .01$) for the trained condition. These results indicated that P2's performance was higher in the Mandarin trained and Mandarin untrained conditions compared to the control condition, and

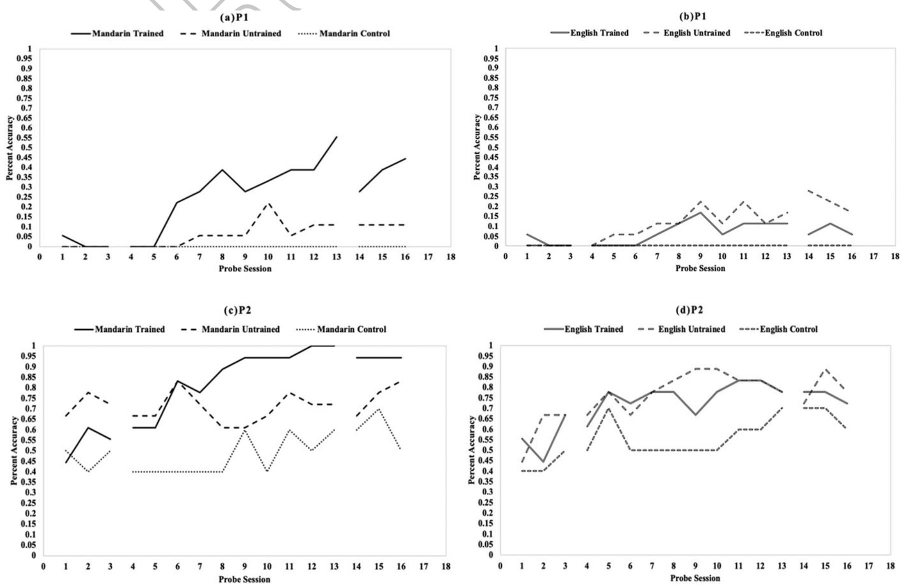

Figure 3. Percent accuracy of probe responses in Mandarin Chinese and English for P1 (a) (b) and P2 (c) (d).

there was a significant improvement in the trained condition over time. Additionally, we found a medium effect size for the Mandarin trained condition from baseline to post-treatment, which was well maintained one month later. There was also a medium effect size for the Mandarin control condition from baseline to post-treatment, which fell to the baseline at the maintenance phase. Results from the logistic mixed-effects model conducted to predict a cross-language generalization effect in P2's English probes did not reveal significance. However, we found a medium effect size for the English trained condition, as well as a small effect size for the English untrained condition from baseline to post-treatment. These results suggest a cross-language generalization effect in the untrained language. Additionally, we observed a large effect size for the English control condition from baseline to post-treatment, and a medium effect size for the same condition from baseline to maintenance, suggesting a practice effect.

Generalization to standardized language measures

Both patients' scores on standardized assessments from pre- to post-treatment are demonstrated in Table 3.

Results from the McNemar's Chi-squared test for P1's naming performance indicated a significant increase of correct noun retrieval in the English BNT (Goodglass et al., 2001) ($\chi^2 = 4.45$, $df = 1$, $p < .05$), and a significant gain of correct verb retrieval ($\chi^2 = 6.40$, $df = 1$, $p < .05$) in the VNT of the NAVS (Thompson, 2012).[1] Measures from P1's Mandarin connected speech showed a mixed pattern of change, such that the percentage of CIUs decreased, but the CIUs/min and the number of total utterances increased. Measures from the English connected speech task showed an increase of the percentage of complete sentences and the percentage of CIUs, which indicated a generalization effect to sentence production at the connected speech level after training in Mandarin VNeST.

P2 showed improvement in the overall aphasia severity based on the WAB-AQ and the ABC-AQ, but these scores did not meet the criteria for clinical significance (Gilmore et al., 2019). Correct lexical retrieval based on the Chinese BNT and the ASPT-NAVS also improved, but they did not reach statistical significance ($p > .05$). Measures from P2's Mandarin connected speech response showed an increase of the percentage of CIUs, but a decrease of the total number of words, the CIUs per min, and the percentage of complete sentences. In the English connected speech task, there was a gain of the total number of words, the MLU, and the number of CIUs, but a decline of the percentage of CIUs and the CIUs per min.

Both of the patients showed slight improvements in the Cognitive Linguistic Quick Test (CLQT; Helm-Estabrooks, 2001). P1 improved in attention, memory, and language, and P2 improved in attention and visuospatial skills. These gains in CLQT might reflect an increase of engagement or attention involved within the therapy context.

