The neuroscientific basis of speech therapy for aphasic word-finding difficulties

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2022

Christopher W.T. Heath

School of Biological Sciences
Division of Neuroscience and Experimental Psychology

Contents

List of Figures 10

The neuroscientific basis of speech therapy for aphasic word-finding difficulties 12

Abstract 12

Declaration 14

Copyright statement 15

Acknowledgements 16

Chapter 1: Introduction 17

Chapter 2: Investigating the neural basis of post-stroke aphasia therapy; a systematic review of neuroscience-based therapy studies 19

Statement of Contribution 20

Abstract 21

1. Introduction 22

2. Methods 29

3. Results 37

3.1. Notable excluded studies 39

3.2. Assessment of Bias 41

3.3. Semantic and phonological therapy 46

3.4. Pantomime cueing 57

3.5. Melodic Intonation Therapy (MIT) 59

3.6. Audio-visual therapy 63

3.7. Intention Therapy 64

3.8. Behavioural treatment + neurostimulation 67

3.9. Targeting other levels of language processing 70

3.10. Imitation-based therapy 72

3.11. Functional and structural neural changes 74

3.12. Exploring relationships within and between studies and assessing robustness of the synthesis 75

4. Discussion 84

4.1. Limitations 90

4.2. Conclusions 94

Acknowledgements: 96

Chapter 3: Structural damage post-stroke and its relationship to therapy-related recovery in aphasia 97

Statement of Contribution 98

Abstract 99
Introduction 100
Methods 104
Pre-processing 107
T-tests and correlations 108
Tract-Based Spatial Statistics 110
Results 111
T-tests and correlations 111
Tract-Based Spatial Statistics 114
Discussion 116
Conclusions 120
Chapter 4: A comparison of different approaches to behavioural treatment of chronic post-stroke aphasia: methodological considerations 121
Statement of Contribution 122
Abstract 123
Introduction 125
Methods 132
Therapy approaches and treatments 132
Assessments 145
Methodology strengths and limitations 148
Disclosure of interest 150
Chapter 5: A comparison of different approaches to behavioural treatment of chronic post-stroke aphasia: Results and Discussion 151
Statement of Contribution 152
Abstract 153
Methods 154
Participants 154
Study Design 155
Results Section 1: Treatment Effectiveness 156
Q1.1: Do any treatments show a significant therapy-related change in naming of trained items? 156
Q1.2: Which, if any, treatments show a stronger therapy-related change in naming than others? 161
Q1.3: Were therapy effects maintained through to followup? 164

Q1.4: Are these therapy-related changes affected by the order the treatments are presented in (are there any order effects?) 166
Results Section 2: Inter-participant variation 167
2.1. Did inter-participant demographic factors modulate the effectiveness of treatment? 169
2.2. Did inter-participant language factors modulate the effectiveness of treatment? 169
2.3: Did inter-participant variations in lesion volume modulate the effectiveness of treatment? 172
Further investigation of Q2.1, 2.2 and 2.3 172
2.4. Did lesion location modulate the effectiveness of different treatments? 176
Results Section 3: Generalisation beyond treated items 181
Q3.1: Did significant therapy effects in naming accuracy generalise to untrained words?181
Q3.2 Did significant therapy effects in naming accuracy generalise to measures of semantic or phonological processing (e.g., changes in error patterns)?...... 183
3.3. Did significant therapy effects in naming accuracy generalise to wider measures of language and cognitive functions (e.g., comprehension, executive skills)? 185
3.4. Did significant therapy effects in naming accuracy generalise to functional language and communication measures? 195
Discussion 202
Treatment Effectiveness 202
Inter-participant variation 206
Generalisation beyond treated items 208
How our findings relate to each other 213
Disclosure of interest 216
Chapter 6: General Discussion 217
Summary of findings 217
Implications of research 222
Treatment Comparison 222
Generalisation 224
Semantic and Phonological ability change 226
Optimisation of treatment based on inter-participant variables 230
Conclusions 239
References 241

Total Word Count: 62,341

List of Tables

Figure 2.1. Overview of the citation selection process. Grey boxes indicate the number of citations remaining at each stage of analysis. The white boxes indicate the number of studies removed due to each criterion. 33
Table 2.1. A brief summary of the studies included in this review...................38
Table 2.2. Results of bias analysis using the Cochrane Collaboration's tool. L = Low bias risk, U = Uncertain bias risk, H = High bias risk. 44
Table 2.3. Adaptations made to Abel et al. (2015) framework. 45
Table 2.4 Methods of studies using semantic or phonological therapies........47
Table 2.5. Table of therapy effectiveness 75
Table 2.6. A summary of key changes of activation over the course of therapy in brain regions discussed as potentially significant by the authors, and the mechanisms which may underlie them (based on those described in Abel et al., 2015)..81
Table 2.6. A summary of key differences of activation compared to healthy controls in brain regions discussed as potentially significant by the authors, and the mechanisms which may underlie them (based on those described in Abel et al., 2015) 82
Table 2.7. A summary of key shifts of laterality in brain regions discussed as potentially significant by the authors, and the mechanisms which may underlie them (based on those described in Abel et al., 2015) 83
Figure 2.2.The number of datasets with post-therapy changes of activation or structure relative to pre-therapy in each hemisphere, grouped by aphasia severity...84
Table 3.1. Characteristics of the 18 PWA included in analyses 105
Figure 3.1. Binarised white matter tract mask overlaid on an average brain – left Inferior Longitudinal Fasciculus. 108
Figure 3.2. Lesion binary mask for a single PWA overlaid over Harvard Oxford atlas. 109
Table 3.2. Notable correlations between connectivity scores and language measures. *Moderately positive correlation (r>0.4). **Strongly positive correlation (r>0.6). 114
Figure 3.3. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- most anterior view. Red and green show skeletonised white matter tracts- areas with a significant difference are displayed in red...115
Figure 3.4. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- middle view. Red and green show skeletonised white matter tracts- areas with a significant difference are displayed in red. 115
Figure 3.5. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- most posterior view. Red and green show skeletonised white matter tracts- areas with a significant difference are displayed in red...115

Figure 4.1. Gantt chart to show overall timeline of study for each participant week by week. 134
Figure 4.2. Study structure and therapies for PWA order groups 1, 2, and 3. 135
Table 4.1. Model-oriented approach prompt levels and examples for the target item "Cat"........ 137
Table 4.2. Alternative modality approach prompt levels and examples for the target item "Cat"........ 139
Table 4.3. Executive demand approach prompt levels and examples for the target item "Cat"........ 142
Figure 4.3. Progression of each treatment week by week. Decreasing prompt hierarchy: Beginning on the easiest prompt and progressing to the hardest. Increasing prompt hierarchy: Beginning on the hardest prompt and progressing to the easiest. Self-selected prompts: Participant-selected prompts and/or order of prompt exposure. See text for further details. 145
Figure 5.1. Participant scores for items trained using the Phonological treatment pre- and post- therapy. 156
Figure 5.2. Participant scores for items trained using the Semantic treatment pre- and post- therapy. 157
Figure 5.3. Participant scores for items trained using the Gesture treatment pre- and post- therapy. 157
Figure 5.4. Participant scores for items trained using the Prosody treatment pre- and post- therapy. 158
Figure 5.5. Participant scores for items trained using the Interfered treatment pre- and post- therapy 158
Figure 5.6. Participant scores for items trained using the Speeded treatment pre- and post- therapy. 159
Table 5.1. Raw scores for each participant (out of a total of 60) before and after (pre and post) each treatment (Pros = Prosody, Phon = Phonology, Sem = Semantic, Spe = Speeded, Int = Interfered, Ges = Gesture). 160
Figure 5.7. Improvement in participant trained item naming scores for each treatment. 163
Table 5.2. Interactions between Treatment and Timepoint effects for each 2x2 ANOVA treatment comparison. 164
Figure 5.8. Scores by PWA at the immediate and delayed posttherapy timepoints. 165
Figure 5.9. Immediate and delayed improvement in PWA scores for each treatment. 166
Table 5.3. Statistics for each CLMM used in Q1 and their likelihood ratio tests to the null model. 167
Figure 5.10. Alluvial plot showing broad degree of improvement in naming trained items by each participant across the different treatments. 168
Figure 5.11. Alluvial plot showing broad degree of improvement in naming trained items by each participant across the different treatments. 168
Table 5.4. Spearman's correlation coefficients between measures of language ability and improvement in different treatments. 171

Table 5.5. Statistics for each CLMM used in Q2 and their likelihood ratio tests with the null model. CLMMs are split into those focused on demographic factors (red), those focused on language factors (green) and those focused on the lesion-related factor (blue), with each branch increasing in complexity to identify the model of best fit. Bold rows show models with lowest AIC value. 174
Table 5.6. Coefficient statistics for each inter-participant factor considered in the penultimate CLMM (including all inter-participant factors). 175
Figure 5.11. VBCM analysis showing correlates to improvement in the Phonological treatment. Clusters significant at reduced threshold are marked by arrow (p-FWE-corr: Left image: .015, Right image: .042). 177
Figure 5.12. VBCM analysis showing correlates to improvement in the Prosody treatment. Significant clusters are marked by arrows (p-FWE-corr: Left image:.006, Right image: .001). 178
Figure 5.13. VBCM analysis showing correlates to improvement in the Speeded treatment. Cluster significant at reduced threshold is is marked by arrow (p-FWW-corr = 0.039). 179
Figure 5.14. Results of VBCM analysis showing neural correlates to improvement in naming for Phonological (Green), Speeded (Blue) and Prosody (Red) treatments. Image thresholded at p < .005. 180
Figure 5.15. Violin plot showing change in naming score (out of a total maximum of 60) over each 3- week approach. 182
Figure 5.16. Change in Dell p score before and after treatment using each approach. 184
Figure 5.17. Change in Dell s score before and after treatment using each approach 184
Figure 5.18. A graph to show delayed improvement for PWA T score averages for each comprehension and each expressive language section in the CAT (maximum possible T score: 75) 187
Figure 5.19. A graph to show delayed improvement for PWA T score averages for subtests relating to expression and comprehension at the word and sentence level on the CAT (maximum possible T score: 75) 187
Figure 5.20. A graph to show delayed improvement for PWA T scores on individual subtests in the CAT (maximum possible T score: 75) 188
Figure 5.21. A graph to show delayed improvement for PWA T scores on comprehension in the CAT (maximum possible T score: 75) 189
Figure 5.22. A graph to show delayed improvement for PWA T scores on individual auditory comprehension subtests in the CAT (maximum possible T score: 75) 189
Figure 5.23. A graph to show delayed improvement for PWA T scores on expressive language in the CAT (maximum possible T score: 75) 190
Table 5.7. CAT T scores used in analyses- columns in bold underwent Wilcoxon tests 191
Table 5.8 CAT T scores used in analyses- columns in bold underwent ilcoxon tests 192

Table 5.9. CAT T scores used in analyses- columns in bold underwent Wilcoxon tests 193
Table 5.10. CAT T scores used in analyses- columns in bold underwent Wilcoxon tests 194
Figure 5.24. Delayed improvement to PWA Scenario Test scores (out of a maximum of 54) 197
Figure 5.25. PWA ANELT Meaningfulness and Intelligibility scores over the delayed improvement timeframe -pre-treatment to post-treatment (out of a maximum of 40) 198
Table 5.11. PWA scores pre- and post-treatment for the ANELT and Scenario Test 200
Table 5.12. Changes in measures of communication for the composite pictures from pre- to post-treatment. Dark green and red denote larger (>1) increases or decreases in each measurement, while light green and red denote smaller (>1) increases or decreases in each measurement. 201

List of Figures

Figure 2.1. Overview of the citation selection process.. 33
Figure 2.2.The number of datasets with post-therapy changes of activation or structure relative to pre-therapy in each hemisphere, grouped by aphasia severity. 84
Figure 3.1. Binarised white matter tract mask overlaid on an average brain – left Inferior Longitudinal Fasciculus. 108
Figure 3.2. Lesion binary mask for a single PWA overlaid over Harvard Oxford atlas. 109
Figure 3.3. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- most anterior view. 115
Figure 3.4. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- middle view. 115
Figure 3.5. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- most posterior view.. 115
Figure 4.1. Gantt chart to show overall timeline of study for each participant week by week. 134
Figure 4.2. Study structure and therapies for PWA order groups 1, 2, and 3 135
Figure 4.3. Progression of each treatment week by week. 145
Figure 5.1. Participant scores for items trained using the Phonological treatment pre- and post- therapy. 156
Figure 5.2. Participant scores for items trained using the Semantic treatment pre- and post- therapy. 157
Figure 5.3. Participant scores for items trained using the Gesture treatment pre- and post- therapy. 157
Figure 5.4. Participant scores for items trained using the Prosody treatment pre- and post- therapy. 158
Figure 5.5. Participant scores for items trained using the Interfered treatment pre- and post- therapy 158
Figure 5.6. Participant scores for items trained using the Speeded treatment pre- and post- therapy. 159
Figure 5.7. Improvement in participant trained item naming scores for each treatment. 163
Figure 5.8. Scores by PWA at the immediate and delayed posttherapy timepoints. 165
Figure 5.9. Immediate and delayed improvement in PWA scores for each treatment. 166
Figure 5.10. Alluvial plot showing broad degree of improvement in naming trained items by each participant across the different treatments. 168

Figure 5.11. Alluvial plot showing broad degree of improvement in naming trained items by each participant across the different treatments. 168
Figure 5.12. VBCM analysis showing correlates to improvement in the Prosody treatment.. 178
Figure 5.13. VBCM analysis showing correlates to improvement in the Speeded treatment.. 179
Figure 5.14. Results of VBCM analysis showing neural correlates to improvement in naming for Phonological, Speeded and Prosody treatments 180
Figure 5.15. Violin plot showing change in naming score (out of a total maximum of 60) over each 3- week approach. 182
Figure 5.16. Change in Dell p score before and after treatment using each approach. 184
Figure 5.17. Change in Dell s score before and after treatment using each approach 184
Figure 5.18. A graph to show delayed improvement for PWA T score averages for each comprehension and each expressive language section in the CAT 187
Figure 5.19. A graph to show delayed improvement for PWA T score averages for subtests relating to expression and comprehension at the word and sentence level on the CAT 187
Figure 5.20. A graph to show delayed improvement for PWA T scores on individual subtests in the CAT 188
Figure 5.21. A graph to show delayed improvement for PWA T scores on comprehension in the CAT 189
Figure 5.22. A graph to show delayed improvement for PWA T scores on individual auditory comprehension subtests in the CAT 189
Figure 5.23. A graph to show delayed improvement for PWA T scores on expressive language in the CAT 190
Figure 5.24. Delayed improvement to PWA Scenario Test scores 197
Figure 5.25. PWA ANELT Meaningfulness and Intelligibility scores over the delayed improvement timeframe -pre-treatment to post-treatment 198

The neuroscientific basis of speech therapy for aphasic word-finding difficulties

Christopher Heath
The University of Manchester
For the degree of Doctor of Philosophy, April 2022

Abstract

A wide range of different behavioural treatments for post-stroke aphasia have been developed and studied in depth over the course of modern aphasia research. Recent work involving neuroimaging has also developed our understanding of how these treatments function in relation to the functional and structural changes in the brain experienced by people with aphasia (PWA). However, understanding of the intricate relationships between different treatment effects and PWA lesion, language and demographic profile is not yet fully developed. This thesis focused on progressing our understanding in this area via investigation of a range of treatments and their effects across various PWA exhibiting differences across these three areas. **Chapter 2** sought to summarise existing research in this area via a systematic literature review. It discussed the research into a range of different neuroscientifically-based treatments for post-stroke aphasia that has occurred since 1997. In addition, it helped to highlight areas in this field which are currently more sparsely researched and could benefit from greater examination. These areas - namely, the relationships between white matter damage and PWA language profile and response to treatment, direct comparison between differing behavioural treatments, the effects of treatments on functional communication, and how and to what extent treatment effects are affected by inter-participant variation, were then explored in greater detail in subsequent chapters. **Chapter 3** investigated the relationships between language deficits and lesion location using analyses of both structural Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI) imaging in 18 participants. The results showed some support for a dual pathway model of language processing, as well as support for the roles of the fornix, frontal aslant tract, and corticospinal tract in language. Tract-Based Spatial Statistics analyses also found differences in integrity between controls and semantically improved PWA in several right hemisphere tracts. **Chapter 4** discussed the methodology developed in order to perform a standardised, direct comparison of three different approaches to aphasia treatment, divided into six individual treatments, for 18 PWA. This includes details of the overall structure and timepoints of the study, as well as the methodology for each treatment, and details of the assessments performed at each timepoint. **Chapter 5** covers the results and discussion from this comparison, including cumulative link mixed models (CLMM) analyses in addition to more standard analyses, in order to try to capture some of the nuances in the interactions between treatment effect and inter-

participant factors. The two executive function focused treatments, Speeded and Interfered naming, were found to outperform the others in terms of the primary outcome measure. While this success was modulated by inter-participant factors to some degree, they were generally the strongest treatments regardless of the inter-participant factors accounted for, although Interfered naming was found to be more resilient to the effects of these factors than Speeded naming. Some evidence supporting length of treatment as an important factor in generalisation to untrained naming was also found.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification or any other university or any other institute of learning.

Copyright statement

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Paul Conroy, Dr Stefanie Bruehl, and Dr Andrew Stewart for their continued support and guidance throughout my PhD, particularly over the course of the COVID lockdowns. I would also like to thanks my examiners; Dr Gorana Pobric and Professor Alex Leff, for taking the time to read and discuss this thesis.

I would also like to thank all the people who gave up their time to participate in this research, as well as their partners and family, many of whom helped facilitate this participation. I also need to thank the researchers at both NARU and PLORAS involved in the maintenance of the participant databases used, and in particular; Professor Cathy Price, Sophie Roberts, and Storm Andrews for all their help and support regarding the London-based participants and enthusiasm for collaboration.

Thanks also to everyone at NARU for their support, as well as the other DNEP PhD students who I have been able to share this experience with and who have offered their help and support whenever needed. In particular, Katherine Gore and Jayesha Chudasama have been there to support, encourage and help me throughout my PhD, both academically and emotionally. I must also thank the friends and family who gave me places to stay whenever necessary during data collection in London; Freya Caudwell and my uncle and aunt, David and Karen Heath.

More generally, I am incredibly grateful to the family and friends that have offered me support and encouragement over the last four years, helping both to pick me up at my lowest points and celebrate the high points with me. In particular, my granny, stepmum and sisters; Edith and Philippa Heath, and Hannah and Alicia Smith, and my friends; Sam Mason, James Ross, Georgia Longmoor, and Sam and Holly Openshaw.

Special thanks go to my friends Paul and Dr Natalie Hetherington and Alex and Jennifer Mackay, who have all supported me unreservedly over the last four years; given me places to stay during housing trouble or COVID lockdowns, given me academic, emotional, physical, and social support without me even needing to ask, and been there for me through some of the worst and best times of my life.

Finally, I also want to thank my sister Caity, who has been there with me through everything.

Chapter 1: Introduction

This thesis is divided into three main projects. Firstly, a systematic review of the existing literature was performed in order to highlight the current status of the research field of post-stroke aphasia therapy. This allowed us to identify both the strengths of the current field, and the areas which had not yet been fully explored, meaning that we could focus our remaining efforts on areas of the field which require development and would benefit from additional research.

One of the areas identified as currently not fully explored was studies incorporating diffusion-weighted imagery. The second project included in this thesis is therefore a study investigating post-stroke structural damage and its relationship to therapy-related recovery in the field. This study focuses primarily on analysis of diffusion-weighted scans and exploration of the effects of white matter damage on language and therapy-related recovery of that language.

The main takeaway from the systematic review, in terms of areas of sparsity in the current research field, was the lack of research performing direct comparisons between different treatments. While many studies investigated different behavioural treatments, these were usually compared to some form of control as opposed to another non-generic treatment, so the best treatment for a given situation is still unclear. The bulk of this thesis, therefore, is taken up by an empirical study performed to address this shortcoming in the current literature. This empirical study splits six different treatments into three approaches and compares their effects in the short and long term in 18 people with aphasia (PWA). In addition to this, it compares the effects of a number of inter-participant factors on treatment effects. As the field was also identified as not yet being at the stage where optimal treatments can be prescribed for individuals based on their pre-treatment language, demographic and lesion profile, further investigation of inter-participant factors' varying effects on treatment efficacy aimed to progress the field in this direction. Finally, investigation of generalisation of treatment effects to untrained items and functional communication was also identified as an area of sparsity in the existing literature. Therefore, a number of measures of improvement of untrained items and different forms of functional communication and overall language and cognition changes were also investigated in this study.

Because of the methodological complexity required for a study including so many independent and dependent factors, this final empirical study is broken into multiple chapters. The first explains the theoretical basis and methodology of the study. The second discusses the data analysis, results and discussion of study findings.

The final chapter is a discussion of the overall findings and strengths and weaknesses of the research included in the thesis. Future directions for research following this thesis are also discussed, as well as the key overall findings.

Chapter 2: Investigating the neural basis of post-stroke aphasia therapy; a systematic review of neuroscience-based therapy studies

Christopher Heath[a,*], Paul Conroy[a], Ilaria Pellegrini[b], Stefanie Bruehl[a,c,d,e]

[a]School of Biological Sciences, University of Manchester, UK

[b]The University of Siena, Siena, Italy

[c]St Mauritius Rehabilitation Centre, Meerbusch, Germany

[d]Clinical and Cognitive Neurosciences, Department of Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany

[e]Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine University, Duesseldorf, Germany.

