Title: The role of language proficiency and linguistic distance in cross-linguistic treatment effects in aphasia

Peggy S. Conner[a], Mira Goral[a,b,c], Inge Anema[d], Katy Borodkin[e], Yair Haendler[f], Monica Knoph[c,g], Carmen Mustelier[a], Elizabeth Paluska[a], Yana Melnikova[h], Mariola Moeyaert[i]

[a]Speech-Language-Hearing Sciences, Lehman College of the City University of New York, Bronx, NY, USA; [b]Speech-Language-Hearing Sciences, Graduate Center of the City University of New York, New York, NY, USA; [c]MultiLing, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo, Norway; [d]Communication Disorders Department, SUNY New Paltz, New Paltz, NY, USA; [e]Department of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; [f]Laboratoire de linguistique formelle, Université Paris Diderot, Paris, France; [g]Department of Speech and Language Disorders, Statped, Oslo, Norway; [h]Parcours d'Enfant, Montreal, Quebec, Canada; [i]Educational Psychology and Methodology, University at Albany, SUNY, Albany, NY, USA

Corresponding Author: Peggy S. Conner, Speech-Language-Hearing Sciences, Lehman College, 250 Bedford Park Blvd. West, Bronx, NY 10468; 718-960-8030 or 914-420-2669; peggy.conner@lehman.cuny.edu,

This is an Accepted Manuscript of an article published by Taylor & Francis in Clinical Linguistics & Phonetics on 08 Feb 2018: http://www.tandfonline.com/10.1080/02699206.2018.1435723

ABSTRACT

Current findings from intervention in bilingual aphasia are inconclusive regarding the extent to which levels of language proficiency and degree of linguistic distance between treated and non-treated languages influence cross-language generalisation and changes in levels of language activation and inhibition following treatment. In this study we enrolled a 65-year-old multilingual speaker with aphasia and administered treatment in his L1, Dutch. We assessed pre- and post-treatment performance for 7 of his languages, 5 of high proficiency and 2 of lower proficiency. We asked whether treatment in L1 would generalise to his other languages or increase interference among them. Forty hours of treatment were completed over the course of five weeks. Each language was tested three times at pre-treatment and at post-treatment. Testing included measures of narrative production, answering questions, picture description and question generation. Dependent measures examined language efficiency, defined as Correct Information Units/min. as well as language mixing, defined as proportion of code-mixed whole words. We found that our participant's improved efficiency in Dutch was mirrored by parallel improvement in the four languages of high proficiency, English, German, Italian, and French. In contrast, in his languages of lower proficiency, Norwegian and Spanish, improved efficiency was limited. An increase in code-mixing was noted in Spanish, but not in Norwegian. We interpret the increased code-mixing in Spanish as indication of heightened inhibition following improvement in a language of close linguistic proximity, Italian. We conclude that an interaction of language proficiency and linguistic similarity affects cross-language generalisation following intervention in multilingual aphasia.

KEY WORDS

multilingual; aphasia; cross-linguistic; language mixing; treatment generalisation; efficiency

INTRODUCTION

Bilingual speakers who acquire aphasia typically receive language intervention in one of their languages. Research studies have provided mixed results to answer the question of whether treatment-related improvement in language production and comprehension of individuals with aphasia generalises to their untreated language(s) (e.g. Faroqi-Shah, Frymark, Mullen, and Wang, 2010; Kohnert and Peterson, 2012). Several research studies have reported robust cross-language generalisation (e.g. Kiran and Iakupova, 2011; Knoph, 2013; Kurland and Falcon, 2011), whereas others have reported limited or non-existent change in the untreated language (Abutalebi, Della Rosa, Tettamanti, Green, and Kappa, 2009; Keane and Kiran, 2015; Meinzer, Obleser, Flaisch, Eulitz, and Rockstroh, 2007; Miller Amberber, 2012). It is also common for such treatment studies to report cross-language generalisation in some, but not all, outcome measures (e.g. Altman, Goral, and Levy, 2012; Laganaro and Overton-Venet, 2001), to some, but not all, untreated languages (e.g. Goral, Rosas, Conner, Maul, and Obler, 2012; Knoph, Lind, and Simonsen, 2015; Miertsch, Meisel, and Isel, 2009) and ---for studies that report on several bilingual participants---to find the generalisation in some, but not all, participants (e.g. Croft, Marshall, Pring, and Hardwick, 2011; Kiran and Roberts, 2010; Kiran, Sandberg, Gray, Ascenso, and Kester, 2013).

For example, Goral and colleagues enrolled a trilingual speaker of Hebrew, English, and French, who was seven years post a left CVA and demonstrated mild aphasia in all his languages. Treatment targeted morphosyntactic structures in sentence production and was conducted in English, his second language (L2). Following treatment, testing revealed improved accuracy on several measures of language production in the treated language. As well, the untreated French (L3) demonstrated improved grammatical accuracy, for example, increased correct use of prepositions and pronouns, but not on other measures, including article use and speech rate in sentence production (Goral, Levy, and Kastl, 2010). Moreover, morphosyntactic measures at the sentence level did not show change in his other untreated language, his L1 Hebrew, while other measures of connected speech production, such as sentence grammaticality and narrative structure, did (as reported in Altman, et al., 2012). Similarly, Knoph, Lind, and Simonsen (2015) treated a multilingual speaker in her late-learned (L4) Norwegian and found mixed evidence for cross-language generalisation to her other languages; for example, gains were noted in her earlier-learned English (L2) in semantic and syntactic subtests of the Bilingual Aphasia Test (BAT; Paradis and Libben, 1987), but not in verb naming nor on sentence and discourse variables in connected speech production.

In a study that examined cross-language generalisation in multiple participants, Croft and colleagues enrolled five Bengali-English bilinguals with aphasia in two phases of therapy, one in each of their languages, sequentially (Croft et al., 2011). All participants were from Bangladesh and were living in England at the time of the study; all were at least six months post a left CVA resulting in aphasia and all experienced lexical retrieval difficulties in both their languages. Treatment targeted word retrieval using semantic and phonological strategies. Testing prior to intervention was extensive; testing following treatment focused on picture naming. Results demonstrated that two participants benefitted from semantic-based treatment and gains were noted in both their languages; four participants benefitted from phonological-based treatment, one of them showing gains only in her first language; and one participant did not show treatment-related improvement. Critically, three of the five participants demonstrated cross-language generalisation to the untreated language following semantic-based treatment. Similarly, mixed results were reported in Kiran and Roberts (2010), with one of their four participants showing clear cross-language generalisation following treatment.

The mixed pattern of results regarding cross-language treatment generalisation raises the questions of what might predict whether cross-language generalisation will be obtained and how it is best measured. Regarding predictors of cross-language treatment effects, several variables have been implicated. These include the domain targeted in therapy (e.g. lexical retrieval vs. morphosyntactic structure), the degree of linguistic similarities between the languages in question (e.g. cognates vs. non-cognate translation equivalents; languages from the same family, such as Spanish and Catalan vs. distant languages, such as English and Chinese), and the participants' relative language proficiency and language use in each language prior to and following the aphasia onset.