Error analysis

Figure 4 illustrates types of speech errors in P1 and P2's responses (sentence probes and control probes). Results in P1's Mandarin responses showed that cross-language ($p < .05$) and semantic errors ($p < .01$) declined over time compared to morphosyntactic errors. Additionally, neologistic ($p < .05$) and phonological errors ($p < .01$) increased significantly over time. Results in P1's English responses exhibited a significant increase of cross-language ($p < .01$) and phonological errors ($p < .01$), as well as a significant decrease of semantic errors ($p < .01$). Because P2's speech errors in Mandarin responses were nearly all semantic errors, we removed types of errors from the Poisson mixed-effects model. Results revealed a significant decline of semantic errors over time in the trained condition ($p < .01$). English responses did not show any significant changes of speech errors.

Figure 5 displays the change of qualitative count of speech errors in P1 and P2's responses, ranging from more severe (.5) to less severe (9.5) according to the rating scale (Appendix IV). P1's Mandarin responses showed a significant change from more severe to less severe errors in the untrained condition over time ($\beta = .26$, SE = .11, t = 2.49, $p < .05$). No significance was found in P1's

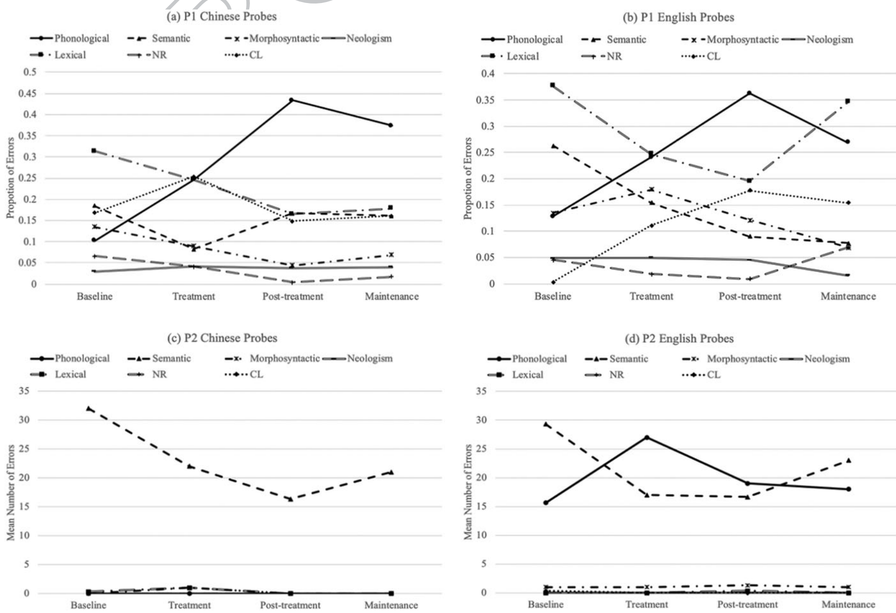

Figure 4. Types of speech errors in Mandarin Chinese and English responses for P1 (a) (b) and P2 (c) (d), respectively. CL = Cross-language; NR = No Response.

English responses ($ps > .05$). P2's responses showed a significant change from more severe to less severe speech errors in both Mandarin trained ($\beta = .25$, SE $= .06$, $t = 4.07$, $p < .01$) and English trained ($\beta = .14$, SE $= .06$, $t = 2.48$, $p < .05$) conditions.

Discussion

The current study aimed to adapt VNeST (Edmonds et al., 2009) into Mandarin Chinese, and examine its feasibility of improving verb retrieval ability in sentence context for Mandarin-English bilingual individuals with aphasia. This study answered whether training in Mandarin VNeST can: (1) improve lexical retrieval of a trained verb, its agent and patient (treatment acquisition effect), (2) generalize to (a) semantically related items (within-language generalization), (b) the untrained language (cross-language generalization), (c) other standardized language tasks, and (3) reduce production of semantic errors over time.

The current study successfully replicated the VNeST treatment protocol when extended to Mandarin-English bilinguals with aphasia. Indeed, this study, to our knowledge, was the first bilingual language treatment study that aimed to improve verb retrieval ability in Mandarin-English bilinguals with aphasia as most studies have focused on Indo-European languages. In general, results corroborated the notion of a shared semantic-syntactic system in bilingual verb processing, and provided clinical evidence for rehabilitating bilingual individuals with verb retrieval deficit. Importantly, this treatment was delivered entirely through videoconference and reinforces emerging trends that effective treatments such as VNeST can be delivered via video conferencing, thus circumventing physical and demographical constraints with in-clinic treatment. In the discussion that follows, we elaborate on these findings.