*Corresponding authors:

Christopher Heath

School of Biological Sciences

University of Manchester

Oxford Road Manchester M13 9PL, UK

christopher.heath-2@postgrad.manchester.ac.uk

Statement of Contribution

Christopher Heath formulated search criteria, parameters and process. Christopher Heath and Ilaria Pellegrini performed independent inspection of search results in order to formulate final paper list. Christopher Heath performed paper analysis and writeup of paper. Stefanie Bruehl and Paul Conroy provided guidance and support throughout the length of the study and repeated feedback on various paper drafts.

Abstract

A variety of therapies for aphasia can be found in the current literature. However, the questions of which changes in the brain are most linked with improvement of language abilities, and how alterations in neural activation are affected by different approaches to therapy, require further exploration. This systematic review therefore aimed to answer the key research question: What are the underlying mechanisms forming the neural and cognitive basis for behavioural treatment of aphasia? Inclusion and exclusion criteria focused on finding studies utilising neuroimaging and language testing before and after neuroscientifically-based behavioural treatment for at least 5 participants with chronic post-stroke aphasia were identified using a 2-stage analysis. Citations were found via Google Scholar, PubMed, and Web of Science searches, The Cochrane Collaboration's tool was used to assess risk of bias. A narrative synthesis was used to compare heterogeneity within and between studies in each treatment area. . The resulting 28 studies included in the review covered therapies ranging in approach from targeting specific stages of language processing, to employing alternative modalities of communication, to facilitating activation of specific regions of the brain. Robustness of the narrative synthesis was limited in some ways by some studies with unclear or high risks of bias. Many studies found changes in both hemispheres following treatment, particularly those with datasets including mild deficits. Overall, this review shows that manifold changes in the brain may occur, stemming from therapy and improvement in language abilities, although which changes are most important in facilitating improvement for participants with different specific profiles of damage and language deficit remains unclear.

Keywords: Aphasia; Treatment; Neuroimaging; Stroke; Recovery; Therapy

1. Introduction

Aphasia, an acquired language disorder after stroke, typically has a hugely detrimental impact on quality of life and affects a significant number of stroke patients worldwide. Estimates have previously suggested it may affect as many as 250,000 people across the UK (Geranmayeh et al., 2014). The language deficits exhibited by people with aphasia (PWA) can significantly affect their day-to-day lives, to the extent that aphasia has been found to have a greater impact on health-related quality of life than potentially life-threatening diseases such as cancer (Lam & Wodchis, 2010). However, PWA often have the potential to regain some degree of their former language abilities even in the post-acute and chronic recovery stages (Brady et al., 2016; Breitenstein et al., 2017). A wide range of behavioural interventions by means of Speech and Language Therapy (SLT) and neurostimulation methods are available to allow realisation of this potential, and recent advances in imaging technology and our increasing understanding of aphasia and its effect on the brain may further enhance therapy outcomes.

This review will therefore systematically evaluate studies in which behavioural treatments are supported by neuroscientific evidence and employed neuroimaging techniques. Our focus will be on anomia therapy studies. Anomia is a symptom of aphasia in which PWA are unable to retrieve the words they want to say (Nardo et al., 2017). These word-finding deficits are an almost universal element of aphasia (Crinion & Leff, 2007) and they are particularly limiting to successful communication of PWA in everyday life (Nardo et al., 2017). Change in severity of anomia may also be used as an easily quantifiable measure of behavioural improvement, allowing studies to relate behavioural changes to neuroimaging results (Crinion & Leff, 2007). This study focuses on both functional and structural results obtained using Magnetic Resonance Imaging (MRI) techniques, as both functional Magnetic Resonance Imaging (fMRI) and Diffusion Tensor Imaging (DTI) sequences have frequently been used in post-stroke aphasia research. Conventional MRI is used to provide structural information (Sebastian & Breining, 2018; Smith et al., 2004; Zatorre et al., 2012) and is frequently used in combination with fMRI and DTI to provide basic structural information such as lesion size and location.

While recovery of language abilities is possible for many PWA given the right circumstances, it is still unclear which exact forms of therapy and neural activations would produce optimal recovery for any given PWA (Cocquyt et al., 2017; Hartwigsen & Saur, 2019; Kiran, 2012; Pierce et al., 2017). Individual differences in the demographic and lesion factors, as well as variations in specific deficits, mean that definitive prognostic predictions are presently very challenging and not likely to be accurate (Plowman et al., 2012; Tippett, 2015). Similarly, the benefits or drawbacks of activation in each hemisphere are also currently still unclear. Studies show greater right hemisphere activation in PWA compared to healthy controls (Cao et al., 1999; Weiller et al., 1995), and more divergence from normative models of activation in the right hemisphere (Sarasso et al., 2010). There is some evidence of beneficial right hemisphere compensation in studies involving language treatment (Blasi et al., 2002; Crosson et al., 2005; Leff et al., 2002; Musso et al., 1999; Raboyeau et al., 2008) and right hemisphere activation during language tasks in PWA (Calvert et al., 2000; Thulborn et al., 1999). However, right hemispheric activation has previously also been associated with a limiting effect on language improvement (Cao et al., 1999; Richter et al., 2008), with inhibition of right hemisphere activity often resulting in improved language (Martin et al., 2004; Naeser et al., 2005). Overactivation of the right hemisphere is theorised to possibly be a maladaptive adaptation strategy (Naeser et al., 2004; Rosen et al., 2000; Sarasso et al., 2010). Some theories instead suggest that increased right hemisphere lateralisation may be more beneficial (and therefore associated with the improvements in language ability demonstrated in the data) in PWA with more severe deficits (Abel et al., 2014; Crosson et al., 2007). This is supported by studies which show varying outcomes depending on lesion size and location (Perani et al., 2003; Vitali et al., 2007). Increased left hemisphere damage may result in greater compensation by contralateral regions in the right hemisphere, meaning facilitation of right hemisphere activation is more beneficial when there is more left hemisphere damage and more compensation is needed (Crosson et al., 2007; Heiss & Thiel, 2006). However, there is also evidence that in the chronic phase, some PWA with severe deficits demonstrate left-hemisphere lateralisation (van de Sandt-Koenderman et al., 2016).

The exact mechanism by which recovery of language occurs is also currently not clear, with multiple contrasting theories on the topic. Redundancy recovery suggests that similar (or even the same) functional representations are distributed across a larger area of the brain, allowing recovery via activation of one of these representations following damage (Zahn et al., 2006). Vicarious functioning suggests that recovery of the function occurs via adaptation of neurons not previously associated with the function (Zahn et al., 2006).

While treatment intensity plays a key role in language recovery (Bhogal et al., 2003; Cherney et al., 2008; Hartwigsen & Saur, 2019; Pulvermuller et al., 2001), different therapy approaches also utilise a variety of strategies in order to facilitate recovery of language production. Studies can therefore generally be placed into broad categories based on the type of approach they adopt. Therapies targeting processing of semantic information (conceptual information relating to a word) and/or phonological information (word form and structure) can often be identified by their use of cues and prompts for word retrieval which provide either semantic or phonological information on the target word, aiming to improve each underlying level of processing separately (Maher & Raymer, 2004). As word-finding difficulties are generally believed to be due to impairments in semantic and/or phonological processing in the left hemisphere (Howard & Gatehouse, 2006; Schwartz et al., 2006), semantic and phonologically-focused therapies are presumed to directly target these potentially impaired areas of processing. Gesture based therapy employs gestures (such as visual portrayals of an object using gesture) or pantomimes to elicit a verbal response (Kroenke et al., 2013; Pierce et al., 2017). Melodic Intonation Therapy (MIT) includes the use of prosody and rhythm in the PWA's vocalisations. The PWA and therapist 'sing' utterances together with rhythmic tapping, using pitch and intonational changes to emphasise the cadence of a given word or utterance (Norton et al., 2009; van de Sandt-Koenderman et al., 2016). Due to their differing methodologies, these therapies were considered separate categories in this review; gesture therapy treatment being notable for its use of the visual modality and non-verbal communication of information, MIT due to the rhythmic and melodic components. However, both of these two therapies aim to utilise unimpaired processes associated with language to help compensate for deficits in language production (Maher & Raymer, 2004). It is theorised that some of the regions behind these processes are located in the right hemisphere (Cocquyt et al., 2017; Gili et al., 2017). Intention therapy is also described as involving the right hemisphere as well, as it aims to directly recruit right hemisphere resources. Methodologically, its distinguishing feature is the use of a complex left-hand action before each language task to facilitate this shift (Crosson et al., 2005; Crosson et al., 2009). Audio-visual therapy (described as speech entrainment) aims to improve language production by allowing PWA to mimic the speech of others. Methodologically, this involves emphasising participants' ability to mimic speech by combining the auditory information with a clear view of the speaker's mouth (often using headphones and visual information presented via a screen) (Fridriksson et al., 2012). A final category of interventions is behavioural therapy applied in combination with neurostimulation. This method aims to facilitate or inhibit activation of different regions of the brain using electrical or electromagnetic stimulation, increasing therapy effectiveness (Hara et al., 2017; Hara et al., 2015), and can be distinguished methodologically by their use of neurostimulation techniques such as repetitive transcranial magnetic stimulation (rTMS) or transcranial direct current stimulation (tDCS) in combination with behavioural therapy.

Previous reviews have discussed factors such as involvement of the right hemisphere in recovery (Cocquyt et al., 2017; Kiran, 2012; Thompson & den Ouden, 2008), or investigated the effects of constraint-induced versus multimodal therapies (Pierce et al., 2017), the general effectiveness of SLT (Brady et al., 2016), aphasia recovery and therapy effects using functional neuroimaging (Crinion & Leff, 2007, 2015), or the effects of therapy on conversational ability (Carragher et al., 2012). However, this is the first systematic review to compare a wide range of interventions with a clear neuroscientific basis, presenting both their behavioural effects and their brain functional/ structural effects. The distinct contribution of the current review is to cover studies that target certain cognitive processing levels in an existing model, language abilities, or pathways/ regions of the brain based on our current understanding of the neural basis of aphasia and how to most effectively facilitate recovery. This allows for a discussion of the neural changes associated with improved behavioural outcomes for individuals with given characteristics in the context of the therapy provided, which is a necessary discussion as we learn more about the effectiveness of different therapies for different individuals (Abel et al., 2014; Anglade et al., 2014; van Hees et al., 2014).

Abel et al. (2015) provided a framework seeking to provide a summary of presumed brain mechanisms which may be responsible for differences in activation between PWA and controls, or activation changes in PWA as a result of therapy. It considers the direction of brain responses to be important for interpretation of the data: presence of more versus less activation versus controls (enhancement versus reduction) as well as presence of more activation after versus before therapy (activation increase versus decrease) along with the locus of brain responses are supposed to allow an identification of the underlying mechanisms which have been reported to date. For this reason, we set out to relate the regional activation changes and possible underlying mechanisms attributed by papers in the review to the categories of Abel et al. (2015). The framework provides potential mechanisms for four major categories of activation difference: Reduced activation, both in relation to controls and post-therapy in comparison to pre-therapy, and increased activation, again both in relation to controls and post-therapy in comparison to pre-therapy. The mechanisms for increased activation focus on increased demands and rewiring, while mechanisms involved in the reduction of activity focus either on left hemisphere malfunction, or alternatively on differences in task performance or higher efficiency. The rationale for this review is to provide an updated overview of the current field of research into behavioural therapies for aphasia, and specifically to investigate the mechanisms explaining the neural and cognitive changes driven by these therapies. Therefore, Population, Intervention, Comparison, Outcomes and Study (PICOS) criteria to define study eligibility are as follows: Population can be defined as adults win the chronic phase of aphasia secondary to stroke/ CVA. Intervention is a speech therapy method that was constructed with the goal of targeting certain cognitive processing levels in an existing model, language abilities, or pathways/ regions of the brain based on our current understanding of the neural basis of aphasia and how to most effectively facilitate recovery with a behavioural (non-neurostimulation or pharmacological) component in the intervention used, excluding pharmacological interventions and described in enough detail as to allow replication. No specific Comparison to control or other treatments was required. Outcomes included language measures as well as neural activation patterns and underlying neural mechanisms. Study design was required to include five or more participants, performance of neuroimaging both before and after behavioural intervention, and consideration of possible language changes before and after treatment. These criteria inform the key research question of: What are the underlying mechanisms forming the neural and cognitive basis for behavioural treatment of aphasia? Furthermore, a secondary research question was included: how does the framework provided in Abel et al. (2015) relate to the key activation differences reported in neuroscience-based therapy data?

2. Methods

Aphasia has a number of potential causes (e.g. neurodegeneration, brain tumour), and the prognosis and deficit profile for a specific PWA can vary dramatically depending on the cause of their aphasia (Jefferies & Lambon Ralph, 2006). However, to limit this variability, this review only considered studies testing PWA in the chronic phase of aphasia post-stroke. In the chronic phase, spontaneous recovery of language function is usually minimal (Berthier & Pulvermuller, 2011), so treatment effects can be easier to detect and measure. Similarly, the review was also limited to studies employing behavioural testing and neuroimaging both before and after treatment, as only these studies can directly explore language and neural changes as a result of the intervention. Studies with fewer than five total participants were also excluded from the review, in an attempt to focus on studies with more robust findings and greater generalisability to the field at large.

A systematic literature review was conducted by searching three different databases in January 2020. Web of Science and Google Scholar were selected primarily due to their size and breadth of research topics included. PubMed was additionally selected for inclusion of a more focused database, and to allow the use of MeSH terms in searching. It was judged that these databases would cover the vast majority of relevant articles on this topic, and were consistent with previous research in a similar area (Cocquyt et al., 2017), so it was not necessary to include any additional databases. A search was carried out on Web of Science for papers including the keywords "aphasia" AND "therapy", AND either "neuroimaging" OR one of the several terms for structural and functional neuroimaging methods commonly used in aphasia research (“MRI”, “fMRI”, “Diffusion Tensor Imaging”). “Diffusion Tensor Imaging” was not abbreviated to “DTI” as, unlike MRI and fMRI, using the unabbreviated name increased the number of search results. A series of searches for papers with the same terms present in their titles was carried out on Google Scholar, and a PubMed search for the MeSH terms “aphasia”, “neuroimaging”, and either “therapy” or “treatment outcome” was also performed (the use of MeSH terms in this database removed the need to use any specific neuroimaging methodology terms). Searches were also restricted to papers which had been published in the last 21 years (since 1997), in order to restrict results to the most currently relevant research, and papers published in English.

Secondly, abstracts of resulting publications were independently inspected by two reviewers and the following criteria for inclusion of a publication were applied: (1) featuring adults with aphasia secondary to stroke/ cerebrovascular accident (CVA), (2) describing application of an intervention (behavioural as well as other treatment methods – neurostimulation, pharmacological etc.), (3) including quantitative comparisons of language measures before in comparison to after treatment, and (4) utilising a widely-recognised neuroimaging method. At this stage, any studies which contained mixed aetiologies of language disorders, e.g., dementia rather than, or as well as, stroke, were excluded from the search. These criteria were focused on objective information likely to be present in the abstract which could be used to reduce the list of results to those which were potentially relevant. After discussion between reviewers, criteria 3 was adjusted for additional clarity to (3) including consideration of possible language changes before and after treatment.

Thirdly, resulting studies which met the abstract inclusion and exclusion criteria were then further checked for relevance by analysis of the full texts. The process of study selection is detailed in Figure 2.1. The additional criteria required for inclusion at this stage were: (5) a total of five or more participants, (6) inclusion of participants in the chronic stage of aphasia (at least four months post stroke), (7) performance of neuroimaging both before and after behavioural intervention, (8) involvement of speech therapy with a behavioural (non-neurostimulation or pharmacological) component in the intervention used, (9) use of a therapy method that was constructed with the goal of targeting certain cognitive processing levels in an existing model, language abilities, or pathways/ regions of the brain based on our current understanding of the neural basis of aphasia and how to most effectively facilitate recovery, and (10) a description of the treatment method with sufficient detail as to allow replication. Additionally (11), any study in which a pharmacological intervention was used was excluded. These criteria were chosen to remove irrelevant studies based on information which might not have been available in their abstracts. While there is some variation in when the chronic phase is considered to begin (Anglade et al., 2014; Hartwigsen & Saur, 2019; Kiran, 2012), we opted for a more lenient time-point (4 months post-stroke) in order to avoid excluding any papers of potential relevance.

For the narrative synthesis (Popay et al., 2006), the studies were grouped by intervention type. A single reviewer collected the relevant data for use in the review from each study meeting the inclusion and exclusion criteria. This included information on intervention length and characteristics, number and characteristics of participants, neuroimaging methods, location and direction of activation differences compared to controls and/or changes over the course of the study (as measured by the specific form of neuroimaging data included in the study), the mechanisms underlying these changes (as described by authors), measures of language ability used, and participant performance on these measures (specifically exploring comparative scores from pre- to post-treatment). Study eligibility for each synthesis subgroup was based on intervention characteristics – these were summarised in Table 2.1 and categorised into distinct approaches for synthesis. These subgroups were then used as the basis for exploration of heterogeneity among study results, namely exploring how heterogeneity in intervention related to heterogeneity in mechanisms underpinning language improvement and subsequent activation changes. As synthesis was narrative and based on outcomes included in the inclusion criteria, no specific methods were required to prepare the data for presentation.

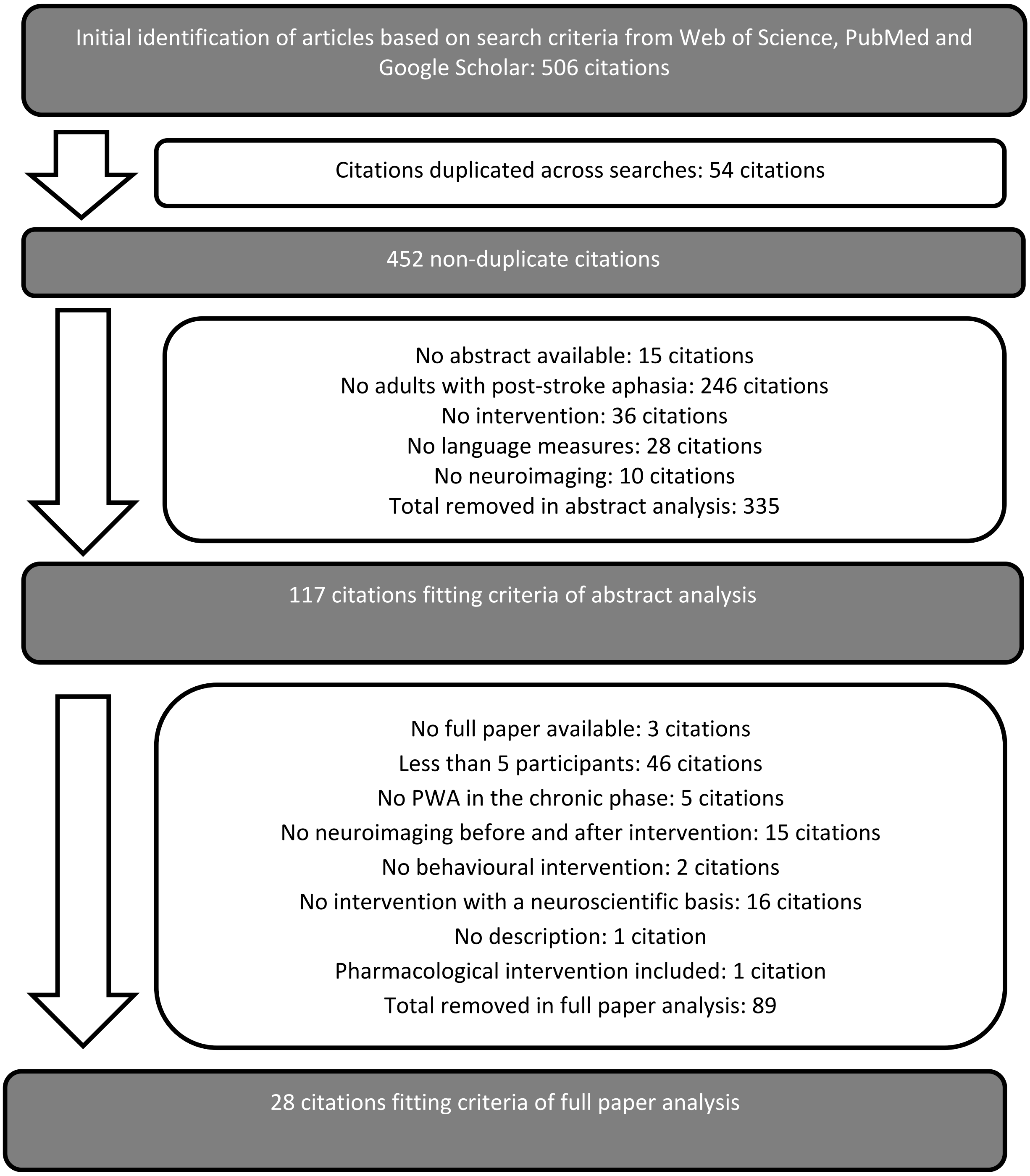

Figure 2.1. Overview of the citation selection process. Grey boxes indicate the number of citations remaining at each stage of analysis. The white boxes indicate the number of studies removed due to each criterion.