Treatments that target underlying processes (for example semantic organization, as in Semantic Feature Analysis, e.g. Boyle and Coelho, 1995) or overall production fluency (as in, for example, Oral Reading and Language for Aphasia [ORLA], e.g. Cherney, 2004), may be predicted to promote generalisation across languages more than might treatments that focus on specific lexical items or syntactic structures. However, most treatment studies have reported treatment outcome in terms of the treated aspects, such as accuracy of lexical retrieval (e.g. Croft et al., 2011; Edmonds and Kiran, 2006; Kohnert, 2004; Lalor and Kirsner, 2001). Few studies have reported untreated aspects, such as overall test scores and severity levels (e.g. Knoph, 2013; Miller Amberber, 2012), or generalisation to connected speech production (Altman et al., 2012; Knoph et al., 2015). Particularly little is known about change in overall efficiency of language production, especially in mild aphasia, among bilingual individuals.

One variable that has been put forward as a critical variable in determining the likelihood of cross-language generalisation following intervention is language proficiency, though researchers diverge in the predictions they have generated and in the findings they have reported. Specifically, it has been suggested that cross-language generalisation is more likely to be observed between languages of comparable proficiency (Edmonds and Kiran, 2006; Goral, 2012), and from a weaker language to a stronger language (Edmonds and Kiran, 2006) or from a stronger language to a weaker one (Goral, 2012; Miertsch et al., 2009), and several studies have demonstrated that proficiency had little predicting value on cross-language effects (Knoph, 2013; Knoph et al., 2015).

Goral and colleagues have suggested that cross-language effects can lead to transfer of treatment benefits but may also lead to an increase in between-language competition, thus yielding temporary negative cross-language effects (Goral, Naghibolhosseini, and Conner, 2013). Such a view, consistent with Green's Inhibitory Control theory (e.g. Green, 1998), can account for the lack of treatment-related improvement reported in several studies (e.g. Abutalebi et al., 2009). That is, it is possible that both influences are at play: cross-language generalising of treatment benefits is modulated by decreased overall activation levels of the untreated language. It is yet to be determined how positive cross-language treatment transfer, on the one hand, and negative effects resulting from increased activation levels of the treated language and decreased activation levels of the untreated language, on the other hand, may work in concert to yield the results reported in the literature.

Another variable that may influence cross-language generalisation is languages' typological distance. This may be particularly relevant for individuals who speak three languages or more, whereby multiple languages may be differentially inhibited or facilitated based on their typological relationship to the more activated (treated) language (Mosca, 2017). Theoretical support for this variable is based on De Bot's 2004 Multilingual Processing Model, which proposes that production in one language yields heightened activation of language-specific elements as well as elements that may overlap among a speaker's languages. The variable outcome of studies of cross-linguistic generalisation of aphasia treatment in bilingual speakers of language with shared typology (e.g., facilitation reported in Kohnert, 2004; inhibition reported in Kurland & Falcon, 2011) may be due to the relationship between the target of treatment (e.g., cognates, semantic features, verb argument) and shared linguistic elements.

In our current work, we had the opportunity to examine the role of language proficiency in cross-language generalisation in a multilingual individual with aphasia by testing his languages of higher and lower proficiency levels following treatment. We generated the following prediction on the basis of psycholinguistic accounts of bilingual representation and processing (e.g. Costa, Santesteban, and Ivanova, 2006; Dijkstra, De Bruijn, Schriefers, and Ten Brinke, 2000; Guo and Peng, 2006). Treatment in a language of high proficiency --- by prompting greater co-activation of languages of comparable (high) proficiency than of languages of differing proficiency levels --- would lead to cross-language generalisation among languages of comparable proficiency. This expectation is commensurate with the findings of non-selective language activation for highly proficient bilinguals (e.g. Dijkstra et al., 2000; Kroll, Bobb, Misra and Guo, 2008).

In addition, language co-activation may result not only in cross-language treatment generalisation for languages of higher proficiency but also in cross-language intrusions for languages of lower proficiency. That is, greater activation of highly proficient languages may result in the intrusion of lexical items from these highly activated languages while attempting to speak in the less activated languages. Several studies have reported language mixing behaviours among their participants (e.g. Ansaldo, Ghazi Saidi and Ruiz, 2010; Croft et al., 2011; Keane and Kiran, 2015), but did not set out to examine instances of code-mixing as a measure of cross-language treatment effects. Keane and Kiran (2015) reported increased cross-language intrusion errors in their participant's naming performance in his second and third languages, English and French, following treatment. The authors noted greater frequency of naming in English during French testing following treatment in English, and in French during English testing following treatment in French. The authors did not observe cross-language treatment generalisation following either treatments.

Whereas such intrusion errors can be taken as aphasia-related behaviour, switching between languages in the same sentence (code-mixing) or between sentences (code-switching) are common phenomena among multilingual speakers, although habits of mixing vary greatly within- and between-individuals (Myers-Scotton, 2002). The production of target words in the non-target language may be especially common between languages that are linguistically similar (e.g. Ecke, 2015, Mosca, 2017). Switching languages among bilinguals with aphasia has been considered both a sign of language impairment (e.g. Perecman, 1984) as well as a strategy to improve communication (Grosjean, 1985; Muñoz, Marquardt, and Copeland, 1999). In a few studies, explicit use of switching, or translation, has been encouraged as a therapeutic tool, i.e. to facilitate word retrieval of translation equivalents (Ansaldo et al., 2010). In contrast, when the switching is unintended, or when the interlocutor does not share that language with the speaker, switching can be detrimental to communication. Thus, decreased numbers of unintended instances of language mixing and switching can serve as a measure of production success among bilingual individuals with aphasia.

In the current research study, we asked whether we could capture cross-language treatment generalisation by measuring efficiency in connected speech production of a multilingual person with mild aphasia. We defined efficiency in two ways, one concerned a calculation of correct information units (CIU; Nicholas and Brookshire, 1993) per minute, the other concerned instances of language switching (frequency of lexical insertion), suggesting cross-linguistic interference. We provided treatment in one language (L1) of the participant and tested his L1 as well as six of his untreated languages. We administered treatment that targeted general processes of efficient language production and we hypothesised that the concurrent activation of highly proficient languages would facilitate cross-language generalisation. In contrast, languages of lower proficiency would be less likely to be active during L1 processing and so would likely benefit less from treatment in L1. As well, we predicted that language similarity may have a positive effect on cross-language generalisation. Nevertheless, it is also possible that the type of treatment we employed, which did not focus on language-specific items and structures, would lead to cross-language generalisation regardless of whether the languages shared more or less syntactic structures or lexical items. Finally, we predicted increased cross-language intrusion errors during production in languages of lower proficiency as outlined below.

The following research questions guided our investigation:

1. Does treatment in L1 generalise to the non-treated languages and is generalisation modulated by language proficiency and linguistic similarity?
2. Does treatment in L1 result in increased co-activation of languages of high proficiency yielding cross-language interference (i.e. increased language mixing) in languages of lower proficiency?

Based on theories of bilingual language activation we predicted the following:

1. Treatment in L1 will generalise to the non-treated languages, such that there will be increased efficiency in language production. Based on theories of proficiency-dependent and similarity-dependent language co-activation, greater generalisation is expected for languages of higher proficiency than for languages of lower proficiency, and for languages that are linguistically similar.
2. Based on the co-activation of highly-proficient languages and in the absence of activation of languages of lower proficiency, an increased interference from languages of higher proficiency into languages of lower proficiency is expected following treatment.