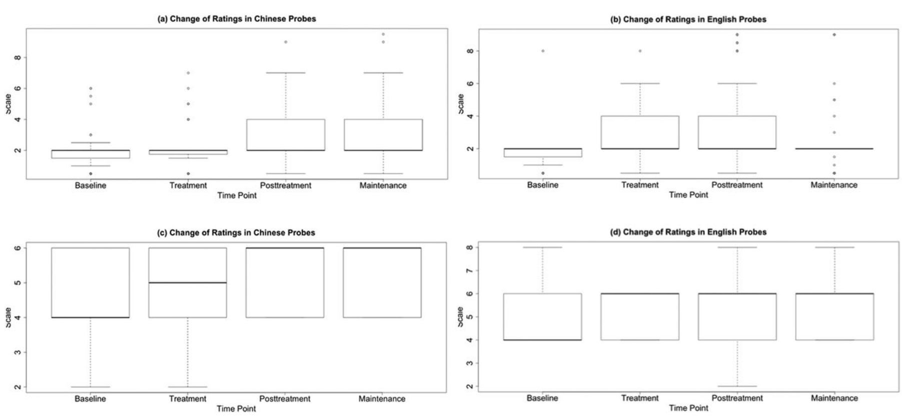

Figure 5. Qualitative ratings of Mandarin Chinese and English responses in P1 (a) (b) and P2 (c) (d), respectively.

Treatment effect

The current study successfully replicated previous VNeST studies (Edmonds et al., 2014, 2009), where training in VNeST improved lexical retrieval of a trained verb, its agent and patient in a sentence context, even when the treatment was provided in Mandarin Chinese. Thus, despite the different organization of Mandarin, improvements in sentence production of trained verbs and their agents and patients were observed due to strengthened connections between the trained verb and its thematic roles through repeated activation and use (Edmonds et al., 2009). Previous VNeST studies have shown positive treatment outcomes in different language populations, including 19 English speaking monolinguals with aphasia (Edmonds et al., 2014, 2009; Edmonds & Babb, 2011; Furnas & Edmonds, 2014), three Korean speaking individuals with aphasia (Kwag et al., 2014), and three Hebrew-English bilingual with aphasia (Lerman, 2020). The current study extended the impact of VNeST to Mandarin Chinese, which is linguistically distant from English. Our results together with previous VNeST studies suggest promising rehabilitation outcomes using verb-focused treatments for individuals with aphasia, pointing to future directions in aphasia treatment.

We additionally observed a gain in P2's responses in the control task. This task was designed to match the complexity of the sentence probe task. One explanation to the unexpected improvement in the control task is that it was related to the mild aphasia severity, which influenced P2's performance on the control task over the course of treatment. Another possible interpretation is an improved mapping process across sentences containing different syntactic structures, which applied to both of the control probe and the sentence probe. Previous studies have shown successful treatment generalization from improved sentence comprehension of locative sentences to comprehension of active declarative sentences (Byng, 1988), predicting a generalization between verb-based syntactic structures and locative structures.

Within-language generalization

Both P1 and P2 illustrated an improvement in verb retrieval in the semantically related untrained items in addition to the trained items, as evinced by the effect sizes. The effect of within-language generalization in VNeST was hypothesized to activate diverse event schemas that were related to the target verb (Edmonds et al., 2009). Therefore, improvement in the Mandarin untrained condition implies spreading activation from the trained verb to the untrained verb and its thematic roles. Findings of a within-language generalization in the current study support the mixed model of bilingual lexical retrieval developed by de Groot (1992), which predicts equally strong connections between lexicons of both languages in relatively balanced bilinguals. However, results from the regression models did not capture a within-language generalization effect. In particular, P1's responses did not reveal any significance because there was a lack of variance in the control condition, hence the effect size changes are meaningful. Likewise, a lack of significant interaction between probe session and the Mandarin untrained condition in P2's responses is possibly associated with the unexpected improvement of the control items, which again limits the interpretation of within-language generalization. It is worth noting that effect sizes and regression models are different ways of assessing the effects of therapy. Effect sizes examined the magnitude of change from pre- to post-treatment, whereas the regression models in the current paper included probe data from each time point over the course of the treatment, which generated more conservative results.

Cross-language generalization

Both P1 and P2 demonstrated modest changes in effect sizes in the English trained and untrained conditions, suggesting a cross-language generalization effect of Mandarin VNeST. Expectation of cross-language generalization in bilingual rehabilitation was based on the premise of a shared semantic system between languages (de Groot, 1992; Francis, 2005; Kroll & Stewart, 1994), which indicates that lexical access depends on proficiencies in both L1 and L2, with connections between each lexicon and between the semantic system and both lexicons strengthened over the course of treatment. Even though these models have mainly been tested on nouns, we assumed that the same notion of a shared conceptual system also applies to verbs. The magnitudes of change in the English conditions across both of the patients support the account of a shared representation system for bilingual verb processing (Salamoura & Williams, 2007). A previous healthy bilingual study implemented a syntactic priming paradigm to investigate the representation of verb argument structure in Greek (L1) advanced learners of English (L2) (Salamoura & Williams, 2007). Results showed a use of L1 structures while producing sentences in L2. Other studies have found that proficient bilinguals share more syntactic structures of their L1 and L2, while less proficient bilinguals start out with separate syntactic representations for new syntactic structures in their L2 (Bernolet et al., 2013).