The studies fitting all criteria which were included in the review and their key details are listed in Table 2.1. The key differences of activation and laterality shifts were placed into the framework presented in Abel et al. (2015). Regional changes or laterality shifts were taken to be those which were interpreted as key by the authors, i.e., which were explicitly mentioned in the abstract, or, if no regions were considered in the abstract, those emphasized in the discussion. Similarly, the categorisation of the changes induced by potential mechanism was based on the authors' interpretations or judged based on the kind of activity change and related behavioural changes if no interpretation was given. For example, if decreased activation in a left hemisphere region was interpreted as a negative factor in recovery by the authors, this activation change would be included in the framework as being the result of persistent left hemisphere malfunctioning rather than increased efficiency (which would be associated with a decrease in activation positively linked to therapy gains). Some reported changes are therefore attributed to several mechanism categories, as authors tended to present multiple possible reasons for a given activation change. The framework was then adapted to remove redundant categories and otherwise summarise the neuroimaging results effectively. The categories of activation change differ from those in Abel et al. (2015) in the following ways: Missing inputs in the left hemisphere, as well as some mechanisms of enhanced demands and rewiring (storage of knowledge in language areas and recognition and familiarity) were removed due to not being discussed as a possibility in any of these studies. For post-therapy enhanced demands and rewiring, storage of knowledge in language areas and increased recognition and familiarity were grouped into a single category due to studies frequently using broader interpretations of new versus existing strategies for language processing (in language and non-language areas). Higher efficiency due to priming and due to learning were also combined into a single category due to frequent broader interpretations of generally higher efficiency. The original framework and the adaptions made to it are shown in Table 2.3. Tables 2.5 and 2.6 show how these results are summarised using the adapted framework for each region of the brain.

A number of papers fell narrowly outside the scope of this review, such as those involving Intensive Language Action Therapy (ILAT) and related therapies (Breier et al., 2006; McKinnon et al., 2017; Meinzer et al., 2008; Meinzer et al., 2009; Mohr et al., 2016; Nenert et al., 2017; Pulvermuller et al., 2005; Richter et al., 2008). Other notable studies were excluded due to neuroimaging occurring at only a single time-point (Bonilha et al., 2016; Campana et al., 2015; Chen et al., 2015; Fridriksson et al., 2011; Kroenke et al., 2013), the sample size falling below the 5-participant cut-off (Al-Janabi et al., 2014; Breier et al., 2010; Crosson et al., 2005; Fridriksson et al., 2007), the use of pharmacological intervention (Berthier et al., 2017), or the use of neurostimulation treatment only, with no behavioural treatment component (Szaflarski et al., 2011).

The outcomes data sought for were primarily treatment methodology used in each study and changes in any measures of language profile from pre- to post- treatment. Any data concerning patterns of neural activation, lateralisation or structure from pre- to post-treatment were also sought. Finally, to explore how severity of damage affects the degree of lateralisation of activity following therapy and how changes in activation are distributed across the brain during therapy in PWA, data concerning the levels of impairment for the PWA included in each study were also sought and compared with the key locations of structural or functional changes post-therapy. As there are some instances in this review of multiple papers being produced from the same original dataset, datasets rather than papers were used to avoid duplication of these results for this comparison. Ratings of impairment were taken either from author categorisation or individual Western Aphasia Battery Aphasia Quotient (WAB-AQ) scores (Kertesz, 1982), and datasets without these ratings or a group mean score only were left unconsidered (n=8). This left a total of 12 datasets for consideration in this comparison. Since all datasets studied more than one category of impairment (mild, moderate, severe), datasets span across categories. For example, as Abel et al. (2015) include PWA with at least moderate impairment, this dataset is included in both the 'both hemispheres, moderate impairment' and 'both hemispheres, severe impairment' categories. This comparison can be found in Figure 2.2. Risk of study bias, as well as risk of study bias due to missing results, was also assessed by a single reviewer using the Cochrane Collaboration's tool (Higgins et al., 2011). A summary of this assessment is shown in Table 2.3.

3. Results

The final number of studies meeting all criteria for inclusion in the review was 28. Scoring agreement between the two reviewers was good (89% agreement on which studies passed the abstract review stage after adjustment of criteria 3, 92% agreement on which studies passed the full paper review stage). Table 2.1 contains the list of studies in alphabetical order, according to the first author, together with some key descriptors of each study. The most obvious and practical way of grouping these studies is by type of intervention used for language rehabilitation. While the largest group of studies by type of intervention is those using semantic and/or phonological therapies, with nine studies, this group includes several different therapeutic methods. Otherwise, the included studies are spread across a variety of behavioural methods and modalities, as well as neurostimulation in combination with behavioural therapy. While our search terms covered aphasia as a whole, the included studies focused to a large extent on anomia and deficits of spoken language production. Despite defining the chronic phase as at least 4 months post-stroke, all of the papers included featured participants at least 6 months post-stroke.

Table 2.1. A brief summary of the studies included in this review.

Study	No. of PWA	No. of healthy controls	Therapy methods	Intervention length in weeks (w) and sessions (s)	Neuroimaging methods (performed both pre and post-therapy)	Time post-stroke (in months)
Abel et al. (2014)	14	0	Semantic/ Phonological	2 w (10 s) each	fMRI	11-72
Abel et al. (2015)	14	14	Semantic/ Phonological	2 w (10 s) each	fMRI	11-72
Barbieri et al. (2019)	19	0	Treatment of Underlying Forms	12 w (24 s)	fMRI	13-104
Benjamin et al. (2014)	14	0	Intention therapy	3 w (30 s)	fMRI	10-112
Cherney, Erickson and Small (2010)	8	0	Epidural cortical stimulation and behavioural therapy	6 w (42 s)	fMRI	12-194
Crosson et al. (2009)	5	5	Intention therapy	6 w (30 s)	fMRI	8-83
Duncan and Small (2018)	12		IMITATE	6 w (108 s)	fMRI	7-124
Fridriksson (2010)	26	0	Semantic/ Phonological	2 w (10 s)	fMRI	At least 8
Fridriksson et al. (2012)	13	20	Audio-visual speech entrainment	6 w (42 s)	fMRI	10-261
Gili et al. (2017)	10	0	Action/ pantomime observation therapy	6 w (30 s) each	fMRI	At least 6
Hara et al. (2015)	50	0	rTMS and behavioural therapy	1.5 w (10 s)	SPECT	Average 55
Hara et al. (2017)	8	0	rTMS and behavioural therapy	1.5 w (10 s)	fNIRS	11-106
Johnson et al. (2019)	26	17	Semantic	12 w (24 s) or until ≥90% accuracy	fMRI	10-170
Kiran et al. (2015)	8	8	Semantic	Until ≥80% accuracy over 2 consecutive w	fMRI	15-157
Laganaro et al. (2008)	4	15	Computer Assisted Therapy	3-5 w (3-5 s)	EEG/ERP	1-4
Leonard et al. (2015)	5	0	Phonological	10 w (30 s)	fMRI	12-36
Marangolo et al. (2016)	9	0	tDCS and behavioural therapy	3 w (15 s)	fMRI	7-96
Marcotte et al. (2012)	9	0	Semantic	Max 6 w (18 s)	fMRI	50-300

Study	No. of PWA	No. of healthy controls	Therapy methods	Intervention length in weeks (w) and sessions (s)	Neuroimaging methods (performed both pre and post-therapy)	Time post-stroke (in months)
Marcotte et al. (2013)	9	10	Semantic	Max 6 w (18 s)	fMRI	50-300
Menke et al. (2009)	8	9	Phonological	2 w (14 s)	fMRI	12-72
Peck et al. (2004)	3	3	Intention therapy	8 w (30 s in 3 2-w phases)	fMRI	8-74
Santhanam, Duncan and Small (2018)	12	>1	IMITATE	6 w (108 s)	fMRI	7-130
Schlaug, Marchina and Norton (2009)	6	0	MIT	75 s	DTI	At least 12
van de Sandt-Koenderman et al. (2016)	9	0	MIT	6 w (30 s)	fMRI	0.5-3 (subacute group), 17-40 (chronic group)
van Hees et al. (2014)	8	12	Semantic/ Phonological	2 w (6 s) each	fMRI	17-170
van Hees et al (2014)	8	12	Semantic/ Phonological	2 w (6 s) each	fMRI	17-170
Wan et al. (2014)	11	0	MIT	15 w (75 s)	DTI	7-110
Yang et al. (2019)	6	0	MIT-C	16 w (32 s)	DTI	10-28

3.1. Notable excluded studies

A number of notable studies investigating the neural bases of aphasia recovery had to be excluded from the results of this review in order to maintain focus on the specific scope of research the review concerns. For example, several studies investigating ILAT and related therapies were considered at the full paper analysis stage of citation selection. These included Breier et al. (2006), McKinnon et al. (2017), Mohr et al. (2016), Nenert et al. (2017), Pulvermuller et al. (2005), Meinzer et al. (2009), Meinzer et al. (2008), and Richter et al. (2008). ILAT involves intensive therapy over a short period, with a focus on spoken communication based on everyday conversation (Mohr et al., 2014). It can be considered to feature a neuroscience-based approach at a broad explanatory level, as it implements neuroscientific principles like coincidence learning and massed practice, behavioural relevance, and focus on remaining language abilities to counteract learned non-use (Pulvermüller & Berthier, 2008). However, as ILAT studies do not target a specific aspect of language processing, modality or region of the brain, they may be considered to be less directly based in neuroscientific theory than other approaches discussed, and do not fit the specific criteria for inclusion in this review.

In addition to ILAT, other therapeutic approaches not considered in this review are nevertheless worth noting. Speeded cueing therapy aims to improve connected speech by increasing the speed of participant response via methods like deadline naming (Conroy et al., 2018). Naming to deadline requires participants to produce the target word before a beep in a picture naming task. The deadline is shortened as participant speed increases (Conroy et al., 2018; Hodgson & Lambon Ralph, 2008). Interfered-naming therapy is based on the finding that, in picture naming, different types of distractors to a target word (for example, words that share a semantic category or phonological similarity with the target word) can facilitate or inhibit successful target word production when presented at different times relative to the picture (Abel et al., 2009). Certain distractors presented at the right time as part of a naming therapy can therefore be used to facilitate recovery of linguistic-executive abilities (Abel & Willmes, 2016). These alternative approaches to cueing therapy are still novel and have comparatively little research dedicated to them, but are so far promising options for neuroscience-based aphasia treatment.

Also excluded from this study were papers discussing aphasia as a consequence of other causes than stroke, such as neurodegeneration, and papers investigating pharmacological treatments of aphasia. Primary Progressive Aphasia (PPA), resulting from neurodegeneration, generally involves progressive worsening of the deficit, as well as frequent appearance and worsening of other cognitive deficits over its development (Gorno-Tempini et al., 2011). Aphasia resulting from neurodegeneration, as well as other causes, such as traumatic brain injury or tumour, may differ broadly in terms of deficit profile from post-stroke aphasia (Glosser & Deser, 1991; Patterson et al., 2006), and so have not been included in this review. For example, Davie et al. (2009) found a higher proportion of anomic aphasia relative to other subtypes than would be expected from stroke victims in a study on PWA with tumours, indicating a general difference in the severity of deficit that might be expected from the two different causes of aphasia.

3.2. Assessment of Bias

Risk of study bias, as well as risk of study bias due to missing results, also assessed by a single reviewer using the Cochrane Collaboration's tool (Higgins et al., 2011) (Table 2.3). The results of this assessment found that most studies in this review had a low bias risk on the majority of bias risk measures. Abel et al. (2014, 2015), Barbieri et al. (2019), Benjamin et al. (2014), Cherney, Erickson and Small (2010), Crosson et al. (2009), Duncan and Small (2018), Fridriksson et al. (2012), Hara et al. (2015, 2017), Johnson et al. (2019), Laganaro et al. (2008), Leonard et al. (2015), Marangolo et al. (2012), Marcotte et al. (2013), Menke et al. (2009), Peck et al. (2004), Santhanam, Duncan and Small (2018), Schlaug, Marchina and Norton (2009), and van de Sandt-Koenderman et al. (2016) were all assessed as having a low risk of bias in at least 4 of 6 measures. Fridriksson (2010), Gili et al. (2017), Kiran et al. (2015), and van Hees et al. (2014a, 2014b) were assessed as being more uncertain in their bias risk level, all showing uncertain bias risks in 3-4 measures. While these studies may have lower risk of bias, lack of clarity regarding certain aspects of their methodologies makes this hard to confirm. However, these studies were also assessed as not having high risk of bias on any measures, so overall risk of bias is not judged to be high. The studies in this review that appear to have the highest risk of bias are Wan et al. (2014) and Yang et al. (2019), which have a high risk of bias in 2 and 3 measures respectively. The findings of these studies should therefore be interpreted with the greatest degree of caution. Overall, random sequence allocation was the measure with the methodological aspect which introduced the most risk of bias for the studies in this review, with 5 studies being assessed as having a high risk of bias in this measure. Selection bias was the main source of risk of bias in the study. Several studies were at a high or unclear risk of bias due to lack of random sequence allocation, as specific allocation sequences were often not specified, and in some studies were definitively not used, for example, studies relying on a retrospective convenience sample, or which allowed participants to decide their treatment (Yang et al., 2019). Allocation concealment was also an unclear level of bias risk in many studies, as measures taken to conceal allocations were often not discussed. Blinding of participants and personnel had a similar lack of clarity in some cases, and in some, this risk was high, in studies where participants were clearly away of the available treatments and which they were being presented. Again, Yang et al. (2019) is a good example of this, where participants were given the choice of which treatment they wished to be given, so could not have been blinded in any way. Blinding of outcome assessment had a low risk of studies for most cases, as the vast majority of outcome assessments were either blinded or quantitative (e.g. Neuroimaging data, scores on established picture naming assessments, etc). Incomplete outcome data was an unclear risk in several studies in which things like completeness of outcome data or reasons for participant attrition were unclear. Finally, the majority of studies had a low risk of selective reporting , as all described outcome measures were generally reported.

Table 2.2. Results of bias analysis using the Cochrane Collaboration's tool. L = Low bias risk, U = Uncertain bias risk, H = High bias risk.

Source of Bias	Random Sequence Allocation	Allocation Concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective Reporting
Abel et al. (2014)	L	L	L	L	L	L
Abel et al. (2015)	U	U	L	L	L	L
Barbieri et al. (2019)	L	U	L	L	L	L
Benjamin et al. (2014)	L	U	L	U	L	U
Cherney, Erickson and Small (2010)	L	L	L	L	U	L
Crosson et al. (2009)	L	L	U	L	L	L
Duncan and Small (2018)	L	L	L	U	U	U
Fridriksson (2010)	U	U	U	L	L	L
Fridriksson et al. (2012)	L	U	L	L	U	L
Gili et al. (2017)	U	U	L	U	U	L
Hara et al. (2015)	L	L	L	L	L	L
Hara et al. (2017)	H	U	L	L	L	L
Johnson et al. (2019)	H	U	L	L	L	L
Kiran et al. (2015)	L	U	U	L	U	L
Laganaro et al. (2008)	L	L	L	L	U	L
Leonard et al. (2015)	U	U	L	L	L	L
Marangolo et al. (2016)	U	L	L	L	U	L
Marcotte et al. (2012)	L	L	L	L	U	L
Marcotte et al. (2013)	L	L	L	L	U	L
Menke et al. (2009)	L	U	U	L	L	L
Peck et al. (2004)	H	U	L	L	L	L
Santhanam, Duncan and Small (2018)	L	L	U	L	L	L
Schlaug, Marchina and Norton (2009)	L	L	L	L	L	U
van de Sandt-Koenderman et al. (2016)	L	L	L	L	L	U
van Hees et al. (2014a)	U	U	U	L	U	L
van Hees et al (2014b)	U	U	U	L	U	L
Wan et al. (2014)	H	U	H	L	U	L
Yang et al. (2019)	H	H	H	U	L	L

Table 2.3. Adaptations made to Abel et al. (2015) framework.

Abel et al. (2015) framework			
Brain signals before therapy (comparison to controls)		**Brain signal changes due to therapy (comparison pre-post or Time x Group interaction)**	
Enhancement of activation	**Reduction of activation**	**Increase of activation**	**Decrease of activation**
Enhanced demands and rewiring: - Storage of knowledge in language areas - New strategies in non-language areas - Maladaption in RH - Recognition and familiarity - More effort for executive control	Different task performance Malfunctioning in LH: - Local brain damage - Disconnection - Missing inputs	Increased demands and rewiring: - Storage in language areas - New strategies in non-language areas - Recognition and familiarity - More effort for executive control	Persistent malfunctioning in LH Higher efficiency: - Priming - Learning

Adapted framework			
Pre-therapy difference (compared to controls)		**Post-therapy activation difference (change over time)**	
Enhancement of activation	**Reduction of activation**	**Increase of activation**	**Decrease of activation**
Enhanced demands and rewiring: - New strategies in non-language areas - Maladaption in RH - More effort for executive control/lower efficiency	Different task performance Malfunctioning in LH: - Disconnection - Local brain damage	Increased demands and rewiring: - Storage in language areas/recognition and familiarity - New strategies for language processing - More effort for executive control	Persistent malfunctioning in LH Higher efficiency (priming or learning)

3.3. Semantic and phonological therapy

Phonological and/or semantic focused therapy, designed with the aim of targeting semantic and/or phonological processing to aid overall improvement in language production, were used in nine studies (Abel et al., 2014; Abel et al., 2015; Fridriksson, 2010; Kiran et al., 2015; Leonard et al., 2015; Marcotte et al., 2012; Marcotte et al., 2013; Menke et al., 2009; van Hees et al., 2014). As semantic and phonologically-focused therapies are presumed to directly target potentially impaired areas of semantic and phonological processing in the left hemisphere. It is theorised that language improvement as a result of these therapies is mostly driven by mechanisms in the left hemisphere, particularly higher efficiency due to priming or learning and storage of knowledge in language areas and increased familiarity.

One of the more common methodologies for targeting either primarily semantic or phonological aspects of language production, or both aspects, is via cueing hierarchies. In cueing hierarchies, further information about the target word (e.g. the fruit 'fig') is provided by means of cues. Semantically or phonologically-based information about word meaning/semantics (e.g. "a type of fruit") or word sound/phonology (e.g. "starts with /fi-/") can assist with naming immediately and longer-term. Cues can be provided in an increasing or decreasing hierarchy; for the former, cues are initially minimal, e.g. a first phoneme, and then become more informative if naming is not achieved, e.g. first-syllable or whole word repetition (Conroy et al., 2009). Alternatively, in the decreasing cue paradigm which is guided by errorless learning principles (Fillingham et al., 2003), whole word cues are used initially and then reduced if naming success can be maintained. These methods have the benefit of providing just enough information for the participant to retrieve the word for the former, or stopping/reverting to an easier level once they cannot do so for the latter (Abel et al., 2005). Other than cueing hierarchies, a common method of targeting the semantic part of language production is via Semantic Feature Analysis (SFA), which involves retrieval of a target word via encouraging production of semantically related words, with the aim of increasing activation of the semantic network around the word to facilitate retrieval (Coelho et al., 2000). Phonology can be targeted via a similar method in Phonological Components Analysis (PCA), where PWA must instead identify phonological aspects of a target word such as the first or last phoneme, or rhyming words (Leonard et al., 2015). The specific treatments targeting semantic, phonological or both aspects of language production are summarised in Table 2.4. It has been argued that most phonological or semantic therapies are not exclusively targeting phonological or semantic processing and overlap in terms of the cognitive processes they evoke, as some information of each form is almost always necessary to provide (Howard, 2000). For example, in SFA, the word form must ultimately be provided if the participant is unable to generate it themselves (Boyle & Coelho, 1995). However, these therapies clearly place emphasis on one form of processing, even if that form is not used exclusively.

Table 2.4 Methods of studies using semantic or phonological therapies

Study	Form of therapy		
	Semantic	Phonological	Graphemic
Abel et al. (2014)	CH	CH	
Abel et al. (2015)	CH	CH	
Fridriksson (2010)	CH	CH	
Johnson et al. (2019)	SF		
Kiran et al. (2015)	SF		
Leonard et al. (2015)		PCA	*
Marcotte et al. (2012)	SFA		
Marcotte et al. (2013)	SFA		
Menke et al. (2009)		CH	CH
van Hees et al. (2014)	SFA	PCA	

SFA= Semantic Features Analysis, PCA= Phonological Components Analysis, CH= Cueing Hierarchy, SF= Semantic Features based rehabilitation. *Some phonological components presented to PWA in written form

The effects of SFA were explored in a pair of studies using the same dataset (Marcotte et al., 2012; Marcotte et al., 2013), in which nine PWA underwent fMRI sessions while performing a picture naming task before and after up to six weeks of SFA. All PWA improved on language ability following therapy, with a mean improvement of 80% from a baseline of 0, as only incorrectly named words were selected for training. The first study of the pair (Marcotte et al., 2012) showed a range of individual responses to SFA, with a better outcome correlating with lack of left inferior frontal gyrus (IFG) damage, as well as with enhanced activation in the left precentral gyrus (PCG) before and after therapy. The left IFG is associated with both phonological encoding as part of the dorsal cortical language pathway and phonological processing (Burton et al., 2000; Schwartz et al., 2012) as well as semantic encoding and processing (Binder et al., 2009; Demb et al., 1995). As the PCG includes both the primary motor and premotor cortex, it has a role in many aspects of motor function and, with regard to language processing, is thought to be involved in articulatory planning and execution, as well as speech perception (Pulvermüller et al., 2006; Watkins & Paus, 2004; Watkins et al., 2003). Improvement was also positively correlated with pre-therapy naming and sentence comprehension scores. Participants who showed better treatment responses also generally showed smaller and/or fewer regions of activation post- compared to pre-therapy, while the worse responding PWA showed more areas of significant activation post-therapy, suggesting that better outcomes are marked by more constrained but effective neural activation. The second study (Marcotte et al., 2013) found differences in network integration between PWA and healthy controls. Modulation of the posterior Default Mode Network (DMN), a network associated with decreased activation during cognitively demanding tasks (Sridharan et al., 2008), was observed following SFA, showing that brain lesions and subsequent plasticity as a result of therapy affect large networks of regions such as the DMN, inducing complex changes in these networks in addition to their effects on specific areas. The authors suggest that improved integration of areas in the posterior DMN involved in language may be responsible for some of the modulation. The precuneus - in which modulation has been found after improvement in chronic aphasia (Fridriksson, 2010; Fridriksson et al., 2007) - and temporal areas involved in object naming (Warburton et al., 1999) and lexical selection (Indefrey & Levelt, 2004), were both considered as part of the posterior DMN.