METHOD

Participant

The participant is a Belgium-born 65-year-old man, DN. He acquired Dutch (Flemish) at home, and was exposed to Dutch, French and German growing up in a trilingual region, becoming fluent in these three languages in childhood. He has continued to use these three languages with family and friends, and for business. He was also exposed to Afrikaans at home, from books and magazines and as it was spoken for business by his father. He learned English at age 15 and Italian at age 16 and became fluent in these two languages within a couple of years. He learned Danish, Swedish, and Norwegian in his mid-30s and Spanish and Portuguese in his 40s. He has continued to use all his languages with friends and family, for travel, and for business. He had started learning Russian but did not gain high proficiency. At the time of the study, DN rated his language proficiency prior to the stroke as 7 or above on a 1-9 scale (where 9=native like proficiency) in five of his languages: Dutch, English, French, German, and Italian; 6 in Spanish, and 4 in Norwegian. He completed high school, college, and a law degree.

Schooling was completed in Dutch with the exception of a couple of classes taught in French. Table 1 provides the age of learning, self-rated proficiency, and the amount of use of each of his languages.

Approximately a year prior to enrolment in the study, DN sustained a left CVA resulting in a lesion extending to the left frontal and Sylvian fissure areas and left internal capsule. He was diagnosed with a transcortical motor aphasia, characterised by good comprehension and good repetition skills, with impaired fluency and word finding, as well as neurogenic stuttering characterised by frequent episodes of sound repetition, prolongation or vocal disruption accompanied by facial and bodily tension. Administration of the WAB in English three and four months post onset revealed an Aphasia Quotient of 75.4 and 89.6, respectively.

Insert table 1 about here

Treatment

Before beginning the study, the testing and treatment procedures were explained to the participant and he signed an informed consent. Forty hours of treatment in Dutch, DN's first language, were completed over the course of five weeks via Skype ™. The number of hours per week ranged from six to ten, to accommodate the participant's availability. Treatment was administered by a licensed speech-language therapist, a native speaker of Dutch. The objectives of treatment were to improve efficiency of spontaneous speech in responsive as well as in conversational contexts. We selected Oral Reading for Language in Aphasia (ORLA; Cherney, 2004) for several reasons. This method is particularly well suited for telerehabilitation and has been used successfully with mild to moderate non-fluent aphasia (Cherney, 2010a; b). In addition, by design this oral reading treatment focuses on connected discourse and provides practice with speech rhythm and pacing, which we expected to be particularly beneficial to our participant given his frequent hesitations and blocking during oral language production (Cherney, 2010b). We modified the original protocol to suit the needs of our participant and employed the following six steps: reading single paragraphs aloud, locating words in the text, using words in sentences, reading multiple paragraphs independently, summarising paragraphs, and answering content questions. The reading material consisted of selected articles from newspapers and magazines covering current events, travel, politics and economics, all areas of interest for the participant. To promote fluency and reduce pausing, the participant was encouraged to read using exaggerated prosody, a technique that was modelled by the clinician and then gradually faded (Norton, Zipse, Marchina, and Schlaug, 2009). See Appendix A.

Pre-and post-treatment measurements

We tested DN in seven of his languages: Dutch, French, English, German, Spanish, Italian, and Norwegian[1]. For pre- and post-treatment measures, each language was assessed on three different days. Each testing session lasted approximately one hour for a total of 21 testing hours pre-treatment and 21 post-treatment. The order of languages was pseudo-randomised to reduce order and fatigue effects (see table 2). Testing was administered by seven clinicians or student clinicians who were native or highly proficient speakers of the respective language. Because some of the testers were not physically located in the same city, all testing was done using computer video calls with Skype™. The computer screen was shared such that the

[1] Although it would have been interesting to include all of DN's languages, we considered clinician availability and the testing time required and selected those that would address our research questions.

participant was able to view the test stimuli on the examiner's screen. All sessions were audio- and video-recorded.

The testing items ---detailed below---were selected to evaluate generalisation of treatment effects using comparable items across languages. Only the first task was practised during treatment, although the questions he answered were unrelated to the treated material.

1. Answering questions: The participant was given six wh-questions and was instructed to answer each question in one sentence.
2. Picture description: The picture description task of the BAT was administered, instructing the participant to produce a story based on a panel of six drawings.
3. Narrative production: The participant was asked to talk for a few minutes about a topic (e.g. a recent vacation).
4. Question generation: The participant was presented with six cartoons and was instructed to generate a question that the person in the picture might be asking.

A different topic was given in each of the three testing days for subtest 3. A different order of the same items was administered for the remaining subtests.

Insert table 2 about here

Outcome measures

All responses to the testing items were transcribed and scored by the examiner or by a trained research assistant highly proficient in the language. Prior to scoring, the transcriptions were coded to blind the scorers to the testing time. The responses to the narrative and picture description tasks were divided into utterances by identifying each Analysis of Speech Unit (AS-Unit; Foster, Tonkyn, and Wigglesworth, 2000, p. 366). The AS-Unit determination considers intonation, pausing, syntax and meaning. (For a complete description see Kempler and Goral, 2011.) The duration of each utterance was then measured manually using sound-editing software. Comments made by the examiner, if present, were excluded for the recorded duration.

To capture post-treatment change across languages we measured Correct Information Units (CIU)/min. as a measure of production efficiency for each of the tasks (Nicholas and Brookshire, 1993). We defined CIUs as words that were intelligible in context, accurate in relation to the picture(s) or topic, relevant, and informative. Comments on the task, false starts, fillers, repetitions or revisions were excluded. Responses did not have to be grammatically correct to be considered in the CIU count, but they did have to be in the target language. The CIUs were summed and divided by the sum of the duration of the utterances as defined by the AS-Unit, multiplied by 60, and then averaged across the three testing times for each task and language.

In addition, we assessed cross-linguistic interference by calculating the percentage of whole words that were code-mixed. Production of an entire word in a non-target language was counted. Although there were some instances of other types of code-mixing (base in target language, stem in non-target language), they occurred minimally in the corpus, and we did not include these in our count. The percentage of code-mixed words was calculated by dividing the number of instances of code-mixing by the total number of verbal units (defined as words or part words in the target or non-target languages).

Statistical analyses:

We employed quantitative non-parametric analysis to assess the effects of the intervention on the target and non-target languages, using nonoverlap statistics. *Nonoverlap of All Pairs* (NAP) was the principal measure of effect size. The NAP is the most complete of the nonoverlap indices as it makes use of all pairwise comparisons (Kratochwill and Levin, 2014) and is appropriate for use with small number of observations. The NAP has demonstrated high calculation accuracy and good external validation with the R^2 standard when compared with other nonoverlap statistics (Parker and Vannest, 2009). Each of the three pre-treatment baseline observations was paired with each of the three post-treatment observations to make 9 pairs (i.e. $N = n_A * n_B$). We counted the number of Positive (P), Negative (N), and Tied (T) pairs, summed P and ½ T, and divided that sum by the total number of pairs.

$$NAP = \frac{n - n_0}{n} * 100 = \frac{P + .5T}{P + T + N} * 100$$

$$NAP\ Rescaled = 2 * \mathrm{NAP} - 1$$

Because NAP is an AUC (area under the curve) statistic, the probability is reduced to chance level at 0.5. Therefore, we rescaled the NAP calculations to reflect a 0 to 1.0 range (with a negative NAP indicative of post worse than pre). Parker and Vannest suggested the following tentative ranges of 0.85--1.00 = strong effects; 0.32--0.84 = moderate effects; below 0.31 small effects. We adopted a stricter criterion and reduced the moderate effect range by 50% yielding a range of 0.58--0.84. We report here moderate and strong effects. Because NAP is new in our field, for reader reference we calculated effect size that has been used traditionally in treatment studies (Beeson and Robey, 2006), according to the formula below.

$$d = \frac{(\mathrm{M}_2 - \mathrm{M}_1)}{\sigma_1}$$

RESULTS

For each language, we assessed language efficiency using CIUs/min for the four production tasks: Narrative, Picture Description, Answering Questions, and Question Generation. The results are presented in table 3 and summarised in figure 1. Language interference was assessed by the percentage of code-mixed words. (See table 4 and figure 2).