Findings of the cross-language generalization are still mixed in the literature (Kohnert, 2009). Premorbid language proficiency and language dominance play roles in bilingual aphasia treatment outcomes (Gil & Goral, 2004; Kohnert, 2009). Both of our patients were highly proficient in Mandarin Chinese and English prestroke according to their self-rated proficiencies. Improvements in the English conditions for both P1 and P2 validate the mixed model (de Groot, 1992), which predicts equally strong connections from conceptual representation to lexicons of both languages in relatively balanced bilinguals. Findings of a cross-language generalization in the current study are also consistent with previous studies in healthy bilinguals and bilingual individuals with aphasia (Edmonds & Kiran, 2006). Both types of studies suggest that activating a target lexical item can generalize to semantically related items both within- and between-languages (Costa et al., 1999; Edmonds & Kiran, 2006; Hermans et al., 1998). These generalization effects also provide evidence in support of the prediction that the conceptual system can spread out activation to both lexicons regardless of the target language (Green, 1998a; Hermans et al., 1998).

We also found differential effects of cross-language generalization across patients, such that P1 showed a larger magnitude of improvement in the English untrained condition than in the English trained condition, suggesting cross-language interference. According to the cross-language interference account in bilingual lexical retrieval (Gollan & Silverberg, 2001; Van Hell & de Groot, 1998), translation equivalents in the non-target language are activated during lexical retrieval in the target language, which may suppress the correct retrieval in the target language. In the current study, the activation threshold for the English trained condition was inhibited from the intensive training of Mandarin VNeST. However, since the non-target language was activated and likely in competition with the trained language, the English untrained condition was activated with a larger magnitude over the English trained condition. When Mandarin VNeST was terminated for one month, there was a decreased response accuracy in the English untrained set from post-treatment (mean accuracy: P1 = 22.2%, P2 = 79.6%) to the maintenance phase (mean accuracy: P1 = 0%, P2 = 72%), suggesting a deteriorated cross-language interference. Nonetheless, this observation warrants future investigation.

Other language and cognitive measures

Widespread generalization beyond the trained verbs was hypothesized because Mandarin VNeST focused on strengthening semantic knowledge underlying noun and verb representations and the neural connections between them. Since verbs allow for diverse possibilities for both agent and patient, they provide additional opportunities for accessing potential agents and patients (Edmonds et al., 2009). After training in Mandarin VNeST, P1 showed an improvement in their overall aphasia severity, and lexical retrieval ability of nouns and verbs in English, according to the WAB-AQ, the English BNT, and the VNT of the NAVS. Consistent with previous findings in other bilingual aphasia treatment studies (Edmonds & Kiran, 2006; Kiran & Roberts, 2010), these improvements suggest that training in Mandarin VNeST facilitated activation of semantic features not only in the trained language, but also in the untrained language. While another potential mechanism of improvement in single verb retrieval may be because of an overlap (six) between verbs in the VNT and verbs used in the treatment probes, only one of these verbs improved from pre- to post-treatment on the VNT. These findings suggest that such improvement was not just due to the overlap in the test stimuli.

Both patients additionally showed different patterns of improvement in the connected speech task, according to measures of complete sentences and correct information unit. In particular, the percentage of complete sentences in P1's English response after treatment was four times larger than pre-treatment scores but there were no changes in Mandarin. The percentage of CIUs increased in English but decreased in Mandarin. In comparison, the percentage of CIUs in P2's connected speech decreased in English while increased in Mandarin, and the percentage of complete sentences decreased in Mandarin and remained stable in English. These mixed results suggest a dynamic but incomplete view of cross language generalization in connected speech. One interpretation of the improvement in this task is that patients had been exposed to producing and listening to sentences during therapy, which might enable them to produce more sentences during this task. Future studies with a larger sample size are needed to systematically investigate the effect of Mandarin VNeST on connected speech production.

Patients further exhibited an improved composite score on the CLQT. P1 remained moderately severe but P2 improved from mild to within the normal limits (WNL) after treatment. Such change of CLQT scores could be a result from an increased engagement or attention during the therapy sessions (Villard & Kiran, 2015). Nevertheless, future studies may examine how attention relates to treatment outcomes of Mandarin VNeST with a larger sample size.

Error analysis

Speech errors in both patients demonstrated an overall decline, which suggest an evolution from more severe errors (*e.g.,* lexical) to less severe errors (*e.g.,* morphosyntactic). Consistent with previous studies (Edmonds & Kiran, 2006; Kiran & Roberts, 2010), both of our patients showed a reduced production in semantic errors in both trained and untrained languages, which further evinced the significance of a treatment acquisition effect. The treatment outcome was generated from strengthened connections between the trained items and their thematic roles through repeated activation and use (Edmonds et al., 2009). Therefore, the amount of semantic errors decreased because Mandarin VNeST aimed to target semantic processing of verbs and reduce the activation of a semantically related competitor.