One study within the nine semantic/phonological subset utilised a semantic feature based rehabilitation program which continued until PWA reached 80% accuracy in two consecutive weekly sessions (Kiran et al., 2015). This study used fMRI to assess activation in eight PWA and eight healthy controls, and supported the findings of Marcotte et al. (2012) regarding the importance of the left IFG and PCG. Observation of regional activation showed the left IFG was found to be the most consistently modulated region in terms of effective connectivity changes across patients following rehabilitation. Both regions were consistently more activated post- compared to pre-therapy, while the right IFG, bilateral superior and middle frontal gyri (SFG and MFG), right superior and middle temporal gyri (STG and MTG), as well as bilateral supramarginal gyrus (SMG) and angular gyrus (AG) were all consistently active as a function of rehabilitation in at least six patients during performance of either a trained picture naming task or a semantic feature verification task, or during both. Several of these regions are believed to play important roles in language function: Subregions of the MTG are thought to be involved in semantic processing (Binder et al., 2009; Binder et al., 1997), retrieval and memory (Xu et al., 2015), and lexical-semantic processing (Choi et al., 2015) and form part of a multimodal semantic network along with regions including the IFG (Jackson et al., 2016). The STG has been linked to high-order auditory processing (Burton et al., 2000; Visser & Lambon Ralph, 2011) and auditory-verbal semantic processes (Hocking & Price, 2009; Scott et al., 2000). Finally, the SMG is primarily associated with phonological processing (Gold et al., 2005; Stoeckel et al., 2009) and mapping between graphemic and phonological information (Booth et al., 2006). All PWA improved post-therapy, showing at least medium rehabilitation effect sizes. A comparison of trained with untrained items found both a significantly larger effect size and percentage change for the trained categories. The authors describe the treatment as emphasising both semantic feature analysis and phonological access, suggesting that this treatment may facilitate diverse processing stages, rather than semantic processing alone. Given that all PWA showed some improvement, and several of these areas were also active in the controls, the authors suggest these brain areas form a base required for normal language processing.

Johnson et al (2019) also used a semantic feature based treatment in which patients were required to name items and complete tasks involving identifying their semantic features. 26 patients completed the treatment (10 completed the untreated condition), which continued for 12 weeks or until 90% accuracy was achieved on 2 consecutive tests. This study also performed fMRI using an overt picture naming task pre and post therapy. Nine PWA did not respond to the treatment- however, those that did showed both activation patterns that were initially closer to those of controls, and a significant increase in activation over the course of treatment, compared to those who did not (who also showed little change in activation over treatment). Controls particularly differed from patients in higher bilateral AG and lower bilateral IFG pars triangularis activation. The authors suggest that an explanation for the varied behavioural responses for treatment could be that those with initial activation patterns more similar to controls had greater capacity to improve over treatment.

Another study within the semantic/phonological subset featured PCA as a therapy method. Leonard et al. (2015) compared two variations of PCA over 10 weeks. In these variations, if the PWA were unable to produce a phonological component spontaneously, they were provided with one by the experimenter. In one condition, two of the five PWA chose one option from a list (Choice condition); while in the other there was no choice element (No Choice condition). Analysis of individual performance showed that treated item naming was significantly improved at both the immediate and one-week follow ups post-treatment in all participants, demonstrating the effectiveness of both variations of PCA. Two PWA (one from each condition) also underwent fMRI scanning. A Partial Least Squares analysis was used on these scans to identify the latent variables (LVs) showing patterns of activation associated with semantic and phonological judgement tasks compared to baseline tasks. In both tasks, the significant activation patterns found in the PWA in the No Choice condition did not significantly differ when compared before and after treatment scans. However, for the PWA in the Choice condition, LVs were identified which seemed to indicate significant changes in activation following therapy. These LVs included both regions in which activation correlated positively with the phonological or semantic tasks and negatively with the baseline task, and vice versa. In the phonological task, these areas were bilateral prefrontal areas, thalamus, precuneus, and right middle occipital gyrus. In the semantic task, they were bilateral prefrontal regions, the left cingulate gyrus and MTG, and right insula. The phonological task was associated mostly with regions in both hemispheres in both patients, with the PWA in the Choice condition showing additional activation in several right hemisphere regions (MFG, medial frontal gyrus, thalamus, SFG and precuneus) post-treatment. While the semantic task was generally associated with less activation in right hemispheric areas than the phonological task, the Choice condition PWA also showed changes in activation post-treatment, this time in bilateral prefrontal regions, the left cingulate gyrus and MTG, and the right insula. Overall, changes in activation were present in the Choice condition PWA, but not the No Choice condition PWA, particularly in frontal regions of both hemispheres, and there was a larger treatment effect size for the Choice condition PWA.

PCA and SFA have also been compared in a study using resting state fMRI (rs-fMRI) by van Hees et al. (2014). This method of neuroimaging involves fMRI scanning without a specific task or stimulus being presented to or performed by the participant, and measures low-frequency fluctuations in activity, which have been found to be associated with functional networks of activity (Takamura & Hanakawa, 2017; van Hees et al., 2014). In van Hees et al. (2014), eight PWA underwent alternating weeks of PCA and SFA over a total of 12 sessions. Half showed significant improvements for SFA treated items; while seven of the eight showed significant improvement for PCA treated items. No significant improvements were found for untreated items. Interestingly, the PWA with primarily semantic deficits showed no significant improvement for SFA items, while PCA was effective both for those with primarily semantic and phonological deficits. Improvement in items treated with PCA was also correlated with a resting state measure of low-level fluctuations in activity in several regions, in both pre- and post- therapy scans: the right MTG pre-therapy; and left MTG and SMG, as well as pars triangularis of the right IFG, post-therapy. It is suggested by the authors that the involvement of the right IFG may have been due to a large number of left IFG lesions in the PWA involved in this study.

A later paper (van Hees et al., 2014) detailed the event-related fMRI findings of the same dataset, in which scanning took place during an overt picture naming task. They found pre-treatment activation in the left caudate to be associated with greater immediate improvement for the items treated using SFA. Greater improvement in items treated using PCA was correlated with increased left SMG activity post-treatment, and decreased activity was found in right hemisphere regions when naming treated items, compared to during incorrect naming of items pre-treatment. The authors suggest this may be due to treatment resulting in either increased right hemisphere efficiency or reduced maladaptive activity. Increases in activation in left hemisphere regions in five PWA in the same circumstances was also suggested to indicate possible increased recruitment of left hemisphere language-focused areas.

Several recent studies have used semantic and phonological cueing hierarchies as the basis for their research. In their study on long-term effects of aphasia therapy, Menke et al. (2009) employed a computer-assisted anomia therapy with a hierarchy of decreasing information provided via phonological and graphemic (written) cues over two weeks with eight PWA. All participants improved on trained items, and average improvement was found to be significantly greater for trained than untrained items. fMRI scans using an object naming task found that improvement immediately post-treatment was best predicted by increased activation in the left hippocampus and the para-hippocampi bilaterally. The authors interpret this as supporting the importance of brain structures related to memory and their connection with language regions in successful recovery. Long-term (8 months post-treatment) improvement was primarily related to activity increase in right middle and superior temporal areas. It is suggested that this may be due to large and varied lesions in the left hemisphere, limiting its potential for recovery and increasing the importance of activation in homologous right hemispheric language regions.

A more recent study by Fridriksson (2010) used both semantic and phonological hierarchies in 26 PWA who underwent an intensive, two-week therapy period with 30 hours of training split between the two hierarchies. There was a significant increase in naming performance overall between the first and last two treatment sessions. While semantic and phonological hierarchies were used on separate sets of items, there was little exploration of the difference in their effects beyond the fact that there was a strong correlation between improvement on the sets employing each. fMRI results showed activation in several areas of the left hemisphere (particularly the MFG, PCG, IFG pars opercularis, precuneus, and superior and inferior parietal lobule) during a picture naming task to have a positive relationship with naming scores, as well as a strong relationship between treatment-related activations in regions in the anterior and posterior parts of the left hemisphere. This is described as demonstrating the importance of recovery-related left hemisphere regions, in which significant damage or lack of functional change may result in limited recovery following treatment.

The differences between the effects of semantic and phonological hierarchies were addressed to a greater extent in Abel et al. (2014). While both the trained and untrained (control) item sets showed significant improvement from pre to post treatment, both semantic and phonologically trained item sets improved significantly more than the control set. Each hierarchy was trained for two weeks in 14 PWA. This study also found links between behavioural improvement and activation of regions in the left hemisphere during a picture naming task. fMRI activation peaks in the left IFG pars opercularis predicted greater improvement and decreases in activation in left superior temporal sulcus (STS), SMG, MTG and paracentral lobule showed associations with reduced benefits of therapy. However, comparison between PWA with predominantly semantic (S-patients) and predominantly phonological (P-patients) deficits demonstrated greater recruitment of right hemisphere regions pre-therapy by S-patients. Both treatments were found to be beneficial; as in van Hees et al. (2014), phonological treatment was especially beneficial for P-patients, while the same was not true for S-patients and semantic treatment, with semantic treatment being similarly effective for both groups. It is suggested that P-patients may focus more on preserved phonological functioning, benefiting more from therapy targeting their specific deficit, while S-patients involve more compensation from the right hemisphere, benefiting more from lexical treatment in general.

A subsequent study (Abel et al., 2015) using the same dataset compared associations between behavioural improvement, lesion location and changes in activation patterns. A joint independent-components analysis identified three components showing a significant group difference. The first showed a large lesion in the left IFG and surrounding areas to be associated with several changes in activation patterns negatively correlated with behavioural improvement. The authors suggest this is due to a large left IFG lesion resulting in disconnection of frontal and posterior areas via the language pathways. The other two components link lesions in the posterior

MTG (sparing the inferior temporal gyrus (ITG)) and the MFG/dorsal IFG pars opercularis with alterations in activation patterns for areas including right superior temporal lobe for the former and right IFG for both components. However, these components were not related to therapy gains.

The set of studies presented in Table 2.4, which utilised therapy focusing on semantic and phonological processing, have helped to progress our understanding of these approaches as well as the neural mechanisms underlying language recovery. They have demonstrated the effectiveness of several forms of semantically and phonologically based therapies, both individually (Kiran et al., 2015; Leonard et al., 2015; Marcotte et al., 2012; Marcotte et al., 2013; Menke et al., 2009) and in combination (Abel et al., 2014; Abel et al., 2015; Fridriksson, 2010; van Hees et al., 2014). Changes in activation in multiple regions of the brain have also been repeatedly linked to therapy-related behavioural improvement by studies in this area. These include the IFG, MTG, MFG, and SMG bilaterally and PCG in the left hemisphere (Abel et al., 2014; Abel et al., 2015; Fridriksson, 2010; Kiran et al., 2015; Leonard et al., 2015; Marcotte et al., 2012; van Hees et al., 2014). Compensatory activations in the right hemisphere have also been found by several of these studies (Abel et al., 2014; Abel et al., 2015; Leonard et al., 2015; Menke et al., 2009; van Hees et al., 2014). Overall, the findings presented here seem to suggest some differing patterns of activation between phonological and semantic therapies which can result in a difference in effectiveness depending on the specific deficits of the PWA being treated.

A preliminary synthesis of the changes in activation and mechanisms underpinning them (Table 2.5) suggests mixed support for this theory; the most common change in activation in these studies was an increase in left hemisphere activation, primarily due to storage of knowledge in language areas and increased familiarity. However, right hemisphere increases in activation were also observed, and a number of mechanisms were discussed by studies, suggesting that language improvements may be driven by a range of mechanisms in this therapy area. It must be noted that three of the studies using this approach to treatment had an uncertain risk of bias in multiple areas, so findings resulting from these studies must be treated with a degree of caution.

3.4. Pantomime cueing

One of the main methods of intervention using non-verbal methods utilises gestural cues, which may aid word retrieval (Pierce et al., 2017). There is debate over the nature and extent of interaction between gesture and language representation systems; for example, whether gesture and language interact during semantic conceptualisation and formulation, or during phonological encoding. As a therapeutic approach, gesture may therefore be used as a compensatory modality for communication, or a cue to aid word retrieval, and each theory has separate implications regarding the specific functions of gesture in communication (Akhavan et al., 2017). The mirror neuron system, a system of neurons critical in learning of actions and gestures from others, has also been found to be strongly linked to language-related communication (Chen et al., 2015). Pantomime cueing, using observation of pantomime actions to facilitate language production via theorised shared conceptual-semantic representations for both, is used in one study (Gili et al., 2017). As pantomime cueing places a focus on activation of non-language regions and sensorimotor regions, we theorise that the majority of activation changes occurring in this area are in the right hemisphere, underpinned by new strategies of language processing and potentially persistent left hemisphere malfunctioning.

To explore the relation between actions/ sensory-motor concepts and language production, Gili et al. (2017) tested the effects of language training using videos of everyday actions in 10 participants with non-fluent aphasia. Participants had to observe and describe the contents of the videos. Each participant took part in training using a) an action observation condition, in which a video reproducing an everyday real-life context with relevant objects (i.e. a train station with people dragging luggage, stamping tickets, entering and exiting the train, etc.) was presented, and b) a pantomime observation condition, with training using a video reproducing an everyday real-life context with no relevant objects visible; instead, actions were pantomimed (i.e. pantomimed actions and interactions at a clothing store). Each training condition lasted six weeks, with 7.5 hours of training each week. Both the action and pantomime observation conditions improved participants' language production abilities at the group level. There were significant increases in production of mean number of correct nouns, verbs, sentences, and phrases with high communicative value during the task from pre- to post-training. However, the mean number of correct nouns produced was significantly higher in the action observation condition. rs-fMRI scans only showed functional connectivity changes in right hemispheric sensory-motor networks in the action observation condition. It is argued that activation in these networks is important in processing action semantics, and so aids in aphasia recovery, and that this effect is greater in action observation, where the context and goals of the action are clearer than in pantomime observation. Preliminary synthesis therefore supports the idea that language improvement related to this treatment is driven primarily by changes in the right hemisphere as a result of the development of new strategies of language processing involving non-language regions. The majority of studies using this treatment had an overall low risk of bias, however, Gili et al. (2017) demonstrated an uncertain risk of bias in several areas, so this preliminary synthesis is appropriately tentative.

3.5. Melodic Intonation Therapy (MIT)

MIT, which involves use of singing-like word and phrase production to highlight the cadence of utterances, combined with rhythmic hand-tapping, aims to encourage right hemisphere activation via music functions. Three studies use MIT as their intervention (Schlaug et al., 2009; van de Sandt-Koenderman et al., 2016; Wan et al., 2014). MIT also places a strong emphasis on activation of non-language right hemisphere regions, meaning its facilitation of language improvement is also most likely driven by right hemisphere changes as a result of the development of new strategies of language processing involving non-language regions.

Task difficulty in MIT is gradually increased via decreased input from the speech therapist and production of longer phrases as the participant progresses (Schlaug et al., 2010; van de Sandt-Koenderman et al., 2016). While emphasising the rhythm of speech may have more benefit than melody for speech production (Stahl et al., 2013; Stahl et al., 2011), generalised language production improvement has also been associated with melodic therapy, suggesting that the combination of melody with rhythm is also important in the generalisation of benefits to untrained items in MIT (Zumbansen et al., 2014).

While the effects of MIT are theorised to be due to right hemisphere activation, its exact role and extent is unclear. Wan et al. (2014) used semistructured interviews to measure Correct Information Units (CIUs) per minute, providing a measure of speech fluency, in 11 patients with chronic nonfluent Broca's aphasia, as well as DTI scans, before and after 75 sessions of MIT. When comparing pre- and post-treatment fractional anisotropy (FA) in a search space of the right hemisphere regions underlying the cortex, a significant reduction in FA in several locations corresponding with regions in and around the right arcuate fasciculus (the white matter underlying the IFG, posterior STG, and posterior cingulum) was found in the treated group (indicating structural changes in the fibre tracts in these areas), but not in nine untreated control PWA. The treated group also showed significant post-therapy improvement in CIUs per minute not seen in the untreated group, which was negatively correlated with FA changes in the white matter underlying the right pars opercularis after correcting for IFG lesion load. The authors suggest that these changes in FA may indicate changes such as more axonal sprouting and less alignment of fibres (Sidaros et al., 2008), as well as more branching (Hoeft et al., 2007) close to the cortical target regions, and that structural changes could support therapy-induced behavioural effects like those found in the treated group.

Schlaug et al. (2009) selected six patients with moderate to severe nonfluent aphasia, also using 75 sessions of MIT treatment. Patients had DTI scans before and after training, with all showing a significant increase in number of fibres in the right arcuate fasciculus post-treatment. This study also used CIUs in several conditions as a measure of speech fluency, again finding a significant increase in CIUs per minute post therapy in all patients. These findings are seen as supporting a role for the right hemisphere in the production of spoken language, as well as the possible success of the melodic or rhythmic elements of MIT in engaging regions of the right hemisphere.

Treatment in combination with DTI scanning was also used by Yang et al (2018) in a study investigating the effects of MIT-C (Melodic Intonation Therapy- Chinese version) versus conventional therapy in Chinese speaking PWA. It must be noted that MIT-C differs from MIT due to the differences in language characteristics involved. Two groups of three participants underwent 32 1-1.5 hour sessions of MIT-C or conventional speech and language therapy respectively. Parametric statistical tests were not performed due to the small sample size, although greater improvements in meaningful word count were found in all three participants in the MIT-C group than any participant in the conventional therapy group. While this study also supports previous findings of right hemisphere FA changes post-therapy, specific changes are somewhat unclear. Two participants in the MIT-C group showed FA increases in tracts in both the dorsal and ventral streams, while the other only showed increased arcuate fasciculus FA. The changes in the right arcuate fasciculus noted in previous studies were also less clear, with two of the participants showing slight increases and the other showing a decrease in FA in this tract.

The link between the right hemisphere and language production recovery, however, is not a clear one. van de Sandt-Koenderman et al. (2016) performed fMRI before and after a six week period of MIT on nine patients with subacute or chronic non-fluent aphasia, comparing activation during a passive listening task and control task to gain a measure of activation related specifically to auditory comprehension, then comparing the activations in the left and right hemisphere to gain an overall lateralisation index (LI) of auditory comprehension. Subacute and chronic aphasia were defined in this study as less than three months post-stroke (participants ranged from 0.5-3 months post-stroke) and more than a year post-stroke (participants ranged from 17-40 months post-stroke) respectively. While all except one of the PWA showed significant improvement in at least one category of language production, different changes in LI were found in different individuals after MIT. Right-hemisphere activation was suggested to be more beneficial in the subacute period, as four of the five patients with subacute aphasia showed a rightward shift in LI. However, this was not the case for patients with chronic aphasia, for which only one patient showed a marginal rightward shift. No significant whole group correlations were found between magnitude of change in LI and improvement of language repetition or verbal communication. The possible importance of time post-stroke as a factor in treatment and recovery is therefore highlighted by the authors.

Preliminary synthesis shows support for the theory that MIT drives changes in the right hemisphere, developing language through new strategies involving non-language regions. Van de Sandt-Koenderman et al. (2016) is the only study which offers alternative findings, suggesting that while right-hemisphere activation was beneficial in the subacute period, only one patient showing a marginal rightward shift in LI in the chronic phase may indicate less involvement of mechanisms in the right hemisphere after the subacute phase. It must be noted that two of the studies of this treatment are at high risk of bias in multiple areas, so this preliminary synthesis must be interpreted with caution.

3.6. Audio-visual therapy

Audio-visual therapy (listening to a speaker while simultaneously deliberately watching their mouth articulate speech), which allows PWA to more effectively mimic speech, was used in one study (Fridriksson et al., 2012). Audio-visual speech perception has been found to result in increased speech production improvement compared to auditory-only speech perception in treatment for nonfluent PWA (Fridriksson et al., 2009), possibly due to increased activity in Broca's area (Fridriksson et al., 2008). Audio-visual therapy (described as speech entrainment) aims to take advantage of this disparity to provide more effective interventions for aphasia. Due to the suspected involvement of Broca's area in this treatment, changes in activation in left hemisphere language regions are expected due to storage of knowledge in language areas. The involvement of multiple simultaneous tasks may also result in activation differences due to the involvement of executive function processes.

Fridriksson et al. (2012) performed a series of experiments testing the effects of audio-visual entrainment on 13 participants with Broca's aphasia. An initial task showed a greater range of words were produced in audio-visual entrainment than during either audio-only entrainment or spontaneous speech. A subsequent fMRI neuroimaging experiment, in which activation during speech entrainment, spontaneous speech and speech perception were compared, featured the same participants and 20 controls. It suggested that entrainment causes greater activation in the ventral pathway, aiding semantic retrieval. A final experiment using audio-visual entrainment therapeutically during a six-week treatment found a significant improvement in word variety during spontaneous speech one week after the treatment phase (and a non-significant improvement six weeks after the treatment) at the group level.