We report below results based on the NAP analyses. The tables include also effect sizes, which were generally small but consistent with the NAP results.

Language efficiency

On measures of CIUs/min, DN showed a significant improvement in the treated language, Dutch, as well as in his four languages of higher proficiency. Average production efficiency pre- to post-treatment increased in Dutch for three of the four tasks: Picture Description and Answering Questions (medium effects), and for Question Generation (strong effects). In untreated languages, cross-linguistic treatment effects were apparent on four tasks in French: Narrative, Picture Description, Answering Questions, and Question Generation (all strong effects); three tasks for Italian: Narrative, Picture Description (medium effects), and Answering Questions (strong effects); two tasks in English: Narrative and Picture Description (strong effects); and two for German: Picture Description and Question Generation (medium effects).

Gains in efficiency were absent, apart from one task, in his languages of lower proficiency. In Spanish, none of the four tasks yielded significant gains. In Norwegian, he improved markedly on the Question Generation task, however not on the other three tasks.

Language interference

Code-mixing of whole words was minimal for DN's languages of higher proficiency, Dutch, English, French, German, and Italian: either no words or 1--2 words were mixed within a given task and therefore change was not examined. As we anticipated, for languages of lower proficiency his frequency of code-mixing on average across the two testing blocks was greater for Norwegian ($M = 0.21$), his weakest language, than for Spanish ($M = 0.05$). However, the change in the frequency of interference pre-to-post treatment was found for Spanish but not for

Norwegian (see table 4). In Spanish, a significant increase in code-mixing was found for two of the four tasks post treatment, and largely to Italian (example 1). In Norwegian there was no significant change and code-mixed words were to multiple languages (example 2).

(1) El marido *cade* del *albero*.
[Italian] [Italian]

The husband falls from the tree.

(2) *der* mand vil *zie* zine */føgel/*[*] [pause] *der* mand vil zine [pause] */føgel/* … *take away*
[Dutch] [Dutch] [Dutch] [English]

The man wants his birds [pause], wants his birds [pause] take away.

Insert tables 3 and 4 and figures 1 and 2 about here

DISCUSSION

In this study, we examined the cross-linguistic effects on six languages in a multilingual man with mild non-fluent aphasia, following treatment of a seventh, his native Dutch (L1). We asked whether improvement in Dutch would transfer to his other languages and whether it would be dependent upon proficiency and/or linguistic similarity to his other languages. We also evaluated the degree and nature of any change in cross-linguistic interference by examining code-mixing behaviour. Following treatment, his improved efficiency for Dutch production generalised to his four other languages of high post-stroke proficiency but minimally to the two of lower post-stroke proficiency. Increased code-mixing was asymmetric, likely due to the interaction between proficiency and linguistic similarity.

---

[*] His pronunciation may be influenced by Danish, Dutch and German.

Language efficiency

To address generalisation of improved language efficiency within and across languages we examined changes in the number of CIUs (relevant words) DN produced per minute on each of four production tasks: Narratives, Picture Description, Question Generation, and Answering Questions in seven of his languages. The treatment we administered targeted overall production fluency, rather than specific lexical items or syntactic structures, and we therefore expected it to generalise to the non-treated languages. As expected, we found change in language efficiency in the treated as well as generalisation to the non-treated languages.

We interpreted the extent of generalisation by examining the number of tasks that showed improvement and the magnitude of that improvement. Accordingly, our prediction of greater generalisation for languages of higher proficiency than for languages of lower proficiency was confirmed. In comparison to the languages of lower proficiency, Spanish and Norwegian, improvement in his highly proficient French, English, Italian, and German paralleled improvement in his treated L1, Dutch. Our results are consistent with a cross-linguistic generalisation of treatment gains reported in the literature (e.g. Edmonds and Kiran, 2006; Goral et al., 2010; Knoph, 2013; Kohnert, 2004; Kurland and Falcon, 2011) and lend strong support to the suggestion that cross-linguistic transfer is more likely to occur in languages of comparable proficiency (Edmonds and Kiran, 2006; Goral, 2012), possibly due to non-selective activation of bilinguals' languages of high proficiency (e.g. Dijkstra et al., 2000; Kroll, Bobb, Misra and Guo, 2008).

We further predicted greater cross-language treatment generalisation between languages that are close, typologically, and share linguistic elements. However, in contrast to this prediction, our results point to greater cross-linguistic transfer to a more dissimilar language.

Specifically, we observed the least robust generalisation from Dutch to German, a language closer in linguistic distance to DN's L1 than his other highly proficient languages, while the more distant, French, showed the most robust gains. This pattern of results may be surprising, given previous findings of cross-language generalisation between similar languages (e.g. Goral et al., 2010; Kohnert, 2004). A possible account for the less robust generalisation to similar languages could be an increased between-language competition resulting from linguistic similarity (e.g. de Bot, 2004; Kurland and Falcon, 2011; Mosca, 2017).

Thus, in answer to our first research question, treatment in L1 generalised to the non-treated languages, as was evident in increased efficiency in language production and, as we predicted, greater generalisation was found for languages of higher proficiency than for languages of lower proficiency. Linguistic similarity appeared to play a role in cross-language treatment effects, but inconsistent with our prediction, we observed greater generalisation to the more dissimilar languages. Taken together, within this multilingual speaker with aphasia, a shared conceptual system may have supported generalisation among languages of equal proficiency, yet word retrieval difficulties in the presence of linguistic similarity may have reduced the extent of that generalisation by inter-language competition.

Our results demonstrate the feasibility of measuring efficiency of language production to document treatment-related improvement in aphasia. This is particularly relevant to individuals with mild aphasia who typically score well on structured tests (Penn and Beecham, 1992). It is also of note that this production-efficiency measure is a useful way to assess generalisation from treatment to elicited connected speech production. We maintain that the cross-language generalisation observed in our participant is substantial because skills generalised from the in-session treatment activities and material to increased efficiency in our outcome measures which were unrelated, unpractised language production tasks.

Language interference

We predicted that due to the increased activation of DN's languages of higher proficiency and the greater word retrieval difficulty in his languages of lower proficiency, those languages of lower proficiency would be more vulnerable to cross-linguistic interference than those of higher proficiency. Indeed, both before and after Dutch treatment, the frequency of code-mixing was negligible in his five highly proficient languages and greater in the languages of lower post-stroke proficiency. Consistently, within his languages of lower proficiency, he code-mixed more when attempting to produce information in the weaker language, Norwegian, than in Spanish. However, a change in cross-linguistic interference following treatment was not parallel in the two languages. Unlike Keane and Kiran (2015) who reported increased intrusion errors in the two treated languages of their trilingual participant, we observed a different pattern of post-treatment language mixing in two untreated languages of our participant. DN's significant increase in frequency of code-mixing in Spanish not present in Norwegian suggests differential effects of L1 treatment on the two languages. A source of potential interference for Spanish is its linguistic similarity to Italian, unlike Norwegian that shares fewer cognates and is more linguistically distant to DN's other languages (Dyen, Kruskal, and Black, 1992). This leads us to suggest that the post-treatment activation and improvement in Italian resulted in greater between-language competition when attempting to speak Spanish, and consequently in greater frequency of code-mixing when responding in Spanish. Taken together, we posit that interference was modulated by an interaction between language proficiency and linguistic similarity.