P1's cross-language errors (English responses) declined during Mandarin probes while they (Mandarin responses) increased during English probes. P2 did not produce any cross-language errors. These results suggest an interesting dynamic where training in Mandarin Chinese led to fewer cross-language errors when the target probe language was the trained language than when the target probe language was the untrained English. Furthermore, the amount of cross-language errors (Mandarin responses) decreased one month following the treatment. One explanation for this result is that cross-language errors increased when the non-target language was not appropriately inhibited (Green, 1998a; Hermans et al., 1998) and in this case, the incompletely inhibited language was the trained language (i.e., Mandarin). This potential observation of a lack of cognitive control/inhibition and resulting cross-language errors subsequent to training requires further examination in future studies.

Limitations and future directions

The current study investigated the feasibility of Mandarin VNeST in two Mandarin-English bilingual patients with chronic aphasia. One limitation is a small sample size and future studies may replicate the findings with a larger sample size. The other limitation is that verb retrieval ability in Mandarin Chinese was not assessed due to a lack of instrument. A standardized battery of assessment, the *Northwestern Assessment of Verbs and Sentences-Chinese* (NAVS-C; Wang & Thompson, 2016) should be included in future studies. Additionally, the interesting finding of a cross-language interference in P1's English untrained condition draws our attention to the underlying mechanism of bilingual language recovery. Future studies can implement different methodologies such as computational modelling and neuroimaging techniques to investigate factors that may contribute to an interference effect. Finally, both patients showed distinct rehabilitation outcomes after treatment, so hypotheses for differential aphasia recovery patterns need to be further examined. Such hypotheses may include separate lexical systems for different languages, and partial or complete inactivation of one language during activation of the other language (Gil & Goral, 2004).

Conclusion

The current study aimed to adapt the *Verb Network Strengthening Treatment* (VNeST) into Mandarin Chinese, and to investigate its treatment effect, within-language and between-language generalization effects, its generalization to other language measures, and its influence on the evolution of speech errors. Significant treatment outcomes in two patients with chronic aphasia suggest that training in Mandarin VNeST improved verb retrieval ability in Mandarin-English bilinguals with aphasia. Results indicate different patterns of within- and between-language generalization effects for both of our patients. Additional improvements in other standardized language tasks in both Mandarin and English substantiate the richness of generalization effects. The error analysis demonstrated a reduced trend of semantic errors in both of our patients, but an increase in cross language errors in one of the two patients. The current study contributes to theories of bilingual verb processing, as well as bilingual aphasia treatment in a population that clearly needs greater treatment guidance.

Note

1. Six of the verb stimuli in the VNT (i.e., “wash”, “read”, “watch”, “pinch”, “throw”, “bite”) overlapped with the treatment stimuli (treated and semantically related untreated), but only one verb on the VNT (“read”) improved from pre- to post-treatment assessment.

Acknowledgments

First and foremost, the authors thank the individuals who participated in this study. We also thank the members of the Aphasia Research Laboratory for their support throughout the development and execution of this study. Special thanks to Maria Varkanitsa for her kind help with drawing the probe stimuli and suggestions on the control task.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The first and third authors were funded by the NIH/NIDCD under [grant number 1U01DC014922]. The authors report no conflict of interest. Swathi Kiran is a scientific consultant for The Learning Corporation (FKA Constant Therapy) but there is no scientific overlap with the work in this project; National Institute on Deafness and Other Communication Disorders.

ORCID

Swathi Kiran http://orcid.org/0000-0003-1586-8913

References

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). *Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, *127*(3), 287–306. https://doi.org/10.1016/j.cognition.2013.02.005

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Byng, S. (1988). Sentence processing deficits: Theory and therapy. *Cognitive Neuropsychology*, *5*(6), 629–676. https://doi.org/10.1080/02643298808253277

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chen, S., & Bates, E. (1998). The dissociation between nouns and verbs in Broca’s and Wernicke’s aphasia : findings from Chinese. *Aphasiology*, *12*(1), 5–36. https://doi.org/10.1080/02687039808249441

Chen, T.-B., Lin, C.-Y., Lin, K.-N., Yeh, Y.-C., Chen, W.-T., Wang, K.-S., & Wang, P.-N. (2014). Culture qualitatively but not quantitatively influences performance in the Boston naming test in a Chinese-speaking population. *Dementia and Geriatric Cognitive Disorders Extra*, *4*(1), 86–94. https://doi.org/10.1159/000360695

Cheung, R. W., Cheung, M. C., & Chan, A. S. (2004). Confrontation naming in Chinese patients with left, right or bilateral brain damage. *Journal of the International Neuropsychological Society*, *10*(1), 46–53. https://doi.org/10.1017/S1355617704101069

Cohen, J. (1988). Statistical power analysis for the behavioural science (2nd Edition). In *Statistical power anaylsis for the behavioural science (2nd Edition)*.