Using the same fMRI comparisons as in the previous experiment, greater cortical activity for entrained compared to spontaneous speech was found in the posterior temporal cortex and anterior insula, which are connected by ventral fibres and suggested to form part of a ventral network involved in conceptual encoding of speech and some homeostatic functions during speech.

Preliminary synthesis shows that, while activation changes did occur in left hemisphere language regions, this was suspected to be due to higher efficiency due to priming or learning, as opposed to activation increases as a result of executive control involvement or storage of knowledge in language areas. The studies using this treatment had an overall low risk of bias, meaning we can have reasonable confidence in this preliminary synthesis.

3.7. Intention Therapy

Intention therapy aims to facilitate activation in the right hemisphere by combining language therapy with a complex left-hand movement in order to activate right hemisphere intention mechanisms associated with action selection. It was used as an intervention in three studies (Benjamin et al., 2014; Crosson et al., 2009; Peck et al., 2004). The complex left hand movement involves choosing a button to press, which is theorised to activate intention mechanisms in the right medial frontal cortex associated with choosing and initiating one action from a selection of competing options (Crosson et al., 2005). The aim is to cause subsequent activation in areas in the lateral frontal cortex such as the IFG and MFG that may play a role in language tasks like picture naming (Benjamin et al., 2014; Crosson et al., 2009). The aim of intention therapy is right hemisphere activation facilitation, so we would expect language improvement to be underpinned by development of new strategies of language processing in right hemisphere regions.

One early use of intention therapy in a neuroimaging study was by Peck et al. (2004), in which it was compared to a treatment designed to activate right hemisphere attention mechanisms. All three PWA improved on language measures (both naming of trained and untrained items and generation of category members) over the eight weeks of treatment, with a participant from each condition showing increased speed of haemodynamic response in the right hemisphere Broca's area homologue, motor cortex and pre-supplementary motor area (SMA). This response was measured using fMRI while participants completed a category-member generation task (naming examples of items in different given categories). The authors claim that this suggests training effects were generalised to non-trained word-finding tasks, and that a faster haemodynamic response indicates a faster response in these right hemisphere regions following therapy.

A subsequent study by Crosson et al. (2009) also studied changes in laterality following six weeks of intention therapy, as well as changes in activation and language, using fMRI with a category-member generation task. Four of the five PWA significantly improved on language measures, and these participants also showed a significant change in lateralisation towards the right hemisphere in lateral frontal areas. These PWA had similar lateralisation to controls pre-therapy. The one PWA who did not show significant behavioural improvement instead showed a lateral frontal change in lateralisation towards the left hemisphere. As she also had severe impairments, it is suggested that use of the intention treatment in recovery via rightwards lateralisation may be less beneficial for more impaired PWA.

Benjamin et al. (2014) attempted to address flaws in previous intention therapy studies by including a control treatment, which used similar methods to the intention treatment but without the complex hand movements, to determine if the lateralisation changes demonstrated in studies like Crosson et al. (2009) were due to the specific use of these movements. A larger sample size of 14 PWA was also used, with seven in each three-week treatment condition. While both groups showed significant improvement in language measures, six PWA in the intention therapy condition and only one PWA in the control condition demonstrated significant improvement on measures of naming members of an untrained category. fMRI scans (again using a category-member generation task) found a consistent rightward shift in lateralisation in the lateral frontal cortex for PWA in the intention condition, along with a medial frontal lateralisation shift in the longer term (three months post-treatment), which were not seen in controls.

Preliminary synthesis did show evidence of activation shift towards the right hemisphere as a result of intention therapy (Table 2.7), underpinned by new strategies for language processing. However, there was also evidence of a role of higher efficiency due to priming or learning, as well as some evidence of activation shifts towards the left hemisphere in some cases as a result of more effort for executive control and new strategies for language processing, meaning the mechanisms underpinning language improvement due to intention therapy may be more complex than theorised. The studies using this treatment had an overall low risk of bias, meaning we can have reasonable confidence in this preliminary synthesis.

3.8. Behavioural treatment + neurostimulation

Neurostimulation such as rTMS deployed in combination with language therapy, with the aim of facilitating or inhibiting certain brain regions and increasing therapy effectiveness, was reported in three studies (Cherney et al., 2010; Hara et al., 2017; Hara et al., 2015).

While neurostimulation methods can be used independently to improve language ability in PWA by altering the activation patterns associated with language (Naeser et al., 2011), they have more recently been used in combination with treatment in order to amplify the effects on language recovery (Cherney et al., 2010; Hara et al., 2017; Hara et al., 2015). Neurostimulation can be used to facilitate (or inhibit) activation in targeted areas to help ensure that language therapy primarily activates the most beneficial regions for recovery in a given patient. For example, facilitating right hemispheric activation may provide a better outcome for some patients in combination with therapies involving compensatory right hemisphere mechanisms (Cipollari et al., 2015). We would therefore expect a range of activation changes and underlying mechanisms depending on which areas are stimulated and how. For example, stimulation with the aim of increasing activation in language regions should be more likely to result in storage of knowledge in language areas or higher efficiency, while stimulation with the aim of increasing activation in non-language areas would be expected to result in activation changes caused by development of new strategies of language processing or more effort for executive control.

A frequent use of rTMS in aphasia rehabilitation is for facilitation of activation in the hemisphere judged to be primarily involved in language production post-stroke. Hara et al. (2015) used fMRI scanning in 50 PWA to determine the primary language compensation regions. Ten sessions of low-frequency (1Hz) rTMS were then used to inhibit activity in either the left or right hemisphere in combination with speech therapy. The aim was to facilitate activity in the primary language compensation regions via reduced inter-hemispheric inhibition. In their more recent study, Hara et al. (2017) applied low (1Hz) or high (10Hz) frequency rTMS to the right hemisphere and speech therapy over 10 sessions in four PWA each (eight in total) to respectively inhibit or facilitate its activity. The aim was again to facilitate activity in primary language compensation regions which were this time determined using functional near-infrared spectroscopy (fNIRS) during a word repetition task. In Hara et al. (2015), more patients with right hemisphere language compensation were classified as severely impaired. Pre- and post-therapy Single Photon Emission-Computed Tomography (SPECT) also found a significant association between a leftward laterality index (measured by comparing regional cerebral blood flow in the lesion and nonlesion sides of the brain) shift of the IFG pars opercularis and language improvement in the right hemisphere stimulation group. In Hara et al. (2017), the fNIRS results showed patients with right hemisphere language compensation had lower scores on the measure of language ability. As expected, the low-frequency stimulation and high-frequency stimulation groups showed reduced and increased right hemisphere activation, respectively. The high frequency group also had an increase in left hemisphere activity. Both studies found significant group language improvement regardless of location or frequency of rTMS.

While rTMS is a popular method of neurostimulation, others have also been used in combination with behavioural therapy. Cherney et al. (2010) used epidural cortical stimulation, in which electrodes are surgically positioned on the surface of the dura mater (a membrane surrounding the brain) (Williams et al., 2018). Cherney et al. (2010) gave eight PWA (four experimental and four control participants) six weeks of intensive daily therapy, including hierarchical linguistic cueing. The experimental participants also received cortical stimulation from an electrode array over the ventral precentral gyrus at 50Hz during therapy. Tasks performed during BOLD fMRI included both observation of and imitation of single syllable production and oral reading. Three out of the four experimental participants were found to have improved more than their matched control, although there was no significant group-level difference. Decreases in whole brain activation were found to correlate with increased improvement on behavioural measures, and more decrease in left hemisphere activation was found in the experimental than the control group. It is suggested that these broad reductions in brain activation are due to greater efficiency and less requirement for conscious effort in completion of these tasks, and are more consistent with patterns associated with long term learning, indicating that epidural stimulation may promote this type of change more than short-term improvement.

A third method of modulating neuronal activation is via tDCS, a non-invasive technique which can be used to increase or decrease cortical excitability (Cipollari et al., 2015). Marangolo et al. (2016) used bilateral tDCS in the IFG in combination with speech therapy for three weeks, to test if this would facilitate language recovery by maximising the activation of perilesional left hemisphere regions. Nine PWA were tested in both tDCS with therapy and sham tDCS with therapy conditions. Response accuracy for word repetition increased significantly in both conditions. However, the real stimulation condition also showed significant improvement in syllable repetition and had significantly greater word repetition accuracy than the sham condition following therapy. PWA also showed increased rs-fMRI functional connectivity in only right hemisphere regions following the sham condition, while in the tDCS condition they showed increased functional connectivity in the left premotor cortex, anterior cingulate cortex, MFG, and precuneus, as well as bilaterally in the cerebellum and in the right frontal cortex and SMA. The authors conclude that this indicates the success of bilateral tDCS in combination with speech therapy in activation of left hemisphere language networks, and that this greater involvement of left hemisphere networks resulted in more behavioural improvement.

As expected, a range of mechanisms were observed by authors in this treatment, depending on the specifics of the neurostimulation being used, including higher efficiency due to priming or learning, new strategies for language processing, and greater involvement of left hemisphere language regions. While treatments mostly targeted increased left hemisphere activation, in some cases there was also evidence of decreases in left hemisphere activation, for example, in the left STG in Cherney et al. (2010). Therefore, while these results broadly fit with those theorised, they also appear to demonstrate additional complexity that had not been accounted for. We can have a reasonable level of confidence in this preliminary synthesis as the studies using this treatment had an overall low risk of bias.

3.9. Targeting other levels of language processing

While the use of semantic and/or phonological information in treatment of aphasia is one of the more common forms of anomia therapy, the use of other types of language information, such as orthographic or graphemic (written language or letters), or lexical and syntactic information providing larger-scale information like word and sentence meaning and construction, is also being explored in several recent studies. It would be expected that these treatments would function mechanistically similarly to semantic and phonological treatments, as both place a focus on assistance via types of language information. For this reason, we would expect left hemisphere activation changes underpinned by higher efficiency or storage of knowledge in language areas in studies using these therapies.

One example of this, used as a treatment in Langanaro et al (2008), is Computer-assisted Therapy (CAT). In its written naming form, CAT involves written picture naming with both graphemic and auditory cues available to PWA. Four PWA underwent a picture naming task, as well as EEG recording during a delayed naming task, before and after therapy. A range of electrophysiological therapeutic responses were found, with some instances of increased similarity to those shown by the control participants, and some instances of divergence. This, along with the fact that only one of the PWA was in the chronic phase of recovery during treatment, makes drawing general conclusions for treatment in the chronic phase challenging. However, the presence of right hemisphere activation at different stages in the study provides some support for a role of the right hemisphere in recovery, with the authors commenting that the appearance and disappearance of right hemisphere activation may support the theory of a sequence of shifts of activation between hemispheres suggested by Saur et al. (2006).

Barbieri et al (2019) used Treatment of Underlying Forms (TUF), which places a focus on the lexical properties of verbs and syntactic mapping, training comprehension and production of passive sentences using language elements like syntax rather than the often used semantic and phonological focuses. 14 treated and 5 untreated PWA were tested on sentence production and comprehension weekly throughout treatment, and fMRI scanning during a sentence comprehension task took place before and after the intervention. Eye tracking during a sentence-picture matching task was also performed. Improved production of trained items in all but 1 treated participant was accompanied by significant generalisation to less complex untreated sentence structures. Neuroimaging results showed a majority of upregulation occurring in the right hemisphere, including in the SPL, AG, SMG, superior lateral occipital cortex (sLOC), postcentral gyrus, MFG and PCG.

While storage of information in language areas was one suspected mechanism underlying participant improvement in this area, unexpectedly, most of the observed activation changes occurred in the right hemisphere, and new strategies in non-language areas were also theorised to be a mechanism underpinning these changes. It is therefore possible that the use of orthographic and graphemic cues in this treatment places it mechanistically closer to treatments employing alternative modalities such as pantomime cueing and MIT. The studies using this treatment had an overall low risk of bias, meaning we can have reasonable confidence in this preliminary synthesis.

3.10. Imitation-based therapy

While imitation plays a role in many forms of speech and language therapy (Duffy, 1995), some approaches, such as IMITATE (Lee et al., 2010) use it as a basis for treatment. Similarly to gesture-based treatments, these use imitation of speech to allow input of the visual system in speech, and specifically to encourage activation of the mirror neuron system so that it may aid language recovery (Lee et al., 2010). These similarities to gesture-based treatments suggest any language improvement is likely due to similar mechanisms, and we would expect right hemisphere changes as a result of new strategies in non-language areas.

Duncan and Small (2018) used IMITATE, an approach focused on intensive, imitation-based treatment administered via computer program (Lee et al., 2010), with twelve PWA over a six week treatment timeline, with rs-fMRI performed at several timepoints. Analysis identified 8 different resting state networks (RSNs) in PWA rs-fMRI; however, the most notable finding of the study was that post-therapy increased production of CIUs during a narrative task was associated with increased time spent in a state of minimal correlation of activation between the different RSNs. The authors suggest this may indicate that increased segregation between different RSNs may benefit recovery, at least in imitation-based treatments, possibly due to increased connections within individual networks and ability to isolate the networks necessary for each specific task allowing more efficient processing.

Santhanam et al (2018) also performed six weeks of treatment on twelve PWA using IMITATE, this time using fMRI to observe activation during a speech observation task, and a repetition test for behavioural language assessment. They found that language improvement was associated with increasing similarity to pathway activations seen in controls, and was found in those with a more moderate amount of time since their stroke, rather than those whose strokes had occurred more than ten years ago, whose activation patterns generally decreased in similarity to controls. In particular, in the left hemisphere the pathways between the ventrolateral premotor cortex and primary sensorimotor cortices, and between the posterior superior temporal gyrus/sulcus and the inferior parietal lobule, became more similar to controls during therapy in the group showing improvement. Right hemisphere changes were not found to be associated with language improvement, and right hemispheric pathways were also found to be less plastic, demonstrating less changes in activation than left hemisphere pathways.

Contrary to expectations, predominantly left hemisphere changes were observed as a result of imitation-based therapy, although these changes were primarily focused on sensorimotor regions. Increased similarity to controls as a result of pathway activations also implies that these changes are not a result of new strategies of language processing and suggests this treatment may instead involve reinforcement of existing language mechanisms. We can have a reasonable level of confidence in this preliminary synthesis as the studies using this treatment had an overall low risk of bias.

3.11. Functional and structural neural changes

Tables 2.5, 2.6, and 2.7 summarise the main functional changes in activation levels and their interpretation described in those papers which discuss regional activation changes. Tables 2.5 and 2.6 show changes in activity by region, both compared to controls and over time, and the mechanisms that may be underlying them, illustrating the underlying activation changes in the brain, how frequently they occur, and which are most crucial to recovery. Table 2.7 shows lateralisation changes in different regions, i.e., shifts in the ratio of activation in left hemisphere language regions and activation in their right hemisphere homologues. These tables show that adjustments were required in order to effectively relate the key differences and changes in activation in this review to, the framework in Abel et al. (2015) . As shown in Table 2.5, many different regions were found to reveal changes in activation post-therapy. However, post-therapy changes in activation in certain regions, such as the IFG, MFG and MTG bilaterally, seem to be particularly common. Similarly, Table 2.7 shows that studies investigating changes in region lateralisation post-therapy also have mixed results, with evidence of changes in lateralisation in both directions being possibly linked with beneficial mechanisms for language improvement.

Figure 2.2 compares the levels of impairment for the PWA included in each study with the key locations of structural or functional changes post-therapy. It appears to show that, at least at a group level, changes frequently occur in both hemispheres rather than only the left or right hemisphere following therapy. Only PWA with moderate or severe deficits were found to undergo changes in one hemisphere only, while all datasets including PWA with mild deficits included changes in activation or structure in both hemispheres.

3.12. Exploring relationships within and between studies and assessing robustness of the synthesis

Therapy effectiveness is shown in Table 2.5.

Table 2.5. Table of therapy effectiveness

Study	Therapy methods	Primary language ability measure	Experimental group pre-post score change	Required number of sessions
Abel et al. (2014)	Semantic/ Phonological	Picture naming	-2-20% of maximally attainable score (p<.001)	10
Abel et al. (2015)	Semantic/ Phonological	Picture naming	-2-20% of maximally attainable score (p<.001)	10
Barbieri et al. (2019)	Treatment of Underlying Forms	Sentence comprehension and production probes	Comprehension: Average 28.5% (p<.001) Production: Averge 70.7% (p<.001)	24
Benjamin et al. (2014)	Intention therapy	Picture naming	Approx. 21%, t = 4.44, p<.005	30
Cherney, Erickson and Small (2010)	Epidural cortical stimulation and behavioural therapy	WAB-AQ	12.3 points	42
Crosson et al. (2009)	Intention therapy	Picture naming	Effect size: -0.48- 6.96	30
Duncan and Small (2018)	IMITATE	Correct Information Units during a narrative task	p=.002	108
Fridriksson (2010)	Semantic/ Phonological	Picture naming	t=4.12, p<.001	10
Fridriksson et al. (2012)	Audio-visual speech entrainment	Percentage correctly produced words from a script	t=2.76, p=.009	42
Gili et al. (2017)	Action/ pantomime observation therapy	Mean number of nouns, verbs, sentences and C-units produced	F=24.07, p<.001	30
Hara et al. (2015)	rTMS and behavioural therapy	SLTA	5.9 in RH-LF-rTMS, 6.6 in LH-LF-r-TMS (p<.01).	10

Study	Therapy methods	Primary language ability measure	Experimental group pre-post score change	Required number of sessions
Hara et al. (2017)	rTMS and behavioural therapy	SLTA	8.5 in LFS (p<.01), 13.5 in HFS (p<.05)	10
Johnson et al. (2019)	Semantic	WAB/WAB-R AQ	3.1 points	24
Kiran et al. (2015)	Semantic	Picture naming	20-49%, F=17.3, p<.001	Variable
Laganaro et al. (2008)	Computer Assisted Therapy	Delayed picture naming	minimum McNemar's Chi-square = 20.632, p < .0001	3-5
Leonard et al. (2015)	Phonological	Picture naming	Effect size: 1.53-3.94	30
Marangolo et al. (2016)	tDCS and behavioural therapy	Syllable/word repetition	F=100.7, p<.001	15
Marcotte et al. (2012)	Semantic	Picture naming	80%, p=.008	18
Marcotte et al. (2013)	Semantic	Picture naming	80%, p=.008	18
Menke et al. (2009)	Phonological	Picture naming	64.4%	14
Peck et al. (2004)	Intention therapy	Time delay between auditory input and verbal response	1.77 seconds	30
Santhanam, Duncan and Small (2018)	IMITATE	WAB-R AQ	Not given	108
Schlaug, Marchina and Norton (2009)	MIT	CIU score	Approx. 350-1750, p<.05	75
van de Sandt-Koenderman et al. (2016)	MIT	AAT, ANELT	AAT: 17.14, ANELT: 6.88	30
van Hees et al. (2014a)	Semantic/ Phonological	Picture naming	Significant improvement (p<.05) for 7 of 8 PWA in PCA, 4 of 8 in SFA	6
van Hees et al (2014b)	Semantic/ Phonological	Picture naming	Significant improvement (p<.05) for 7 of 8 PWA in PCA, 4 of 8 in SFA	6
Wan et al. (2014)	MIT	CIUs/min	T=6.34, p<.001	75
Yang et al. (2019)	MIT-C	Number of meaningful words produced	57.5 words	32

As is evident from Table 2.5, the majority of studies in this review found a significant improvement in language measures in PWA following treatment. However, the number of sessions used to reach this significant improvement varied greatly between studies, suggesting a large variety in terms of time efficiency for both PWA and those administering the treatment. Abel et al. (2014, 2015), van Hees et al. (2014a, 2014b) and Laganaro et al (2008) all produced significant improvement in PWA performance on language measures using a very small number of treatment sessions (10 or less), demonstrating high efficiency in terms of time required to facilitate language improvement. Conversely, Wan et al (2014), Schlaug, Marchina and Norton (2009), and Duncan and Small (2018) all produced statistically significant improvements in PWA language, but used a much larger number of sessions (75 or more) to achieve this improvement, suggesting that they may have had lower time efficiency for both the PWA and administrators of the treatment.

Preliminary syntheses of each treatment area show a range of activation patterns and underlying mechanisms across the different treatments. While our theoretical understanding of the mechanisms by which each category of treatment facilitated language improvement predicted heterogeneity in mechanism and activations between treatment approaches, much of the heterogeneity observed did not match our initial expectations.

We expected semantic and phonological therapy and treatments targeting other levels of language processing to function relatively similarly mechanistically, with a focus on mechanisms reinforcing existing left hemisphere language systems. Audio-visual therapy was expected to utilise similar mechanisms to these, with potential additional focus on executive control. We also expected pantomime cueing, MIT, intention therapy and imitation-based therapy to function relatively similarly, with more right hemisphere activation and focus on new strategies in non-language areas. Neurostimulation in combination with behavioural treatment was expected to differ from study to study depending on the specifics of the neurostimulation employed.

Contrary to our expectations, studies targeting other levels of language processing aligned more closely with the more right-hemisphere focused treatments than semantic and phonological therapy, with more right hemisphere activations and new strategies in non-language areas than semantic and phonological therapies. It is possible that the involvement of orthographic and graphemic cues created a greater focus on areas not associated with speech, despite the focus on levels of language processing. As expected, audio-visual therapy was similar to semantic and phonological therapy (and differed from the other therapies) in terms of a focus on left hemisphere changes in activation. However, it differed mechanistically, being underpinned by higher efficiency due to priming or learning. It is possible that observation of audio-visual speech functioned as a primer for language production rather than as a cue for executive control development.