This interaction may explain a lack of generalisation, reported in the literature, between languages that are more closely related to each other. For example, Abutalebi and colleagues examined treatment effects for JRC, a Spanish-Italian bilingual, using fMRI and behavioural measures (Abutalebi et al., 2009). The participant's Italian (L2), improved but apparently at the expense of his Spanish (L1) which showed no generalisation and decreased accuracy on portions of the Bilingual Aphasia Test. The authors concluded an impairment of language selection and control negatively influenced JRC's performance in Spanish due to interference from the treated and more active Italian. In contrast, many of the studies that reported cross-language generalisation found it between languages of varying linguistic distance (e.g. Bengali and English in Croft et al. 2011; Arabic and English in Knoph, 2013)

We contrast our DN's antagonistic Italian-Spanish recovery with the lack of change in code-mixing in his other language of lower proficiency, Norwegian. Unlike the Spanish-Italian proximity, Norwegian, although Germanic (like German, Dutch and English), has greater lexical distance relative to any of DN's five languages of higher proficiency. Indeed, whereas the primary language of code-mixing was to Italian for Spanish, for Norwegian code-mixed words came from various languages. Thus, in response to our second research question and consistent with our prediction, increased interference in the non-treated languages appeared greater for languages of lower proficiency than for language of higher proficiency, and between languages that are linguistically similar. It is also possible that Norwegian, for which DN had reported only minimal change post the aphasia onset, was less active even prior to the stroke and was thus less affected by the increased activation of the other languages following the treatment we administered. Future studies may be able to dissociate the contribution of proficiency, activation, and language similarity to cross-language effects following treatment.

In summary, as a multilingual individual with aphasia, DN has given us a unique opportunity to observe three effects of treatment generalisation on languages operating in concert. The first is transfer of treatment effects of improved efficiency, which can be taken as evidence for co-activation (Kroll et al., 2008) and overlapping neural representation (Green and Abutalebi, 2008) of these languages. In this case, treatment in L1 benefitted not just one but multiple languages of comparable proficiency. The second process is inhibition. That is, improvement in one language may occur at the expense of another (cf. Abutalebi et al., 2009). In this case, improvement in Italian may have temporarily inhibited DN's Spanish. The third process is the interaction between linguistic proficiency and similarity and is reflected in our data in two ways. Improvement in Dutch transferred to all his languages of high proficiency, benefitting French --- the language most dissimilar to Dutch --- the most, and German --- the language most similar --- the least. In the case of cross-linguistic generalisation, thus, comparable proficiency increased the likelihood of benefit and linguistic similarity decreased the likelihood of benefit. In the case of cross-linguistic interference, our data suggest that unequal proficiency may increase the probability of interference for languages with linguistic similarity.

Conclusion

Following treatment in Dutch, DN's first language, his efficiency of language production improved in the treated language as well as in his other languages of higher proficiency, with reduced or minimal gains for his languages of lower proficiency. In addition to the benefit that the treatment provided, we also found evidence for increased interference as reflected in greater word mixing for one of his languages of lower proficiency but not the other. We suggest that this selective interference resulted from an interaction between language proficiency and language similarity.

This study has limitations. Evaluating performance in seven languages posed logistical hurdles in managing data collection and scoring. The 21 sessions our participant generously completed pre- and post-treatment yielded three sessions per language, limiting the number of data points for analysis. In addition, it was challenging to obtain further clinicians to rescore the data to complete interrater reliability in all the languages. The efficiency measure we employed included measuring the duration of the participant's production in each language and was labour intensive and time consuming, and therefore, although highly effective in capturing treatment-related change, may be less useful clinically. As with any single subject design, there is limited generalisability of data from a single individual and additional data from similar and dissimilar languages of varying levels of proficiency used by multilingual individuals with aphasia are needed to support our interpretation. As well, the increased activation of DN's multiple languages during the week of testing may have contributed to the patterns of interference observed. Nevertheless, because code-mixing did not seem to be present in the languages of higher proficiency, despite the testing schedule, we are comfortable attributing the interference observed during Spanish production to the unique combination of increased efficiency in Italian and the similarities between Spanish and Italian. The extent to which shared linguistic elements increase the likelihood of cross-language interference and may diminish cross-language generalisation warrants further study.

ACKNOWLEDGEMENTS

Acknowledgements to be provided after review.

STATEMENT OF INTEREST

The authors have no financial or non-financial declarations of interest.

REFERENCES

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

Altman, C., Goral, M., and Levy, E. S. (2012). Integrated narrative analysis in multilingual aphasia: The relationship among narrative structure, grammaticality, and fluency. *Aphasiology, 26,* 1029–1052.

Ansaldo, A. I., Ghazi Saidi, L., and Ruiz, A. (2010). Model-driven intervention in bilingual aphasia: Evidence from a case of pathological language mixing. *Aphasiology, 24,* 309–324.

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

Boyle, M. and Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American journal of speech language pathology, 4,* 94–98.

Cherney, L. R. (2004). Aphasia, alexia, and oral reading. *Topics in stroke rehabilitation*, *11*(1), 22-36.

Cherney, L. R. (2010a). Oral reading for language in aphasia: impact of aphasia severity on cross-modal outcomes in chronic nonfluent aphasia. *Seminars in speech and language*, *31*(1), 42-051.

Cherney, L. R. (2010b). Oral reading for language in aphasia (ORLA): Evaluating the efficacy of computer-delivered therapy in chronic nonfluent aphasia. *Topics in stroke rehabilitation, 17*(6), 423-431.

Costa, A., Santesteban, M., and Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of experimental psychology: Learning, memory, and cognition*, *32*(5), 1057.

Croft, S., Marshall, J., Pring, T., and Hardwick, M. (2011). Therapy for naming difficulties in bilingual aphasia: Which language benefits? *International journal of language and communication disorders, 46*, 48–62.

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of multilingualism*, *1*(1), 17-32.

Dijkstra, T., De Bruijn, E., Schriefers, H., and Ten Brinke, S. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and cognition*, *3*(1), 69-78.

Dyen, I., Kruskal, J. B., and Black, P. (1992). An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American philosophical society*, *82*(5), iii–132.

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: language and cognition*, *18*(2), 145-162.

Edmonds, L., and Kiran, S. (2006). Effect of semantic naming treatment on crosslinguistic generalization in aphasia in bilingual or multilingual speakers. *Journal of language speech, and hearing, 49*, 729–748.

Faroqi-Shah, Y., Frymark, T., Mullen, R., and Wang, B. (2010). Effect of treatment for bilingual individuals with aphasia: A systematic review of the evidence. *Journal of neurolinguistics, 23,* 319–341.

Foster, P., Tonkyn, A., and Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, *21*(3), 354–375.

Goral, M. (2012). Cross-language treatment effects in multilingual aphasia. In M. R. Gitterman, M. Goral, and L. K. Obler (Eds.), *Aspects of multilingual aphasia* (pp. 106–121). Bristol: Multilingual Matters.

Goral, M., Levy, E. S., and Kastl, R. (2010). Cross-language treatment generalisation: A case of trilingual aphasia. *Aphasiology, 24,* 170–187.