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Connell, P. J., & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders*, *51*(3), 214–225. https://doi.org/10.1044/jshd.5103.214

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical Selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001–1018. https://doi.org/10.1037/0278-7393.18.5.1001

Edmonds, L. A. (2014). Tutorial for Verb Network Strengthening treatment (VNeST): Detailed description of the treatment protocol with corresponding theoretical rationale. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, *24*(3), 78–88. https://doi.org/10.1044/nnsld24.3.78

Edmonds, L. A., & Babb, M. (2011). Effect of verb network strengthening treatment in moderate-to-severe aphasia. *American Journal of Speech-Language Pathology*, *20*(2), 131–145. https://doi.org/10.1044/1058-0360(2011/10-0036)

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. https://doi.org/10.1044/1092-4388

Edmonds, L. A., Mammino, K., & Ojeda, J. (2014). Effect of verb network strengthening treatment (VNeST) in persons with aphasia: Extension and replication of previous findings. *American Journal of Speech-Language Pathology*, *23*(2), S312–S329. https://doi.org/10.1044/2014_AJSLP-13-0098

Edmonds, L. A., & Mizrahi, S. (2011). Online priming of agent and patient thematic roles and related verbs in younger and older adults. *Aphasiology*, *25*(12), 1488–1506. https://doi.org/10.1080/02687038.2011.599527

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology*, *23*(3), 402–424. https://doi.org/10.1080/02687030802291339

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*(4), 319–341. https://doi.org/10.1016/j.jneuroling.2010.01.002

Faroqi-Shah, Y., & Waked, A. N. (2010). Grammatical category dissociation in multilingual aphasia. *Cognitive Neuropsychology*, *27*(2), 181–203. https://doi.org/10.1080/02643294.2010.509340

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 251–267). Oxford University Press.

Furnas, D. W., & Edmonds, L. (2014). The effect of computerised verb network strengthening treatment on lexical retrieval in aphasia. *Aphasiology*, *28*(4), 401–420. https://doi.org/10.1080/02687038.2013.869304

Gao, S. (1993). *Shiyuzheng [Aphasia]*. Peking Union Medical College and Peking Union Medical College.

Gao, S. (1996). Clinical diagnosis of aphasia in Chinese. *Stroke and Nervous Diseases*, *3*, 57–59.

Gil, M., & Goral, M. (2004). Nonparallel recovery in bilingual aphasia: Effects of language choice, language proficiency, and treatment. *International Journal of Bilingualism*, *8*(2), 191–219. https://doi.org/10.1177/13670069040080020501

Gilmore, N., Dwyer, M., & Kiran, S. (2019). Benchmarks of significant change after aphasia rehabilitation. *Archives of Physical Medicine and Rehabilitation*, *100*(6), 1131–1139. https://doi.org/10.1016/j.apmr.2018.08.177

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*(1), 63–83. https://doi.org/10.1017/S136672890100013X

Goodglass, H., Kaplan, E., & Weintraub, S. (2001). *Boston Naming Test Record Booklet*. Lippincott Williams & Wilkins.

Green, D. W. (1998a). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81. https://doi.org/10.1017/S1366728998000133

Helm-Estabrooks, N. (2001). *Cognitive linguistic quick test: CLQT*. Psychological Corporation.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*(3), 213–229. https://doi.org/10.1017/S1366728998000364

Hoeffel, E. M., Rastogi, S., Kim, M. O., & Hasan, S. (2012). *The asian population: 2010*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Kastenbaum, J. G., Bedore, L. M., Peña, E. D., Sheng, L., Mavis, I., Sebastian-Vaytadden, R., Rangamani, G., Vallila-Rohter, S., & Kiran, S. (2018). The influence of proficiency and language combination on bilingual lexical access. *Bilingualism: Language and Cognition*, *22*(2), 300–330. https://doi.org/10.1017/S1366728918000366

Kendall, D. L., Pompon, R. H., Brookshire, C. E., Minkina, I., & Bislick, L. (2013). An analysis of aphasic naming errors as an indicator of improved linguistic processing following phonomotor treatment. *American Journal of Speech-Language Pathology*, *22*(2), 240–249. https://doi.org/10.1044/1058-0360(2012/12-0078)

Kertesz, A. (2007). *Western aphasia battery: Revised*. Pearson.