Studies within the semantic and phonological therapy subgroup were also more heterogenous than expected, with variation from the expected focus on left hemisphere language areas. This may be due to inter-participant variation, as all the studies utilised a similar focus on semantic and phonological information. However, PWA recruited in the studies covered a fairly large range in terms of time post-stroke and pre-treatment language and lesion profile, which may have resulted in differing responses to treatment.

Pantomime cueing, MIT and Intention therapy were all relatively similar in having treatment responses predominantly driven by activation in non-language right hemisphere regions. However, there was some inter-study heterogeneity, with van de Sandt-Koenderman et al. (2016) suggesting that right-hemisphere activation may have only been beneficial in the subacute period. This was likely due to being one of the only studies including PWA in the subacute period, and therefore one of the only studies able to comment on relative efficacy of right hemisphere activation in the subacute versus chronic periods. It was also the only MIT study exhibiting significant left hemisphere changes, and so able to compare efficacy of left and right hemisphere changes. This is possibly also due to the use of an overall laterality index and use of fMRI, as opposed to observation of white matter tracts via DTI which was more common in the other studies. Intention therapy also differed in terms of indication of a role of higher efficiency due to priming or learning, as well as some evidence of activation shifts towards the left hemisphere in some cases as a result of more effort for executive control and new strategies for language processing. It is possible that the hand movements involved in intention therapy require enough executive control that this additional effort shifted activation towards executive control regions in the left hemisphere.

Behavioural treatment and neurostimulation demonstrated a range of mechanisms and inter-study heterogeneity, depending on the specifics of the neurostimulation being used. Mechanisms included higher efficiency due to priming or learning, new strategies for language processing, and greater involvement of left hemisphere language regions. Similarly, a lot of heterogeneity within this subgroup was observed in terms of changes in activation in the left versus right hemisphere, which again was expected due to the heterogeneity in terms of the methodology and aims of the neurostimulation methods used.

Overall, risk of bias in most subgroups was low. However, the semantic and phonological therapy and MIT subgroups in particular contained multiple studies with several aspects of high or uncertain risk of bias, so risk of bias within these subgroups is likely to be higher, impacting the robustness of these aspects of the narrative synthesis. These studies were not specifically given less weight as a result of their limited trustworthiness, as for MIT, the two high risk studies had similar findings in terms of right hemisphere differences to Schlaug et al. (2009), the only other MIT study utilising DTI.

Similarly, the majority of studies of both low and uncertain risk demonstrated within-study heterogeneity in terms of activations and mechanisms underlying treatment effect. For this reason, giving less weight to the studies of uncertain risk narratively was deemed not to be necessary. However, this lack of acknowledgement of studies of limited trustworthiness may have affected the overall robustness of the synthesis. Another factor which can affect narrative synthesis robustness is uncertainty over reviewers having enough information to judge whether individual studies meet inclusion criteria. However, this review included multiple specific, objective inclusion criteria, limiting confusion over whether or not to include individual studies, as evidenced by high inter-rater agreement rates in terms of which studies to include in each stage of selection for this review. Overall, therefore, this synthesis is likely to exhibit decent robustness, despite certain aspects potentially introducing (or failing to limit) risk of bias

Table 2.6. A summary of key changes of activation over the course of therapy in brain regions discussed as potentially significant by the authors, and the mechanisms which may underlie them (based on those described in Abel et al., 2015)

	Post-therapy activation increase (change over time)			**Post-therapy activation decrease (change over time)**	
Region	**Storage of knowledge in language areas/increased familiarity**	**New strategies of language processing**	**More effort for executive control**	**Persistent LH malfunctioning**	**Higher efficiency due to priming or learning**
Left ACC				2	
Left AG	5,7,10	5,10		1	6
Left Caudate				1,2	
Left FG	11		11		
Left Hippocampus	11				
Left IFG	5,7,8	5		1	7
Left MFG	1,8	1	1		7
Left MTG	7,8,11			1	
Left PHG	11				
Left PCG	5,8	5		1	7
Left PoG				1	
Left Precuneus	5	5		1	2
Left SFG	1,8	1	1		
Left SMG	5,7,10	5,10,12		1	6
Left SPL	5	5			
Left STG	11			1	4
Left thalamus				1	
Right AG	3,7	3			
Right caudate				1	
Right CG	1,11	1	1,11		
Right FG	11		11		
Right IFG	7,8,9	7,8,9			
Right sLOC	3	3			
Right MFG	3,8	7,8			
Right MTG	8	7			2
Right PHG	11	11			
Right PCG	3	3,7			
Right Precuneus	1,11,12	1	1,11		2
Right rolandic operculum				1	
Right SFG		8			
Right SMG	3	3			
Right SPL		1			
Right STG	11	11			4
Right thalamus				1	

Table 2.6. A summary of key differences of activation compared to healthy controls in brain regions discussed as potentially significant by the authors, and the mechanisms which may underlie them (based on those described in Abel et al., 2015)

	Pre-therapy enhancement (compared to controls)			**Pre-therapy reduction (compared to controls)**		
Region	**More effort for executive control/lower efficiency**	**New strategies in non-language areas**	**Maladaption in RH**	**Different task performance**	**LH local brain damage**	**Disconnection in LH**
Left ACC						2
Left AG				7		
Left Caudate						2
Left IFG	7					
Left SMG					3	3
Left STG						2
Right AG				7		
Right IFG		7	7			
Right PCG				2		
Right STG				2		

1= Abel et al. (2014), 2= Abel et al. (2015), 3= Barbieri et al. (2019), 4= Cherney, Erickson and Small (2010), 5= Fridriksson (2010), 6= Fridriksson et al. (2012), 7= Johnson et al. 2019), 8=Kiran et al. (2015), 9=Leonard et al. (2015), 10= Marcotte et al. (2012), 11=Menke et al. (2009), 12= van Hees et al. (2014) ACC= Anterior cingulate gyrus, FG= Fusiform gyrus, PoG= Postcentral gyrus, SPL= Superior Parietal Lobule, CG = Cingulate Gyrus.

Table 2.7. A summary of key shifts of laterality in brain regions discussed as potentially significant by the authors, and the mechanisms which may underlie them (based on those described in Abel et al., 2015)

	Post-therapy LI shift towards LH (change over time)			**Post-therapy LI shift towards RH (change over time)**		
Regions	Storage of knowledge in language areas/increased familiarity	New strategies for language processing	More effort for executive control	Persistent LH malfunctioning	Higher efficiency due to priming or learning	New strategies for language processing
AG	2	2				
IFG	2	2,3*	3*	2	3*	1,3*
ITG	2	2		2		
MFG		3*	3*	2	3*	1,3*
MTG	2	2				
OFC	2	2				
PCG	2	2,3*	3*		3*	1,3*
SFG		3*	3*	2	3*	1,3*
STG				2		

1= Benjamin et al. (2014), 2= Hara et al. (2015), 3= Crosson et al. (2009). *Grouped into lateral frontal lobe region by paper. LH= Left Hemisphere, RH = Right Hemisphere, LI= Laterality Index.

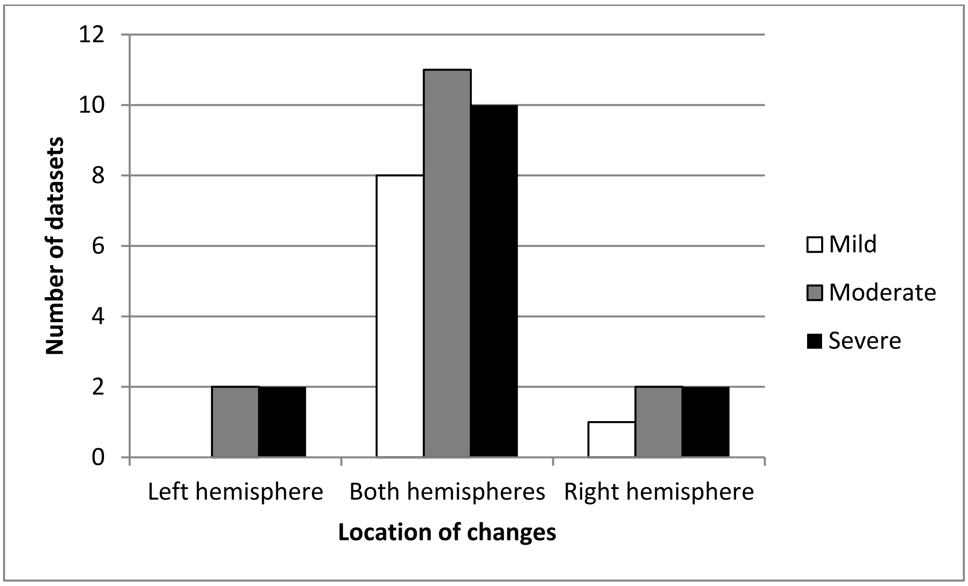

Figure 2.2.The number of datasets with post-therapy changes of activation or structure relative to pre-therapy in each hemisphere, grouped by aphasia severity.

4. Discussion

The criteria and search terms used in this systematic literature review returned a total of 28 studies, all of which were within the scope of the review. These criteria were able to isolate a reasonable number of studies which cover a range of therapy approaches motivated by neuroscientific theory and the links between their behavioural and neural effects. The inclusion of therapies with graphemic, lexical and syntactic elements as well as those focusing on aspects such as phonology illustrates the effectiveness of these criteria in covering a wide range of behavioural interventions. Thus, the criteria were appropriate and effective for the topic of this review.

The common activation of regions such as the IFG, MFG and MTG bilaterally in Table 2.5 supports existing views of these areas as playing a role in language production and its recovery in PWA (Crinion & Leff, 2007; Meinzer et al., 2007; Spielmann et al., 2016; Yagata et al., 2017). There is support for a role for areas in both the left and right hemispheres in improvement via therapy. Similarly, Table 2.6 shows evidence of post-therapy changes in lateralisation towards both the right and the left hemispheres. These results of varied patterns of activation and lateralisation support the long-standing calls for the use of more context-dependent models of recovery which take individual differences in behaviour, lesion location and size into account when considering how PWA will respond to different treatments and changes in activation in different regions of the brain (Crinion & Leff, 2007).

All in all, the data covered in this review were found to fit into the framework discussed in Abel et al. (2015) (Table 2.3). Mechanisms such as different task performance, LH local brain damage, disconnection in LH, persistent LH malfunctioning, and more effort for executive control were all discussed as potentially underlying key differences in activation by authors and were unchanged in the updates made to the framework. However, some mechanisms were grouped together since they were rarely referred to in isolation (storage of knowledge in language areas grouped with increased recognition and familiarity; higher efficiency due to priming grouped with higher efficiency due to learning). This broadening of categorisation may be beneficial in ensuring this framework effectively reflects the key mechanisms involved in activation change. Overall, we found that the framework discussed in Abel et al. (2015) adequately explained the key differences in activation and laterality in the data; with some adjustments, the framework was able to offer explanatory potential underlying mechanisms for all these key differences. The key differences between the studies in our review and the existing framework were that certain mechanisms were generally grouped when discussed in the studies in this review. Unlike in the Abel et al. (2015) framework, storage of knowledge in language areas, increased recognition and familiarity, higher efficiency due to priming and higher efficiency due to learning were not often referred to as individual mechanisms explaining the key differences in laterality and activation in the data. However, these mechanisms were sometimes included within a larger explanation. Table 2.3 therefore demonstrates the key points of difference between the Abel et al. (2015) framework and explanations given by papers in this review, as well as an amended framework to better align included mechanisms with those displayed in the literature. This amended framework grouped storage of knowledge in language areas with increased recognition and familiarity, and higher efficiency due to priming with higher efficiency due to learning. This reflects the presence of these mechanisms within the explanations for key differences within this review, while acknowledging that they are generally not present in isolation.

Despite the high variability in brain responses, our review illustrates that some mechanisms within the framework appear to be more crucial across PWA. For comparisons pre-post therapy, from most to least common by paper number, the mechanisms are ordered: new strategies for language processing (Abel et al., 2014; Benjamin et al., 2014; Cipollari et al., 2015; Crosson et al., 2009; Fridriksson, 2010; Hara et al., 2015; Kiran et al., 2015; Leonard et al., 2015; Marcotte et al., 2012; Menke et al., 2009), the storage of knowledge in language areas/ recognition and familiarity (Abel et al., 2014; Fridriksson, 2010; Hara et al., 2015; Kiran et al., 2015; Leonard et al., 2015; Marcotte et al., 2012; Menke et al., 2009), higher efficiency due to priming and learning (Abel et al., 2015; Cherney et al., 2010; Crosson et al., 2009; Fridriksson et al., 2012), more effort for executive control (Abel et al., 2014; Crosson et al., 2009; Menke et al., 2009), and persistent malfunctioning of the left hemisphere (Abel et al., 2014; Abel et al., 2015; Hara et al., 2015). A comparison between patients and healthy controls in our review has only been discussed in terms of its underlying mechanisms in three studies (Abel et al., 2015; Barbieri et al., 2019; Johnson et al., 2019), only one of which reported enhanced activations (Johnson et al., 2019); however, a recent study found enhancement representing higher demands and rewiring (Abel & Willmes, 2016). Additional theories of activation change associated with recovery, such as redundancy recovery and vicarious functioning (Zahn et al., 2006), and the implications these wider changes in strategy of language processing have for them, are considered in greater detail later in the discussion.

Figure 2.2 seems to indicate that therapy for mild deficits generally involves bilateral changes, while therapy for more severe deficits may sometimes result in changes more limited to a single hemisphere, either the left or the right. One possible explanation for this is that mild aphasia may result in more flexible compensation options due to less structural damage. For example, in van de Sandt-Koenderman et al. (2016), the 4 PWA with the mildest deficits in the ANELT (A1, A3, C1 and C4) showed a range of changes in lateralisation, with examples of increased right and left hemisphere lateralisation. Musso et al. (1999) also found that the right STG and left precuneus were best correlated with training-induced improvement in a study of participants with Wernicke's aphasia, suggesting that rehabilitation of aphasia with limited severity may involve and benefit most from bilateral changes. Conversely, moderate or severe deficits could increase reliance on either preserved left hemisphere language processes or completely new language processing strategies which may involve areas in the right hemisphere previously not primarily concerned with language processing. Naeser et al. (2005) found improved language ability following inhibition of right hemisphere activity via TMS in a case study of a participant with global aphasia, suggesting a possible link between language rehabilitation and increased reliance on the left hemisphere in more severe aphasia types. An important future area of research would be an investigation of these 3 possibilities via large scale trials.

There is evidence supporting activations and structural changes in many different regions being associated with improvement in language measures post-intervention, with the role of the right hemisphere in particular being a complex one. Several studies have shown results suggesting a beneficial effect of increased right hemisphere activation (Benjamin et al., 2014; Cipollari et al., 2015; Crosson et al., 2009; van de Sandt-Koenderman et al., 2016) or white matter connectivity (Schlaug et al., 2009), and others have shown positive effects of therapies believed to facilitate increased right hemispheric activation, such as pantomime observation (Gili et al., 2017), MIT (Wan et al., 2014) and intention therapy (Peck et al., 2004). However, others have found evidence of increased activity primarily in the left hemisphere also being effective (Fridriksson, 2010; Marangolo et al., 2016; Marcotte et al., 2012), or of increased activity in regions in both hemispheres occurring with improvement (Kiran et al., 2015; Menke et al., 2009).

The inclusion of healthy controls in some of the studies covered in this review also allows them to shed some light on the extent to which recovery after stroke is a result of redundancy recovery versus other theories, such as vicarious functioning. Studies such as Kiran et al. (2015), which found increased activations for PWA following therapy (and behavioural improvement) in brain regions also active in controls, support aphasia recovery as a result of redundancy recovery, as the regions showing increased activation in connection with recovery are regions that were already active in healthy controls, not unassociated regions which have been found to "adapt" to language processing. However, Tables 2.5, 2.6 and 2.7 indicate a number of studies which find activation changes in areas not traditionally seen as being linked to language, which may be the result of new strategies of language processing (Abel et al., 2014; Benjamin et al., 2014; Cipollari et al., 2015; Crosson et al., 2009; Leonard et al., 2015; Menke et al., 2009). This may support the theory of vicarious functioning (Zahn et al., 2006), as it implies the adaptation of neurons in regions not previously related to language.

The benefits of phonological and semantic therapies for PWA with primarily phonological or semantic deficits are also complex. Studies employing both have found a disparity between their effects on semantic and phonological deficits. van Hees et al. (2014) found that PWA with primarily phonological deficits significantly improved for both treatments, while those with primarily semantic deficits significantly improved for PCA, but not SFA. Abel et al. (2014) found overall significantly greater improvement in trained items for PWA with phonological deficits compared to those with semantic deficits, despite no significant differences in improvement between items trained using phonological versus semantic hierarchies. However, PWA with semantic deficits showed a training advantage compared to control items for phonologically and semantically trained items, while those with phonological deficits only showed this advantage for phonologically trained items. Abel et al. (2014) suggest this may be due to a disparity in the compensation strategies employed by PWA with primarily phonological or semantic deficits. It is also unclear how much compensation may be possible for each processing area via changes like increased right hemisphere activation. Gold and Kertesz (2000) suggest that the right hemisphere is more able to contribute to (and compensate for deficits in) certain areas of processing, such as lexical-semantic processing, compared to others, such as phonological processing. Ueno et al. (2011) suggest compensation may also be possible across the primarily semantic and phonological pathways, via changes to the division of processing within and between the two. It is therefore unclear exactly which compensation strategies are most common, and whether these strategies differ between semantic and phonological processing.

4.1. Limitations

The research presented in these studies was recent enough to be highly relevant to this subject area, with the majority of the studies having been published in the last five years. Most also had the benefit of including well-defined, replicable interventions which include a high number of sessions taking place over several weeks, ensuring that the interventions reach the intensity and volume required for potential efficacy. However, much of the literature has fairly low participation numbers of PWA, and there is often no healthy control group. The neuroimaging methodologies employed by the literature were fairly homogenous, with most of the studies solely employing fMRI. While this allows more comparison between studies, it also means that findings relating to brain structure rather than functional changes were sparse. Differences in spatial resolution between the different neuroimaging modalities used may also have affected the comparative precision of findings regarding individual brain regions. While the studies using fMRI are likely to be highly spatially precise, those using other modalities, such as SPECT or TMS-EEG, may be less so (Lystad & Pollard, 2009).

While it was necessary to restrict the activations covered in Tables 2.5, 2.6, and 2.7 to the key ones discussed by the authors in order for them to convey meaningful information, this method may also have caused a bias. Authors are more likely to discuss results for which a clear interpretation is available, or results which pertain to their specific hypotheses, possibly resulting in the tables over-representing certain mechanisms of recovery. However, the use of author interpretations also allows for alternate understandings of how regions may function in recovery and provides some evidence of interpretation of results relating changes in individual regions to more widespread differences. For example, activation in regions such as the left IFG or left MTG, which are normally considered to be part of the healthy language network, have sometimes been interpreted as the result of new strategies of language processing. This is due to increased activation of these regions being interpreted in the context of possible wider changes in strategy of language processing, rather than implying that they were previously unused in language processing and are ‘new’ locations of activation. Without the use of author interpretations, these possible alternate explanations for reactivation of parts of the healthy language networks may not have been considered.

A number of notable studies investigating the neural bases of aphasia recovery had to be excluded from the results of this review in order to maintain focus on the specific scope of research the review concerns. For example, several studies investigating ILAT and related therapies were considered at the full paper analysis stage of citation selection. These included Breier et al. (2006), McKinnon et al. (2017), Mohr et al. (2016), Nenert et al. (2017), Pulvermuller et al. (2005), Meinzer et al. (2009), Meinzer et al. (2008), and Richter et al. (2008). ILAT involves intensive therapy over a short period, with a focus on spoken communication based on everyday conversation (Mohr et al., 2014). It can be considered to feature a neuroscience-based approach at a broad explanatory level, as it implements neuroscientific principles like coincidence learning and massed practice, behavioural relevance, and focus on remaining language abilities to counteract learned non-use (Pulvermüller & Berthier, 2008). However, as ILAT studies do not target a specific aspect of language processing, modality or region of the brain, they may be considered to be less directly based in neuroscientific theory than other approaches discussed, and do not fit the specific criteria for inclusion in this review.

In addition to ILAT, other therapeutic approaches not considered in this review are nevertheless worth noting. Speeded cueing therapy aims to improve connected speech by increasing the speed of participant response via methods like deadline naming (Conroy et al., 2018). Naming to deadline requires participants to produce the target word before a beep in a picture naming task. The deadline is shortened as participant speed increases (Conroy et al., 2018; Hodgson & Lambon Ralph, 2008). Interfered-naming therapy is based on the finding that, in picture naming, different types of distractors to a target word (for example, words that share a semantic category or phonological similarity with the target word) can facilitate or inhibit successful target word production when presented at different times relative to the picture (Abel et al., 2009). Certain distractors presented at the right time as part of a naming therapy can therefore be used to facilitate recovery of linguistic-executive abilities (Abel & Willmes, 2016). These alternative approaches to cueing therapy are still novel and have comparatively little research dedicated to them, but are so far promising options for neuroscience-based aphasia treatment.