Goral, M., Naghibolhosseini, M., and Conner, P. S. (2013). Asymmetric inhibitory treatment effects in multilingual aphasia. *Cognitive neuropsychology, 30*, 564–577. doi:10.1080/02643294.2013.878692

Goral, M., Rosas, J., Conner, P. S., Maul, K. K., and Obler, L. K. (2012). Effects of language proficiency and language of the environment on aphasia therapy in a multilingual. *Journal of neurolinguistics, 25,* 538–551.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition, 1*, 67–81.

Green, D. W. and Abutalebi, J. (2008). Understanding the link between bilingual aphasia and language control. *Journal of neurolinguistics*, *21*(6), 558–576.

Grosjean, F. (1985). Polyglot aphasics and language mixing: A comment on Perecman (1984). *Brain and language*, *26*, 349–355.

Guo, T., and Peng, D. (2006). Event-related potential evidence for parallel activation of two languages in bilingual speech production. *NeuroReport*, *17*(17), 1757-1760.

Keane, C., and Kiran, S. (2015). The nature of facilitation and interference in the multilingual language system: Insights from treatment in a case of trilingual aphasia. *Cognitive neuropsychology*, *32*(3-4), 169-194.

Kempler, D., & Goral, M. (2011). A comparison of drill-and communication-based treatment for aphasia. *Aphasiology*, *25*(11), 1327-1346.

Kiran, S., and Iakupova, R. (2011). Understanding the relationship between language proficiency, language impairment and rehabilitation: Evidence from a case study. *Clinical linguistics and phonetics, 25(6–7)*, 565-583.

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

Kiran, S., Sandberg, C., Gray, T., Ascenso, E., & Kester, E. (2013). Rehabilitation in bilingual aphasia: evidence for within-and between-language generalization. *American journal of speech-language pathology*, *22*(2), S298-S309.

Knoph, M. I. K. (2013). Language intervention in Arabic–English bilingual aphasia: A case study. *Aphasiology, 27,* 1440–1458.

Knoph, M. I., Lind, M., and Gram Simonsen, H. (2015). Semantic feature analysis targeting verbs in a quadrilingual speaker with aphasia. *Aphasiology*, *29*(12), 1473–1496.

Kohnert, K. (2004). Cognitive and cognate-based treatments for aphasia in bilingual or multilingual speakers: A case study. *Brain and language, 91,* 294–302.

Kohnert, K., and Peterson, M. (2012). Generalization in bilingual aphasia treatment. In M. R. Gitterman, M. Goral, and L. K. Obler (Eds.), *Aspects of multilingual aphasia* (pp. 89–105). Bristol: Multilingual Matters.

Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances.* American Psychological Association.

Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta psychologica*, *128*(3), 416-430.

Kurland, J., and Falcon, M. (2011). Effects of cognate status and language of therapy during intensive semantic naming treatment in a case of severe nonfluent bilingual aphasia. *Clinical linguistics and phonetics*, 25(6-7), 584–600.

Laganaro, M., and Overton Venet, M. (2001). Acquired alexia in multilingual aphasia and computerassisted treatment in both languages: Issues of generalisation and transfer. *Folia phoniatrica et logopaedica, 53,* 135–144.

Lalor, E., and Kirsner, K. (2001). The role of cognates in aphasia in bilingual or multilingual speakers: Implications for assessment and treatment. *Aphasiology, 15,* 1047–1056.

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

Miertsch, B., Meisel, J. and Isel, F. (2009). Non-treated languages in aphasia therapy of polyglots benefit from improvement in the treated language. *Journal of neurolinguistics 22,* 135–150.

Miller Amberber, A. (2012). Language intervention in French–English bilingual aphasia: Evidence of limited therapy transfer. *Journal of neurolinguistics, 25,* 588–614.

Mosca, M. (2017). *Multilinguals' language control.* Dissertation. Doctoral thesis. urn:nbn:de:kobv:517-opus4-398912.

Muñoz, M. L., Marquardt, T. P., and Copeland, G. (1999). A comparison of the codeswitching patterns of aphasic and neurologically normal bilingual speakers of English and Spanish. *Brain and language, 66*, 249–274.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press on Demand.

Nicholas, L. E., and 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.

Norton, A., Zipse, L., Marchina, S., and Schlaug, G. (2009). Melodic intonation therapy. *Annals of the New York academy of sciences*, *1169*(1), 431–436.

Paradis, M., and Libben, G. (1987). *The assessment of bilingual aphasia.* Hillsdale, NJ: Erlbaum

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior therapy*, *40*(4), 357-367.

Perecman, E. (1984). Spontaneous translation and language mixing in a polyglot aphasic. *Brain and language*, *2,* 43–63.

Penn, C., and Beecham, R. (1992). Discourse therapy in multilingual aphasia: a case study. *Clinical linguistics and phonetics*, *6*(1-2), 11–25.

Appendix A

Treatment Protocol:

Objectives:
1. Increase the duration of maintenance of production rate both in reading and in speaking.
2. Improve word retrieval in responsive and spontaneous speech.
3. Decrease difficulties with speech initiation.