Kiran, S., & Gray, T. (2018). Understanding the nature of bilingual aphasia. *Bilingual Cognition and Language: The State of the Science Across its Subfields*, *54*, 371–400. https://doi.org/10.1075/sibil.54.17kir

Kiran, S., & Roberts, P. M. (2010). Semantic feature analysis treatment in spanish-English and French-English bilingual aphasia. *Aphasiology*, *24*(2), 231–261. https://doi.org/10.1080/02687030902958365

Kohnert, K. (2009). Cross-language generalization following treatment in bilingual speakers with aphasia: A review. *Seminars in Speech and Language*, *30*(3), 174–186. https://doi.org/10.1055/s-0029-1225954

Kong, A. P. H., & Law, S. P. (2009). A linguistic communication measure for monitoring changes in Chinese aphasic narrative production. *Clinical Linguistics and Phonetics*, *23*(4), 255–269. https://doi.org/10.1080/02699200802673234

Kroll & Stewart. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Jounal of Memory and Languages*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 13. https://doi.org/10.18637/jss.v082.i13

Kwag, E. J., Sung, J. E., Kim, Y. H., & Cheond, H. J. (2014). Effects of verb network strengthening treatment on retrieval of verbs and nouns in persons with aphasia. *Communication Sciences & Disorders*, *19*(1), 89–98. https://doi.org/10.12963/csd.14119

Lai, H. M. (2004). *Becoming Chinese American: A history of communities and institutions (Vol. 13)*. Rowman Altamira.

Law, S. P., & Or, B. (2001). A case study of acquired dyslexia and dysgraphia in Cantonese: Evidence for nonsemantic pathways for reading and writing Chinese. *Cognitive Neuropsychology*, *18*(8), 729–748. https://doi.org/10.1080/02643290143000024

Lerman, A. (2020). Direct and Indirect Treatment Effects in Multilingual People with Aphasia.

Lerman, A., Goral, M., & Edmonds, L. (2017, November 5–7). *Within- and cross- language generalization after Verb Network Strengthening Treatment (VNeST) in a Hebrew-English bilingual with aphasia*. Conference Abstract: Academy of Aphasia 55th Annual Meeting, Baltimore, USA.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, *39*(2), 192–198. https://doi.org/10.3758/BF03193147

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, volume II: The database*. Psychology Press.

McCulloch, C. E., & Neuhaus, J. M. (2015). Generalized linear mixed models. *Encyclopedia of Biostatistics*, *4*. https://doi.org/10.1016/B978-0-08-097086-8.42017-9

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384. https://doi.org/10.2307/2344614

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, *36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Perron, B., & Ruffolo, M. (2011). A review of a web-conferencing technology: GoTo meeting.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Rösler, F., Streb, J., & Haan, H. (2001). Event-related brain potentials evoked by verbs and nouns in a primed lexical decision task. *Psychophysiology*, *38*(4), 694–703. https://doi.org/10.1111/1469-8986.3840694

Rstudio Team. (2016). RStudio: Integrated development for R. *RStudio. Inc., Boston MA. RStudio*, https://doi.org/10.1007/978-3-642-20966-6

Ryan, C. (2013). Language use in the United States: 2011. *American Community Survey Reports, US Census Bureau*, (August), 16. doi:ACS-22.

Salamoura, A., & Williams, J. N. (2007). Processing verb argument structure across languages: Evidence for shared representations in the bilingual lexicon. *Applied Psycholinguistics*, *28*(4), 627–660. https://doi.org/10.1017/S0142716407070348

Thompson, C. K. (2012). Northwestern Assessment of Verbs and Sentences (NAVS).

Van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. https://doi.org/10.1017/S1366728998000352

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Villard, S., & Kiran, S. (2015). Between-session intra-individual variability in sustained, selective, and integrational non-linguistic attention in aphasia. *Neuropsychologia*, *66*, 204–212. https://doi.org/10.1016/j.neuropsychologia.2014.11.026

Wang, H., & Thompson, C. K. (2016). Assessing syntactic deficits in Chinese Broca's aphasia using the *Northwestern Assessment of Verbs and Sentences-* Chinese (NAVS-C). *Aphasiology*, *30*(7), 815–840. https://doi.org/10.1080/02687038.2015.1111995

Appendices

Appendix I. Probe stimuli (Mandarin and direct translations in English) and target responses.

Verb Set 1:

Target verb	English translation	Agent_CH	Agent_EN	Patient_CH	Patient_EN
接	catch	运动员	athlete	棒球	baseball
量	measure	兽医	veterinarian	狗	dog
写	write	女儿	daughter	作文	essay
看	watch	爷爷	grandpa	电视	TV
蒸	steam	妈妈	mom	包子	buns
洗	wash	饲养员	zookeeper	猴子	monkey
喝	drink	婴儿	baby	牛奶	milk
缝	sew	设计师	designer	裤子	trousers
拉	pull	农民	farmer	猪	pig
修	fix	工人	worker	汽车	car
咬	bite	狗	dog	猫	cat
偷	steal	贼	thief	钱包	wallet
砍	chop	屠夫	butcher	肉	meat
搅/和	mix	糕点师	pastry cook	面	dough
抱	hug	圣诞老人	Santa Claus	邻居	neighbor
打	hit	老师	teacher	学生	student
捏	pinch	哥哥	brother	表弟	cousin
举	lift	魔术师	magician	兔子	bunny

CH = Mandarin Chinese; EN = English.