Also excluded from this study were papers discussing aphasia as a consequence of other causes than stroke, such as neurodegeneration, and papers investigating pharmacological treatments of aphasia. Primary Progressive Aphasia (PPA), resulting from neurodegeneration, generally involves progressive worsening of the deficit, as well as frequent appearance and worsening of other cognitive deficits over its development (Gorno-Tempini et al., 2011). Aphasia resulting from neurodegeneration, as well as other causes, such as traumatic brain injury or tumour, may differ broadly in terms of deficit profile from post-stroke aphasia (Glosser & Deser, 1991; Patterson et al., 2006), and so have not been included in this review. For example, Davie et al. (2009) found a higher proportion of anomic aphasia relative to other subtypes than would be expected from stroke victims in a study on PWA with tumours, indicating a general difference in the severity of deficit that might be expected from the two different causes of aphasia.

While it is possible that the number of studies included was limited by only searching three databases, and limiting results by language and year published, the studies returned cover a wide range of therapies and methodologies, suggesting that no large areas of study were likely to have been omitted due to these restrictions. Another potential limitation of the review is that only studies with 5 or more total participants were included. This criterion was included in order to place a focus on larger group studies rather than selections of case studies, which rarely use more than 5 total PWA and controls, allowing consideration of wider patterns in the relationships between rehabilitation method and improvement. This focus may have limited more subtle findings based on specific individual differences in language and lesion profiles, as lesions can be highly variable in PWA, meaning it can be hard to draw concrete conclusions from data averaged over groups. However, several studies included discussion of differences between PWA and their effects on therapy outcome, meaning that the effects of individual differences could be considered to some extent, while keeping the focus on the wider trends in the data.

While the field of studies covered in this review includes a range of different therapies, it also disproportionately focuses on those targeting semantic and/or phonological aspects of language production separately. There are comparatively few studies investigating different approaches, such as MIT, gesture, and intention therapies. There is also relatively little comparison between therapies, with many studies either lacking a control intervention or using the same intervention with some aspects removed or altered as a control, such as in Benjamin et al. (2014), which included an intention treatment with the complex hand movements removed, or Gili et al. (2017), in which pantomime was compared to actions using physical objects. While these controls are useful for isolating the beneficial elements of specific treatments, the lack of consistent comparison makes it difficult to draw conclusions regarding the comparative benefits of any specific therapy in any specific circumstance.

4.2. Conclusions

In conclusion, the results of this review highlighted several key features of the current literature on this topic. While some of these features have been discussed in previous reviews on the topic of treatments for aphasia (Cocquyt et al., 2017; Crinion & Leff, 2007, 2015; Hartwigsen & Saur, 2019; Heiss & Thiel, 2006; Kiran, 2012; Thompson & den Ouden, 2008), this review provides a comprehensive, updated overview of the current gaps in the literature which may be explored in the future. Firstly, while a wide range of therapies are covered in the literature, a large number of studies investigate methods such as semantic and phonological therapy, while others utilising alternative approaches, such as intention therapy and pantomime observation, are comparatively rarely studied. Also, most studies focus on a single method, often comparing it to a similar control which does not include potentially key features. While this is beneficial for determining that such therapies have an effect, it is still unclear which is most beneficial in any given circumstance. Thirdly, much of the literature includes smaller sample sizes. Due to the variability of aphasia, use of smaller samples with the individual analyses which are sometimes included can be advantageous in some respects. However, larger-scale studies which can provide more information about the overall impact of these therapies are currently relatively sparse. Finally, the variety of regions of activation and structural change identified in the literature and their links to increased or decreased improvement highlight the importance of context in rehabilitation. It is still unclear how the effectiveness of different therapy methods and regions of activation are modulated by the individual characteristics of the PWA undergoing the rehabilitation. Our recommendations for the field are therefore focused on addressing these gaps in the literature. The field would benefit from more studies assessing alternative approaches to semantic and phonological therapy, to provide a greater balance in terms of the degree of research into each approach to treatment. In combination with this, future studies in this area should also place more emphasis on providing context to the therapy benefits being investigated, through comparison to meaningful control treatments and ideally to other treatments with previously researched, quantified effects. More larger-scale studies are also necessary to allow performance of group-level analyses with sufficient power, to properly investigate the overall impact of different treatments. Inter-participant variation also needs to be examined in more detail in future studies, in order to develop our understanding of how to optimise treatment to individual PWA. Larger-scale studies would also allow this inter-participant variation to be explored in greater depth. Overall, while the current literature in this area has many strengths, application of our recommendations for future research would help to fill some of the gaps in the field and generate more findings with direct practical applications.

Acknowledgements:

The review was not registered, and a separate review protocol was not prepared.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations of interest: none

Data used for analyses in this review is available upon request from the corresponding author; no code was used in this review.

Chapter 3: Structural damage post-stroke and its relationship to therapy-related recovery in aphasia

Christopher Heath[a*], Paul Conroy[a], Natalie Busby[b], Stefanie Bruehl[a,c,d,e]

[a]School of Biological Sciences, University of Manchester, UK; [b]Department of Communication Disorders, University of South Carolina,Columbia, South Carolina; [c]St Mauritius Rehabilitation Centre, Meerbusch, Germany; [d]Clinical and Cognitive Neurosciences, Department of Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany; [e]Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine University, Duesseldorf, Germany

*Corresponding author:

Christopher Heath

School of Biological Sciences

University of Manchester

Oxford Road Manchester M13 9PL, UK

christopher.heath-2@postgrad.manchester.ac.uk

Statement of Contribution

Christopher Heath performed all of the preprocessing and conducted all the analyses that were undertaken in this chapter. Natalie Busby verified preprocessing and analysis. Stefanie Bruehl shared the neuroimaging and behavioural data of one of her previous projects as presented in this chapter. Stefanie Bruehl and Natalie Busby provided guidance and support throughout the analysis. Stefanie Bruehl, Natalie Busby and Paul Conroy provided feedback on the final draft.

Abstract

There is a great deal of research exploring the effects of various forms of behavioural treatment of the language processing disorder aphasia, and how this is mediated by neural factors such as the extent and location of damage, or regions of activation, in the brain. However, the majority focuses primarily on either grey matter structural changes, or functional scans showing changes in activation. While our understanding of the links between white matter connectivity and language processing has been advanced by a number of studies, research in the context of post-stroke aphasia treatment is limited. Our study therefore aimed to investigate the relationships between language deficits and white matter connectivity via analysis of T1 and DTI scans of 18 participants, most of whom underwent speech and language treatment. Both correlations and TBSS analyses were performed. Our findings showed some support for a dual pathway model of language processing, as well as for previous findings regarding the role of the FAT, fornix and corticospinal tract in language. This study underlines the potential benefit of white matter analyses in furthering our understanding of treatment-related neural changes after stroke.

Introduction

Aphasia, an acquired language disorder which may occur as a result of damage to the brain post-stroke (Budd et al., 2010), has a significant impact on the quality of life of those it affects (Lam & Wodchis, 2010). However, it is often possible for some level of recovery of language abilities to occur, both spontaneously and as a result of interventions such as Speech and Language Therapy (SLT) (Cappa, 1998; Doogan et al., 2018; Goldenberg and Spatt, 1994). These behavioural changes are generally accompanied by physical changes in the brain, which may include both functional changes in patterns of activation, and structural changes as neuroplasticity causes altered patterns of connectivity (Kiran & Thompson, 2019; Thompson, 2000).

A common way of observing the effects of both damage as a result of stroke, and recovery of language ability on the structural and connectivity level is via Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI). Structural MRI allows segmentation of the brain into different types of matter and measurement of matter volume and location (Smith et al., 2004), allowing identification of the size and shape of grey matter, white matter and lesion locations; subsequently, damaged brain regions that overlap with crucial connectivities can be determined. DTI measures the diffusion of water molecules in the brain. As the myelination of axonal tracts prevents equal diffusion and thus affects the direction of diffusion (Aung, Mar, & Benzinger, 2013), DTI can be used to judge information such as the location, size, and integrity of these tracts (Schlaug et al., 2009). It is therefore a useful way of observing structural white matter changes in a meaningful way, giving a quantitative measurement of connectivity in a given region.

Observation of structural profiles of people with aphasia (PWA) is important for a number of reasons. Possibly the most obvious is that comparing regions of damage and reduced connectivity to the specific deficits exhibited by a PWA may shed light on the roles these connections play in language production (Geva, Correia, & Warburton, 2011). As PWA may vary significantly in both their language deficits and the size and location of their lesions (Crinion & Leff, 2015), investigating the relationships between these factors can afford us insights into what damage is associated with different deficits in language production. Moreover, DTI scans before and after an intervention may detect changes in connectivity that relate to behavioural improvement (or comparative lack thereof). This allows us to learn more about the link between connectivity and language deficits, but also gives an insight into the benefits of different structural changes in the context of rehabilitation (Schlaug et al., 2009). Pre-therapy structure alone may also be informative in therapy studies, as initial white matter integrity has in past studies been found to be related to the chance of successful recovery (Geva et al., 2011). Investigating the relationship between pre-therapy structure and PWA's responses to different forms of therapy may allow us to form predictions regarding not only the connections in which preservation is most closely linked to reduced impairment, but also the connections in which preservation is related to greater potential for recovery of language abilities. This may also explain how these relationships are affected by the type of intervention used. Due to the variability of both deficit and lesion profiles in aphasia, an understanding of how structure relates to deficit across all stages of aphasia rehabilitation is crucial for providing the best intervention and prognosis for each individual PWA (Crinion & Leff, 2015).

The size and location of lesions in aphasia are thought to play a large role in both initial deficit profile and in the conditions required for optimum recovery (Perani et al., 2003; Vitali et al., 2007). For example, several regions in the left hemisphere are also frequently associated with core language processes, such as the Inferior Frontal Gyrus (IFG), Middle Temporal Gyrus (MTG), and Middle Frontal Gyrus (MFG) (Abel et al., 2014; Cipollari et al., 2015; Fridriksson, 2010; Kiran et al., 2015; Leonard et al., 2015; Menke et al., 2009). Damage to these regions would also be likely to have a disproportionately severe effect on language ability. The role of right hemisphere activation in recovery (specifically, whether or not it is beneficial) has also been previously debated in the literature (Cipollari et al., 2015; Gold & Kertesz, 2000; Schlaug et al., 2010). It is possible that the benefit of right hemisphere compensation may depend on the size/location of the lesion, with PWA with less preserved left hemisphere language regions and pathways benefitting more from right hemispheric compensation, however when there is more integrity in the left hemisphere, this compensation may inhibit effective activation in perilesional areas (Crosson et al., 2007). To understand the most effective forms of treatment for different circumstances, it is important to investigate how initial structural damage may affect both recovery and response to different interventions.

Investigation of the effects of specific white matter damage, as opposed to more general regions of damage, may be particularly important in aphasia as it has been theorised that different components of language may be processed separately, along different white matter pathways (Saur et al., 2008; Yang et al., 2017). The degree of deficit in semantic and phonological ability may differ in individual PWA (Abel et al., 2014), suggesting some separation in the structures involved in their processing. This separation was incorporated into the dual pathway model (Ueno et al., 2011). In this model, primarily semantic (the input-semantic-output mapping; involved in comprehension and speaking/naming) and primarily phonological (acoustic-phonological and phonological-motor processes; involved in repetition) processing is performed by a ventral pathway (primarily connected via the middle longitudinal fasciculus and the extreme capsule) and a dorsal pathway (primarily connected via the arcuate fasciculus and possibly also the superior longitudinal fasciculus) (Saur et al., 2008) respectively. In practice, repetition and comprehension could be considered as the key abilities associated with the dorsal and ventral pathways respectively, and regional activations for repetition and comprehension tasks have even been used as a basis for identifying the most probable anatomical pathways of the two tracts (Saur et al., 2008). It is currently unclear whether the two pathways are fully distinct or if compensation between them is possible. PWA with primarily phonological and semantic deficits have been found to have differing responses to some forms of intervention (Abel et al., 2014; van Hees et al., 2014). Understanding how white matter damage affects the different components of language processing and their ability to recover may therefore have important implications for the development of effective interventions.

A specific area of interest for these analyses is regarding the role of the left frontal aslant tract (FAT) in language processing. Damage to this tract is correlated with decreased verbal fluency in Primary Progressive Aphasia (PPA), and for right-handed subjects it is lateralised to the left hemisphere, which also implies that it may be involved in language (Catani et al., 2013). Neurostimulation in regions overlapping with the left FAT has also been found to consistently induce stuttering (Kemerdere et al., 2016). However, research into the correlation between damage and fluency in post-stroke aphasia is currently limited (Basilakos et al., 2014). Identifying more specific tests that correlate with FAT connectivity may help to clarify its role in language processing.

A number of strands of research have made great progress identifying the links between white matter and language processing, as well as the structural pathways involved in semantic and phonological processing. However, we feel that there are still questions to be answered regarding how semantic and phonological processing relate to structural differences and changes throughout recovery from stroke, and in the context of PWA undergoing treatment to address semantic and phonological deficits. With this in mind, we hypothesise that PWA with damage to key language regions including the IFG, MFG and MTG (which have been proposed to play key roles in the language processing pathways previously discussed (Abel et al., 2014; Cipollari et al., 2015; Fridriksson, 2010; Kiran et al., 2015; Leonard et al., 2015; Menke et al., 2009)) will have significantly lower scores on language tests. Secondly, we hypothesise that connectivity measures for the tracts primarily involved in the ventral and dorsal pathways will respectively correlate with scores on tests for comprehension and repetition tasks, and semantic and phonological ability. Our third hypothesis is that connectivity values for right hemisphere homologues of the tracts in the dual pathways will also correlate with scores for these tests, due to compensation. Finally, we also hypothesise that connectivity in the left FAT will be correlated with measures of semantic and phonological fluency.

Methods

The data analysed in this section was collected at RWTH Aachen University Hospital as part of a series of scanning, therapy and behavioural tests (see Abel et al., 2015; Abel, Willmes & Binkofsky, 2021a, 2021b). A total of 20 PWA underwent T1 and/or DTI scans, resulting in 18 participants with both scans in addition to language testing. T1 scans were acquired in 1x1x1mm resolution. These 18 PWA were included in the analyses in this section. This group was composed of 14 males and four females, all right-handed, with a mean age of 49.9 years (Standard Deviation: 13.07). All had experienced a left hemisphere stroke at least 4 months previously (so were in the chronic phase of recovery (Anglade et al., 2014)). Full characteristics of the participants are provided in Table 3.5. Initial 'Pure' and 'Interfered' naming scores (from a maximum score of 480), the Aachen Aphasia Test (AAT) battery, and impairment measured by the computational Dell model pre-therapy were collected for all 18 PWA, while scores on the Regensburg Word Fluency Test (RWT) pre-therapy were collected in 13 PWA, and 12 PWA post-therapy.

Treatment consisted of 4 weeks of interfered-naming therapy, consisting of treatment of both lexical and executive processing. Naming was assisted by a semantic cueing hierarchy for 2 of these weeks, and a phonological cueing hierarchy for the other two, with the order of hierarchy presentation pseudo-randomised between participants. Pictures from Snodgrass and Vanderwart (1980) were selected as items. Participants were confronted with the pictures and required to name them. Nearly simultaneously, an auditory distractor was presented, but PWA were to still focus on naming the item. If they were unable to produce the correct response, they were given increasing assistance using the semantic or phonological cues. Each item was presented 12 times for interfered-naming and twice for comprehension. In order to facilitate this naming task, it was preceded by a comprehension task in which PWA were to point at the correct item related to the heard word. These words were taken from the set of auditory distractors for the later interfered-naming task.

Dell impairment scores were also collected post-therapy in the same 12 PWA (those who underwent the therapy study). A group of 22 healthy controls was also included for some analyses. Error pattern/ performance values were prepared at the RWTH Aachen University Hospital, Germany, and were handed over to researchers at the University of Manchester, UK, for further analyses in combination with brain imaging data (see below).

This group underwent both T1 and DTI scans, had an average age of 52 and an average of 12 years in education. Analysis of this pre-existing T1 and DTI data originating from RWTH Aachen, was then performed by researchers at the University of Manchester. All brain imaging data discussed in this section was collected at RWTH Aachen University Hospital before undergoing pre-processing and analysis (after anonymisation) at the University of Manchester.

Table 3.1. Characteristics of the 18 PWA included in analyses

PWA	Sex	Age (years)	Years of education	Months post-onset	Lesion localisation	Pure naming (max 480)	Interfered naming (max 480)
P1	M	50	13+U	5	left MCA, PCA: TO & W	401	389
P2	M	47	10	29	left MCA: T & BG & W	435	433
T-P3	M	54	13+U	10	left MCA: TP & W	348	303
T-P4	M	61	9	26	left MCA: T, BG & W	306	276
T-P5	M	54	13+U	38	left MCA: FT, BG & W	262	187
T-P6	M	21	12	30	left MCA: FTP	313	309
P8	M	61	10	49	left MCA: T, BG & W	361	345
T-P10	M	61	13	38	left MCA: FT, BG & W	355	343
T-P11	F	35	13+U	28	left MCA: FTPO, BG & W	158	90
T-P12	M	57	9	19	left MCA: BG & W	140	163
T-P13	M	48	13+U	25	left MCA, ACA: FTPO, BG & W	325	308
T-P15	F	28	13	61	left MCA: FTP, BG & W	300	275
T-P16	M	74	13+U	63	left MCA:FP, BG & W	229	130
T-P17	M	52	9	20	left MCA: FTP	322	319
T-P18	F	63	10	4	left MCA: TP	368	375
T-P19	F	39	13+U	6	left MCA: FT & BG	261	255
T-P20	M	49	13+U	8	left MCA: TP	297	243
T-P21	M	45	13	9	left MCA: TP	437	451
Mean		49.94		26		312.11	288.56
Standard Deviation		13.07		18.34		81.65	99.14

P = person with aphasia, T = received therapy; +U = Attended or graded at university. ACA= Anterior Cerebral Artery, MCA= Middle Cerebral Artery,

PCA= Posterior Cerebral Artery, F= Frontal, T=Temporal, O= Occipital, P= Parietal, BG= Basal Ganglia, W= White Matter.

Pre-processing

Anonymised raw T1 and DTI data was received from RWTH Aachen University Hospital, Germany. All subsequent pre-processing and data analysis stages (see below) were performed by researchers at the University of Manchester. After receiving the anonymised raw T1 and DTI data, several stages of pre-processing were performed in FSL (the Functional MRI of the Brain (FMRIB) software library) (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) and SPM12 (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007) before analyses took place. Brain extraction from the skull was performed using Optimized Brain Extraction (optiBET) software (Lutkenhoff et al., 2014), and binarised brain and lesion masks were generated. Manual changes to the mask were then performed to dilate the mask and fill in any holes (such as those caused by lesions). The images and mask were then normalised in T1 space in SPM12, using the ALI toolbox (Seghier et al., 2008).

For the DTI data, brain extraction (BET) was performed in FSL (Smith, 2002), followed by eddy correction (Andersson & Sotiropoulos, 2016) to account for data distortions and to produce of brain masks. DTIFit was used to fit a diffusion tensor model at each voxel. Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques (BEDPOSTX) was used to run Markov Chain Monte Carlo sampling to build up distributions on diffusion parameters at each voxel (Behrens, Berg, Jbabdi, Rushworth, & Woolrich, 2007; Behrens et al., 2003) using the following parameters: number of fibres per voxel = 3, model=3 [deconvolution model with zeppelin], burn-in period=3000, number of jumps = 1250. This models the probability of a white matter tract in each direction for every voxel. A probabilistic Anatomical Connectivity Map (ACM; Bozzali et al., 2011) could then be produced for each PWA using FSL's probtrackx2 (Behrens et al., 2007; Behrens et al., 2003) function (parameters included: P(number of streamlines per voxel) = 50, mask = binary whole brain mask obtained using BET on the B0 volume, tracking mask = thresholded MD mask outputted from FSL's DTIFit, and the exclusion mask = the inverted tracking mask). This resulting image is a whole-brain connectivity map, with higher values representing a larger number of probabilistic streamlines passing through the voxel (Bozzali et al., 2011). These ACMs were then converted to MNI (standardised) space via T1 space using FSL's FLIRT (FMRIB's Linear Image Registration Tool) (Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson & Smith, 2001). PWA head size (and subsequent number of streamlines initiated in the ACM) was controlled for by dividing the ACM by the number of voxels present in the brain mask (excluding areas such as cerebrospinal fluid (CSF) in which connectivity is irrelevant). Finally, the ACMs were smoothed. To calculate connectivity scores (ACM scores) for individual tracts, masks of white matter tracts (including tracts from both Catani & de Schotten (2008) and Catani & de Schotten (2012)) were overlaid onto individual ACMs and average connectivity scores were extracted. This enabled connectivity scores to be extracted for every tract in each participant. Figure 3.1 demonstrates this for the left Inferior Longitudinal Fasciculus (ILF) mask – the area covered by the red binary mask would be included in the ACM value for this tract.

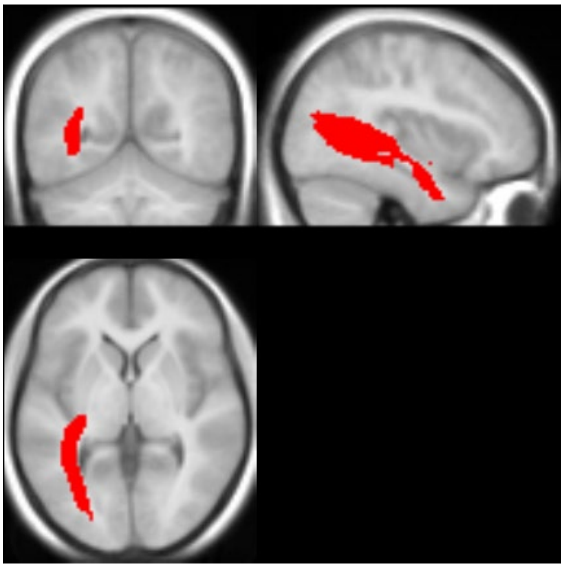

Figure 3.1. Binarised white matter tract mask overlaid on an average brain – left Inferior Longitudinal Fasciculus.