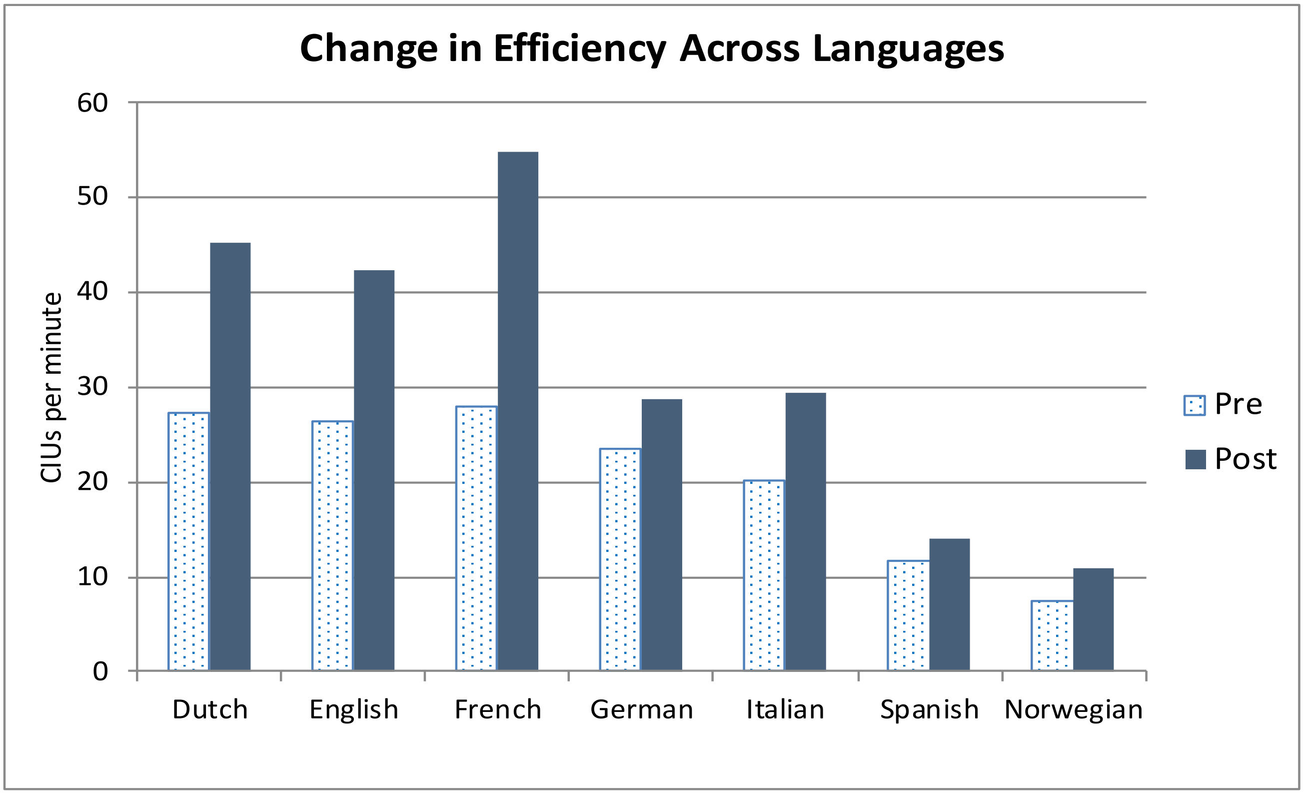
General Instructions

Prior to the activity the clinician models reading with exaggerated prosody. Introduction of the method can be implemented using multisyllabic words, then phrases and then sentences. Cueing consists of modelling the modified (exaggerated) prosody and asking for a repetition and/or cueing for faded or independent use of the modified prosody.

In cases of prolonged pausing, the clinician will offer 1 of 3 options: 1) ask him to start a different sentence, 2) give him a choice of 2 words to use, and 3) direct him to the written text to find the word he is searching for. In addition, if the participant uses a reduced or more laboured rate after a period of fluency, the clinician will remind him to use exaggerated prosody.

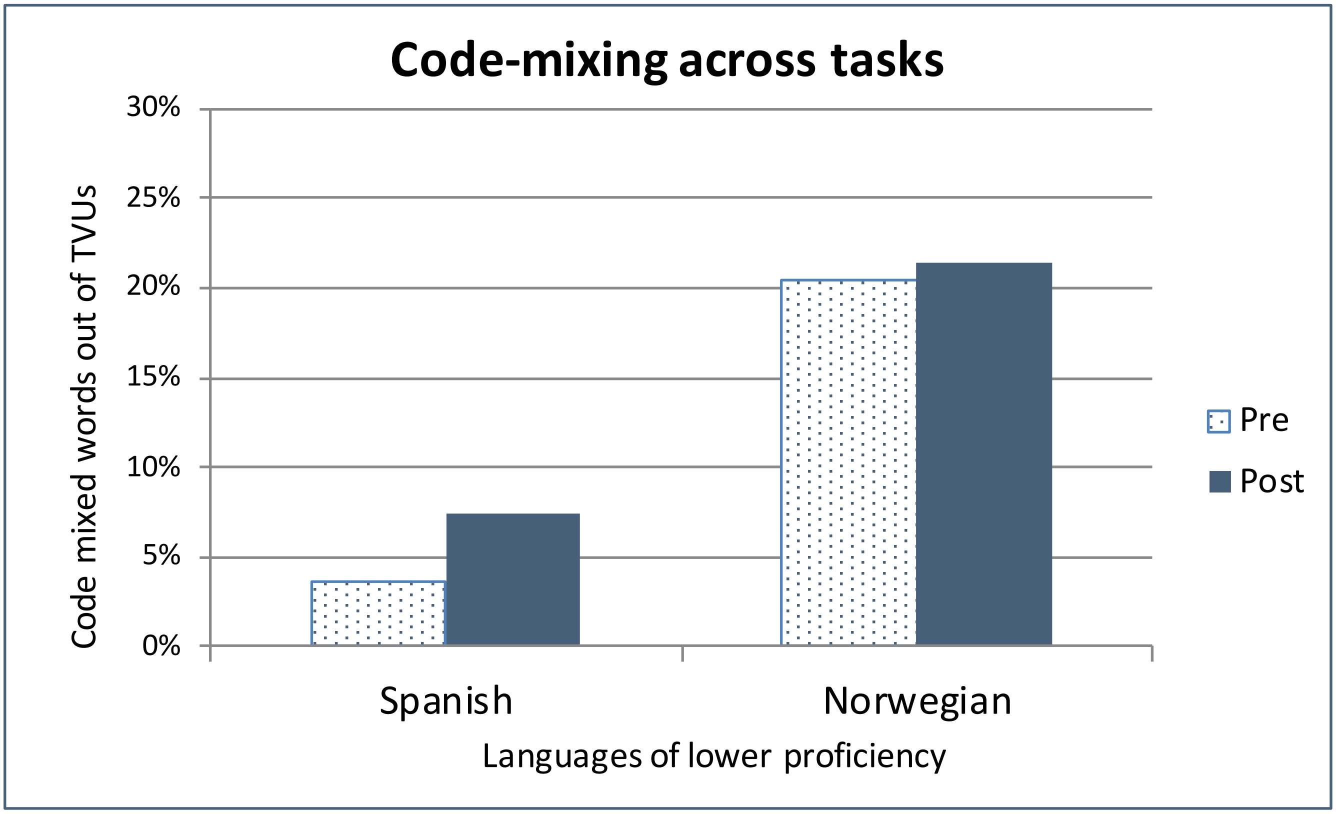
Activities	**Objective**	**Procedure**
Reading aloud	Read separate paragraphs with uniformity of rate and modified prosody	The participant reads the paragraph independently with exaggerated prosody.
Locating words in text	Coherent and fluent description of the location of words; phrase or sentence level production with exaggerated prosody	For each separate paragraph, the participant and clinician take turns selecting a word and asking the other person to locate the word. Once the word has been located, the sentence is read aloud.
Using words in sentences	Fluent use of words in sentences; sentence level production	For each separate paragraph, the participant and clinician take turns selecting a word from the text and asking the other person to use it in a sentence. The sentence is produced with exaggerated prosody.
Reading complete text (i.e. three paragraphs) independently	Practise reading multiple paragraphs with uniformity of rate and modified prosody	The participant reads the paragraphs independently with exaggerated prosody.
Summarising/retelling the paragraphs	Summarisation of the main idea of each paragraph.	The participant gives a one to two sentence summary of each paragraph of the reading passage, using exaggerated prosody.
Answering content questions	Fluent spontaneous production of sentences	The participant and clinician take turns asking open-ended questions about information provided in the text and answering with 1-2 sentences using the words used in the paragraphs they read.