Verb Set 2:

Target verb	English translation	Agent_CH	Agent_EN	Patient_CH	Patient_EN
扔	throw	青少年	teenager	飞盘	frisbie
称	weigh	护士	nurse	婴儿	baby
读	read	学生	sister	小说	novel
听	listen	叔叔	uncle	收音机	radio
煮	boil	室友	roommate	水	water
擦	scrub	农场主	farmer	马	horse
吃	eat	顾客	customer	面条	noodles
织	knit	奶奶	grandma	围巾	scarf
推	push	饲养员	zookeeper	大象	elephant
装	install	程序员	programmer	软件	software
舔	lick	兔子	rabbit	猫	cat
抢	rob	强盗	robbers	银行	bank
切	slice	厨师	chef	面包	bread
摇	shake	调酒师	bartender	瓶子	cup
亲	kiss	新郎	groom	新娘	bride

(*Continued*)

Continued.

Target verb	English translation	Agent_CH	Agent_EN	Patient_CH	Patient_EN
踢	kick	学生	student	足球	soccer
戳	poke	阿姨	aunt	侄子	niece
拎	carry	保姆	nanny	小孩	child

CH = Mandarin Chinese; EN = English.

Control Set

Target Prep.	English translation	Obj1_CH	Obj1_EN	Obj2_CH	Obj2_EN
下面	below	鸟	bird	飞机	airplane
上面	on	蛋糕	cake	桌子	table
对面	across	人	person	房子	house
外面	outside of	狗	dog	狗屋	kernel
旁边	beside	男孩	boy	冲浪板	surfboard
周围	around	孩子们	children	大树	tree
里面	in	足球	soccer	盒子	box
前面	in front of	月亮	moon	云	cloud
中间	between	熊	bear	狗	dogs
后面	behind	女孩	girl	椅子	chair

Prep. = Preposition; Obj = object; CH = Mandarin Chinese; EN = English.

Appendix II. Homework examples

Participants were asked to put three words (agent, verb, patient) into the correct order to generate a phrase. There were 54 trials in this exercise, including three trials for each trained item. The examiner provided correct feedback to patients, who were asked to spend 30 min every day to practice reading these correct phrases. The goal of this exercise was to provide daily exercise in addition to the treatment.

"请把下列每组词的顺序重新排列。然后每天朗读所有正确排序的句子。" (Please put the following words into the correct order. Then practice reading the correct phrases.)

Example: 小明 汽车 驾驶 (Xiaoming car drive)

Correct response: 小明 驾驶 汽车 (Xiaoming drive car)

Exercise trials for the target verb: *catch*

1. 篮球 接 运动员 (basketball catch athlete)
2. 学生 飞盘 接 (student frisbee catch)
3. 接 孩子 气球 (catch child balloon)

Appendix III. Treatment steps for Mandarin VNeST

Step 1	Showed textboxes with words "谁" (*who*) and "什么" (*what*), asked participants to generate multiple scenarios around the trained verb.
Step 2	Participants read the generated agent-verb-patient triads aloud.
Step 3	Asked participants to choose one scenario, then showed textboxes with *wh*-question words "为什么" (*why*), "什么时候" (*when*), and "在哪里" (*where*) one at a time, and asked participants to answer these questions. The clinician typed answers on the screen, and asked participants to read them aloud.
Step 4	The clinician read four simple, active sentences containing the target verb, and participants were asked to decide whether each sentence was semantically accurate or not (Yes or No).
Step 5	Asked participants what verb/action they had been working on.
Step 6	Repeated Step 1 without providing any cues.

Appendix IV. Qualitative count of speech errors

10	complete/grammatical sentence
9.5	complete/grammatical sentence with neologism(s)
9	CLT complete/grammatical sentence
8.5	CLT complete/grammatical sentence + neologism (s)
8	sentence with morphosyntactic errors other than argument omission
7.5	sentence with morphosyntactic errors other than argument omission + neologism(s)
7	sentence with CLT morphosyntactic errors other than argument omission
6.5	sentence with CLT morphosyntactic errors other than argument omission + neologism(s)
6	sentence with semantic error(s) – semantically related substitution(s)
5.5	sentence with semantic error(s) – semantically related substitution(s) + neologism(s)
5	CLT with semantic error(s) – semantically related substitution(s)
4.5	CLT with semantic error(s) – semantically related substitution(s) + neologism(s)
4	sentence with semantic error(s) – semantically unrelated substitution(s)
3.5	sentence with semantic error(s) – semantically unrelated substitution(s) + neologism(s)
3	CLT with semantic error(s) – semantically unrelated substitution(s)
2.5	CLT with semantic error(s) – semantically unrelated substitution(s) + neologism(s)
2	sentence with lexical error(s)
1.5	sentence with lexical error(s) + neologism(s)
1	unintelligent speech
.5	no response

CLT = Cross-language Transfer/Generalization.