Figure 1



Caption: Efficiency measure (correct information units/minutes) averaged across the four tasks for each language prior to and following treatment in Dutch (L1)

Figure 2 :



Caption: Frequency of code-mixing in languages of lower proficiency prior to and following treatment in Dutch (L1)

Table 1 – Language History

	Dutch	German	French	English	Italian	Norwegian	Spanish
Age learned	Birth	Early	Early	15	16	30s	40s
Manner learned	Acquired	Acquired	Learned/ Acquired	Learned/ Acquired	Immersed	Immersed	Immersed
Use pre CVA	Frequently	Regularly	Frequently	Regularly	Regularly	Rarely	Regularly
Self-rating pre CVA	9/9	8/9	7/9	7/9	7/9	4/9	6/9
Self-rating post CVA	8/9	7/9	7/9	6/9	7/9	3/9	3/9

Table 2 – Testing Schedule

Pre-Treatment

Day1	Day2	Day3	Day4	Day5
English	German	Italian	Dutch	French
Italian	Spanish	Norwegian	Norwegian	English
Spanish	Norwegian	Dutch	English	Spanish
German	Dutch	German	Italian	
	French	French		

Post-Treatment

Day1	Day2	Day3	Day4	Day5
German	Norwegian	Italian	Dutch	Norwegian
Dutch	Italian	English	Norwegian	German
Spanish	German	Spanish	English	Italian
English	Spanish	Dutch	French	French
		French		

Table 3 - Results for CIUs/min

Measures by Language	Pre 1	Pre 2	Pre 3	Post 1	Post 2	Post 3	Pre- to Post-mean	NAP	NAP-rescaled	Effect Size post-pre/SD_pre
Dutch										
Narrative	23	24	17	22	25	42	21.64 SD 3.7 to 29.35 SD 10.9	0.78	0.56	2.08
BAT Pic Desc*	**42**	**18**	**24**	**28**	**47**	**57**	**28.21 SD 12.6 to 43.94 SD 14.6**	**0.89**	**0.78**	**1.25**
Answering Q*	**28**	**26**	**39**	**69**	**47**	**35**	**31.37 SD 7.1 to 49.92 SD 17.2**	**0.89**	**0.78**	**2.62**
Q Generation**	**27**	**20**	**36**	**57**	**48**	**68**	**27.74 SD 7.7 to 57.67 SD 10.1**	**1.00**	**1.00**	**3.88**
English										
Narrative**	**25**	**n/a**	**13**	**34**	**30**	**41**	**18.95 SD 8.6† to 35.13 SD 5.4**	**1.00**	**1.00**	**1.88**
BAT Pic Desc**	**15**	**25**	**13**	**50**	**39**	**50**	**17.44 SD 6.7 to 46.28 SD 6.5**	**1.00**	**1.00**	**4.31**
Answering Q	30	24	25	27	27	61	26.35 SD 3.3 to 38.28 SD 19.9	0.78	0.56	3.67
Q Generation	38	61	29	40	38	72	42.7 SD 16.4 to 49.98 SD 19.1	0.78	0.56	0.45
French										
Narrative**	**13**	**17**	**21**	**35**	**26**	**48**	**16.81 SD 3.9 to 36.41 SD 11.0**	**1.00**	**1.00**	**4.99**
BAT Pic Desc**	**33**	**29**	**16**	**66**	**61**	**69**	**26.28 SD 9.1 to 65.17 SD 4.0**	**1.00**	**1.00**	**4.30**
Answering Q**	**33**	**39**	**22**	**41**	**48**	**56**	**31.19 SD 8.5 to 48.24 SD 7.6**	**1.00**	**1.00**	**2.01**
Q Generation**	**19**	**52**	**42**	**55**	**63**	**90**	**37.66 SD 17.2 to 69.29 SD 18.3**	**1.00**	**1.00**	**1.84**
German										
Narrative	21	11	11	18	25	11	14.44 SD 6.0 to 18.10 SD 6.7	0.78	0.56	0.61
BAT Pic Desc*	**18**	**22**	**35**	**32**	**44**	**35**	**25.23 SD 8.8 to 37.28 SD 6.2**	**0.89**	**0.78**	**1.37**
Answering Q	22	38	45	36	28	36	34.73 SD 11.8 to 33.39 SD 5.1	0.33	-0.33	-0.11
Q Generation*	**23**	**15**	**21**	**22**	**24**	**33**	**19.56 SD 4.5 to 26.16 SD 6.2**	**0.89**	**0.78**	**1.47**
Italian										
Narrative*	**11**	**12**	**18**	**22**	**14**	**30**	**13.67 SD 3.9 to 22.03 SD 7.8**	**0.89**	**0.78**	**2.16**
BAT Pic Desc*	**20**	**20**	**33**	**48**	**42**	**32**	**24.28 SD 7.3 to 40.42 SD 8.2**	**0.89**	**0.78**	**2.23**
Answering Q**	**24**	**15**	**25**	**30**	**40**	**36**	**21.61 SD 5.6 to 35.16 SD 5.0**	**1.00**	**1.00**	**2.42**
Q Generation	11	11	41	15	25	20	20.81 SD 17.4 to 20.04 SD 5.3	0.67	0.33	-0.04
Spanish										
Narrative	17	8	9	12	16	12	11.42 SD 5.3 to 13.30 SD 2.5	0.67	0.33	0.36
BAT Pic Desc	21	13	10	14	18	10	14.38 SD 5.4 to 13.85 SD 4.2	0.44	-0.11	-0.10
Answering Q	10	11	21	26	14	15	13.87 SD 6.5 to 18.63 SD 6.8	0.78	0.56	0.73

Q Generation	8	10	2	14	11	8	6.84 SD 4.0 to 10.73 SD 3.0	0.78	0.56	0.99
Norwegian										
Narrative	3	5	11	5	6	7	6.56 SD 4.2 to 5.78 SD 1.2	0.56	0.11	-0.19
BAT Pic Desc	8	5	5	10	8	8	6.32 SD 1.9 to 8.79 SD 1.0	0.78	0.56	1.33
Answering Q	11	10	7	14	5	7	9.05 SD 2.0 to 8.62 SD 4.5	0.44	-0.11	-0.21
Q Generation*	**5**	**6**	**13**	**15**	**16**	**32**	**8.05 SD 4.7 to 20.72 SD 9.4**	**1.00**	**1.00**	**2.73**

† represents two sessions rather than three.

NAP rescaled: *medium effects: .58–.84; **large or strong effects: .85–1.0
NAP: *medium effects: .79–.92; **large or strong effects: .93–1.0
Bolded items represent medium or strong effects for NAP

Table 4 - Results for Code-Mixed Words/Total Verbal Units

Measures by Language	Pre			Post			Pre- to Post-mean	NAP	NAP-rescaled	Effect Size (post-pre) /SD_pre
	1	2	3	1	2	3				
Spanish										
Narrative*	**1.19**	**0.00**	**5.21**	**3.88**	**12.77**	**7.32**	**2.13(2.73) to 7.99(4.48)**	**0.89**	**0.78**	**2.15**
BAT Pic Desc*	**4.17**	**7.89**	**1.35**	**6.90**	**8.11**	**10.14**	**4.47(3.28) to 8.38(1.64)**	**0.89**	**0.78**	**1.19**
Answering Q	0.00	6.52	1.05	3.30	1.98	7.92	2.52(3.50) to 4.40(3.12)	0.78	0.56	0.54
Q Generation	4.92	4.44	6.52	11.84	3.80	10.58	5.29(1.09) to 8.74(4.33)	0.67	0.33	3.16
Norwegian										
Narrative	17.65	25.35	11.69	15.38	15.05	31.09	18.23(6.85) to 20.51(9.17)	0.56	0.11	0.33
BAT Pic Desc	14.29	23.64	25.97	29.09	21.67	25.00	21.30(6.18) to 25.25(3.72)	0.67	0.33	0.64
Answering Q	17.65	23.08	18.00	18.75	13.38	32.77	19.57(3.04) to 21.63(10.01)	0.56	0.11	0.68
Q Generation	36.00	26.47	5.00	26.47	7.69	20.93	22.49(15.88) to 18.36(9.65)	0.39	-0.22	-0.26

NAP rescaled: *medium effects: .58–.84; **large or strong effects: .85–1.0
NAP: *medium effects: .79–.92; **large or strong effects: .93–1.0
Bolded items represent medium or strong effects for NAP