T-tests and correlations

Lesions were then separately overlaid onto the Harvard-Oxford atlas (Desikan et al., 2006) in MRIcron (http://mricro.com/mricron) to identify which grey matter regions were directly affected by each lesion (demonstrated in Figure 3.2).

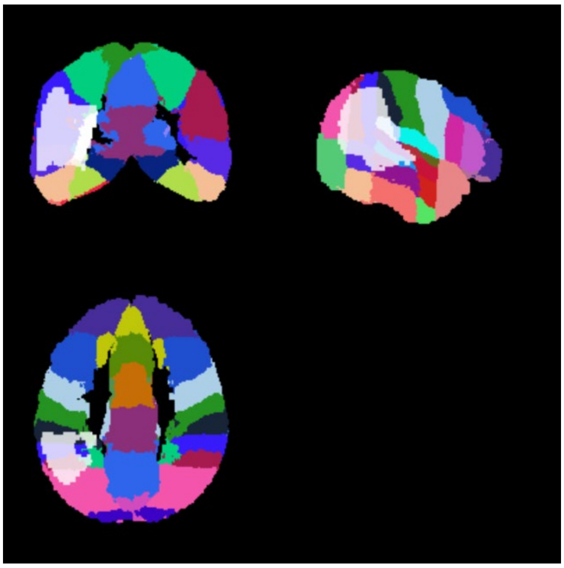

Figure 3.2. Lesion binary mask for a single PWA overlaid over Harvard Oxford atlas.

Once this anatomical data had been calculated, both t-tests and Spearman correlations were performed in order to explore the relationships between damage or connectivity in different regions or tracts and language ability.

Independent-samples t-tests were carried out comparing language scores in PWA with vs without a lesion in a given region. This analysis focused on left hemispheres regions which have been implicated in the dual pathway model of language: the SMG, insular cortex (IC), STG, IFG, and regions involving the ventral anterior temporal region (vATL).

Spearman correlations were carried out between ACM values of left-hemisphere tracts, and different language battery scores to explore the relationships between the connectivity in different white matter tracts and language ability. Again, of particular interest were tracts believed to make up the dorsal and ventral language pathways – the arcuate fasciculus (AF), extreme capsule (included in the uncinate fasciculus (UF) in the tract atlas being used (Catani & de Schotten, 2008; Catani & de Schotten, 2012)), the 3 components of the superior longitudinal fasciculus (SLF I, SLF II and SLF III) and inferior longitudinal fasciculus (ILF) (which is believed to contribute to both pathways (Saur et al., 2008)). As the middle longitudinal fasciculus (MLF) was not included in Catani's atlas, it was not considered in this set of analyses. The Inferior Fronto-Occipital Fasciculus (IOFF) was also included, due to its spatial proximity to both the ILF and MLF, and its inclusion in the ventral pathway in some models (Yang et al., 2017). Finally, the relationship between the FAT and fluency was calculated. The pre-therapy language scores used in these analyses were for Pure and Interfered Naming, AAT naming, repetition, and comprehension language subtests, AAT semantic structure and phonological structure spontaneous speech scores, RWT phonological and semantic fluency, and Dell model semantic and phonological impairment scores (Bruehl et al., submitted). Due to the number of tests being performed, and the variability of the p value in this situation, we focused primarily on r value to gain an understanding of which relationships between tract connectivity and language measures. Moderately or strongly positive or negative correlations were highlighted in Table 3.2.

Tract-Based Spatial Statistics

In addition to the basic Spearman correlations, voxelwise statistical analysis of the FA data was also carried out using TBSS (Tract- Based Spatial Statistics (Smith et al 2006), part of the FMRIB software library (FSL) (Smith et al., 2004)). TBSS involves creation of an average FA skeleton from individual scans. FA images are created and nonlinear registration of these images into standard space is performed. An average FA skeleton is constructed using the cores of each tract, with the tract edges being averaged into this 'skeleton'. Data from, in this case, individual PWA can then be projected onto this skeleton to identify significant voxels relative to the average skeleton via cluster analysis. This can result in difficulties in performing comparisons effectively in post-stroke aphasia, as inclusion of scans containing large lesions in the initial mean FA skeleton can result in lesioned tracts not appearing in the skeleton and thus not being included in group-level analyses. Therefore, we used control scans to create the FA skeleton, allowing us to compare different PWA populations to each other and to the healthy control group while ensuring all tracts were included.

A total of 8 group comparisons were performed using TBSS, with groups based on which PWA produced improved semantic, phonological, or overall ability from pre-post treatment. This change in ability was determined by observing change in number of semantic and nonword errors during a picture naming task. The comparisons were as follows:

1. PWA vs. healthy controls
2. Semantically improved vs. semantically non-improved PWA
3. Phonologically improved vs. phonologically non-improved PWA
4. Semantically improved PWA vs. healthy controls
5. Phonologically improved PWA vs. healthy controls
6. Overall improved PWA vs. overall non-improved PWA
7. Overall improved PWA vs. healthy controls
8. Overall non-improved PWA vs. healthy controls

These comparisons were performed in both ‘directions’ – e.g., for comparison 1, the PWA group was compared to a healthy control ‘baseline’, and the healthy control group was compared to a PWA ‘baseline’, in order to ensure that any significant differences in FA in either direction would be identified. Semantically improved PWA were those who made fewer semantic errors post-treatment in comparison to pre-treatment, while phonologically improved PWA were those who made fewer nonword errors post-treatment in comparison to pre-treatment. Overall improvement was determined by pure naming scores pre- and post-treatment.

Results

T-tests and correlations

Significant differences in language scores between the group of PWA with/without a lesion in the left IFG were found: Pure Naming (p = .005), Interfered Naming (p = .008), and the AAT subtest comprehension (p = .003). Significant differences in Pure Naming (p = .004), Interfered Naming (p = .011) and AAT subtest comprehension (p = .005) between the group of PWA with a lesion in the left SFG and the group of PWA without a lesion in the left SFG were also found.

Initial Pure and Interfered naming significantly correlated with connectivity in the left AF (Pure: r = .546, p = .019, Interfered: r =.538, p = .021), UF (Pure: r = .564, p = .015, Interfered: r = .534, p = .023), IFOF (Pure: r = .635, p = .005, Interfered: r = .633, p = .005), SLF II (r = .494, p = .037), and SLF III (r = .680, p = .002). Connectivity in the AF positively correlated with semantic impairment as measured by the Dell model (r = .522, p = .026), as was connectivity in the SLF III (r = .565, p = .015). Phonological fluency scores on the RWT also significantly positively correlated with connectivity in the left IFOF, SLF III and UF, while semantic fluency score correlations with the IFOF and UF approach significance (r = .763, p =.004 and r = .518, p = .07 for phonology and semantics respectively for the IFOF, r = .763, p =.004 and r = .550, p = .052 for phonology and semantics respectively in the UF, r = .767, p = .002 for phonology in the SLF III). Sub scores of the AAT repetition and comprehension subtests were correlated similarly in the left AF, (comprehension: r = .725, p = .001 repetition: r =.706, p = .001), SLF II (comprehension: r=.663, p=.003, repetition: r = .588, p = .01), and SLF III (comprehension: r = .735, p = .001, repetition: r = .629, p = .005). In the left ILF comprehension showed a more positive correlation than repetition (comprehension: r = .750, p < .001, comprehension: r = .592, p = .01). The AAT repetition subtest also correlated with connectivity in the left SLF I (r =.561, p = .016).

Connectivity scores in the right hemisphere homologues of these tracts did not significantly correlate with initial Pure or Interfered naming. However, a positive correlation was found between connectivity in the right UF and AAT spontaneous speech phonological structure scores (r = .537, p = .022).

Connectivity in the left FAT also significantly correlated with initial Pure (r = .759, p < .001) and Interfered (r = .756, p < .001) naming and the Dell measure of semantic impairment (r = .538, p = .021), as well as RWT semantic (r = .750, p = .003) and phonological (r = .735, p = .004) fluency scores, and scores on the AAT subtest Comprehension (r = .595, p = .009). These results are summarised in Table 3.2.

It should be made clear that these initial analyses were performed in an exploratory capacity in order to gain a broad understanding of the dataset in preparation for the use of TBSS analyses. While they may provide some benefit in terms of looking at the overall ‘shape’ of the data in combination with TBSS results, the use of multiple comparisons means that significant weight should not be placed on individual results in these exploratory analyses. They are included primarily as an initial exploration of the data, to help guide subsequent analyses.

*Table 3.2. Notable correlations between connectivity scores and language measures. *Moderately positive correlation (r>0.4). **Strongly positive correlation (r>0.6).*

Language measure	WM tract	R score	P value
Pure Naming	Left AF	0.546*	0.019
Pure Naming	Left UF	0.564*	0.015
Pure Naming	Left IFOF	0.635**	0.005
Pure Naming	Left SLF II	0.494*	0.037
Pure Naming	Left FAT	0.759**	<.001
Interfered Naming	Left UF	0.534*	0.023
Interfered Naming	Left AF	0.538*	0.021
Interfered Naming	Left IFOF	0.633**	0.005
Interfered Naming	Left FAT	0.756**	<.001
Interfered Naming	Left SLF II	0.68**	0.002
Dell s	Left AF	0.522*	0.026
Dell s	Left SLF III	0.565*	0.015
Dell s	Left FAT	0.538*	0.021
RWT Phonological fluency	Left IFOF	0.763**	0.004
RWT Phonological fluency	Left SLF III	0.767**	0.002
RWT Phonological fluency	Left UF	0.763**	0.004
RWT Phonological fluency	Left FAT	0.735**	0.004
RWT Semantic Fluency Score	Left IFOF	0.518*	0.07
RWT Semantic Fluency Score	Left UF	0.55*	0.052
RWT Semantic Fluency Score	Left FAT	0.75**	0.003
AAT Repetition	Left AF	0.706**	0.001
AAT Repetition	Left SLF II	0.588*	0.01
AAT Repetition	Left SLF III	0.629**	0.005
AAT Repetition	Left ILF	0.592*	0.01
AAT Repetition	Left SLF I	0.561*	0.016
AAT Comprehension	Left AF	0.725**	0.001
AAT Comprehension	Left SLF II	0.663**	0.003
AAT Comprehension	Left SLF III	0.735**	0.001
AAT Comprehension	Left ILF	0.75**	<.001
AAT Comprehension	Left FAT	0.595*	0.009

Tract-Based Spatial Statistics

When performing TBSS, a number of differences were discovered across different tracts. However, when considering only the corrected *p*-value at a threshold of $p < .05$, only the comparisons between semantically improved PWA and healthy controls retained significance. Several tracts in the right hemisphere were found to have significantly higher connectivity in healthy controls than in semantically improved PWA pre-therapy. Large clusters of significantly different FA values were found in the right corticospinal tract and right fornix. Smaller clusters were also found in the right ILF and anterior arcuate.

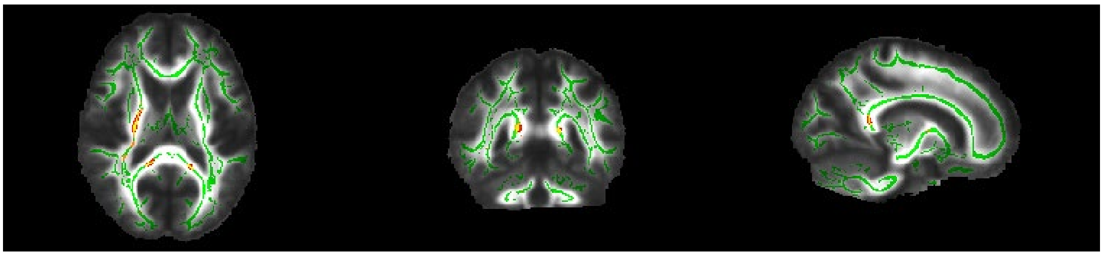

Figure 3.3. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- most anterior view. Red and green show skeletonised white matter tracts- areas with a significant difference are displayed in red.

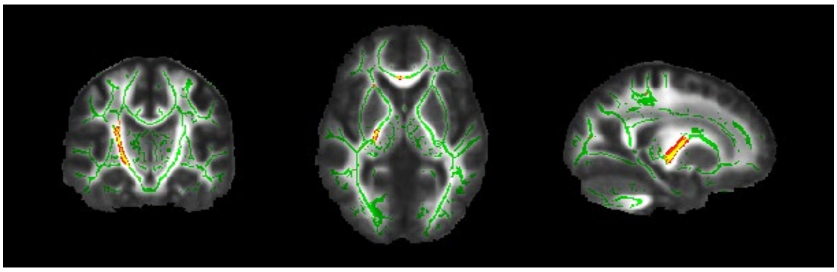

Figure 3.4. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- middle view. Red and green show skeletonised white matter tracts- areas with a significant difference are displayed in red.

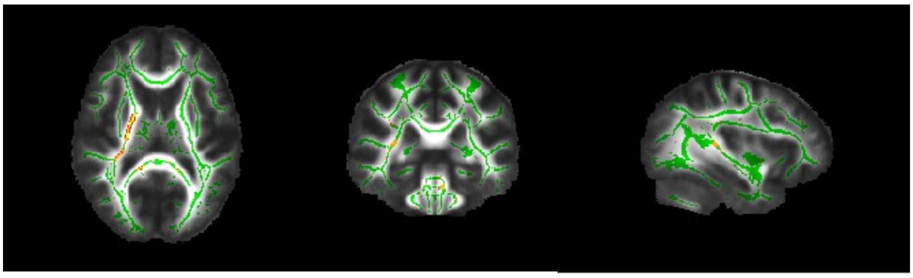

Figure 3.5. Significant differences in FA values between healthy controls and PWA who improved in semantic ability over the course of treatment- most posterior view. Red and green show skeletonised white matter tracts- areas with a significant difference are displayed in red.

Discussion

TBSS analysis returned interesting findings, showing a difference in diffusion between controls and PWA who would improve semantically following treatment in several tracts- the right corticospinal tract, fornix, ILF, and anterior arcuate. This previous research shows that while not all are homologues of left hemisphere tracts theorised to be directly involved in the ventral pathway, all their homologues have some links to semantic processing. The ILF is generally theorised to play a role in the semantically-focused ventral pathway (Saur et al., 2008). While the corticospinal tract is not traditionally implicated significantly in either pathway, its degeneration has more recently been found to be associated with semantic dementia (Josephs et al., 2013). Fractional anisotropy values in the left ILF and corticospinal tracts were also found to correlate positively with oral naming in brain damaged participants in a study examining semantic processing (Han et al., 2013). The role of the ILF in semantic processing is also supported by studies in healthy participants as well as the majority of studies involving PWA (Cocquyt et al., 2020), with negative correlations being found between MD values and both semantic learning (Ripolles et al., 2017) and semantic autobiographical memory (Hodgetts et al., 2017) in healthy participants. There is also some limited evidence of involvement of the fornix in semantic memory (Solca et al., 2013), although there is also evidence to the contrary (Hodgetts et al., 2017). Finally, while the arcuate fasciculus is generally associated primarily with the dorsal tract, 40% of studies examined in Catani et al (2020) found correlations between arcuate fasciculus DTI parameters and behavioural semantic performance. The anterior segment of the arcuate fasciculus, in which we found a difference between the aforementioned groups, has also previously been linked to vocalisation of semantic content (Catani et al., 2005). This difference in FA therefore included several tracts which, while not considered to be primary tracts in the ventral pathway but nonetheless are linked to semantic processing and speech production. Differences in FA in the ILF were also found to be very limited in comparison to the corticospinal tract and fornix. It may therefore be the case that these pathways are those which, while not playing primary roles in the ventral pathway, cannot have their roles effectively transferred from the left hemisphere to right hemisphere homologues. This could mean that PWA with damage to these areas in the right hemisphere were still able to effectively improve via neuroplasticity or changed activation patterns in the left hemisphere. Alternatively, as many of these tracts are not primarily involved in semantic processing in the ventral pathway, these may be the tracts in which damage can be more effectively bypassed via changed activation patterns without requiring a complete change in activation to the right hemisphere.

One alternate explanation for these results, however, is that they are at least partially due to the nature of the statistical analysis performed. Without caution, this could potentially lead to inaccurate interpretation of which tracts are 'responsible' for differences between groups. The relationships between these tracts' FA values in groups with different behavioural outcomes may be mislocalised or distorted as a result of statistical test used. For example, statistical power when identifying differences in FA (or other measures of damage) varies across the brain. In aphasia research, this variability is particularly notable between the left and right hemispheres, as greater variability in damage in the left hemisphere results in greater statistical power here. This leads to statistical 'biasing' towards the detection of significant differences in areas which have higher power, potentially skewing results in terms of which significant differences are shown by these analyses. Similarly, mislocalisation of damage may occur as a result of statistical analyses not accounting for the relationships between damage to different voxels and assuming that damage is independent between each voxel (Mah et al., 2014). This, again, is influenced by variability of damage, and may occur where lesions may result in common overlap of damage between functionally relevant and irrelevant areas due to factors such as the distribution of the vascular tree (Mah et al., 2014). As the results of the TBSS analysis mostly show differences in the Middle Cerebral Artery (MCA) territory, it is possible that these differences have been mislocalised. We must therefore consider that the TBSS results may have been affected by uneven statistical power distribution and mislocalisation and may therefore be inaccurate in identifying the exact tract regions most linked to semantic improvement.

Differences in group-level language scores were also found between PWA with and without lesions in the left IFG and SFG. In the left IFG, this association between preservation and improved naming scores supports existing literature. The IFG has been implicated in many previous studies as playing a key role in language processing (Abel et al., 2014; Fridriksson, 2010; Kiran et al., 2015). It is also often incorporated as part of the ventral aspect of the dual pathway model. The effects of its damage on comprehension, which is more strongly associated with the ventral pathway, therefore also supports the assertions of this model (Saur et al., 2008; Ueno et al., 2011). Similarly, the improved naming scores in participants without an SFG lesion support previous literature- the left SFG has previously been found to have increased activation after speech and language therapy (Abel et al., 2014; Kiran et al., 2015), and in a past review was suggested to be involved in semantic processing, being more activated for semantic than phonological decisions (Price, 2012). Greater evaluation of the specific role of the SFG in language processing may be a worthwhile line of investigation in future studies. Future studies would also benefit from including more individual treatment, given our focus in this study on group-level analyses. Lesions in a number of other regions which have previously been associated with language processing were not found to be significantly related to language test score differences. 'Lesion-defined' approaches using patients with a shared region of injury (Bates et al., 2003) has yielded results in the past (Chao & Knight, 1998), as has the specific method of separating participants into groups based on whether or not their lesion extends into a given region, and then comparing their behavioural scores (Friedrich, Egly, Rafal, & Beck, 1998). However, this is a basic method of analysis which does not capture much nuance in judging damage. The results of this form of analysis should therefore be interpreted with caution, and used as a starting point to direct more subtle analyses.

In the analysis of white matter, a more nuanced measure of the damage caused to a given tract can be used in the form of a connectivity value, which takes into account factors like the size and integrity of remaining fibres to provide a single score for the overall effectiveness of the connections formed by this tract. The connectivity values for the left hemisphere tracts proposed to form the ventral and dorsal pathways correlated with measures of PWA language ability. Only one tract only showed a significant correlation with one of the scores out of the AAT repetition and comprehension scores, rather than showing a correlation with either both or neither of them. This was the SLF I, which was correlated with the AAT repetition subtest, but not comprehension scores, fitting its potential role as part of the phonologically-focused dorsal pathway. Otherwise, there seems to be a large amount of overlap between phonology and semantics in the two pathways, although it is unclear if this is a result of compensation or if this overlap would also be present in healthy controls.

There are several key questions regarding the dual pathway model which have not yet been fully explored in the literature. These include the degree of overlap between the processes performed by both pathways, and the possibility of compensation between the two in the instance that one pathway is damaged. These results appear to indicate a clear overlap between the pathways in this group of PWA in the chronic stage of recovery. There is some indication of the pathways having different foci; for example, the ILF, part of the ventral pathway (Ueno et al., 2011; Yang et al., 2017), was more positively correlated with comprehension than repetition. However, correlations were found for phonological fluency and semantic ability in the IFOF and UF. Also, the AF showed similar correlations for both comprehension and repetition, suggesting that there is also a large amount of overlap in the roles of these pathways in this group of PWA.

Some weak correlations were also found between connectivity in right hemisphere ventral tracts and scores for repetition and phonology. Assuming that these PWA, being in the chronic stage, have already recovered some of their language ability since their initial stroke, this does not support claims that right hemisphere activation is part of a temporary shift in lateralisation which reverts back to the left hemisphere later in recovery (Saur et al., 2006). It instead supports right hemisphere activity being a more permanent form of compensation (Crosson et al., 2007), as right hemisphere pathways appear to still be playing a role in language production. These correlations may also be an indication that compensation between the ventral and dorsal pathways are possible, as right hemisphere ventral tracts, which would not be expected to play a role in language production at all, are related to processes like repetition, which would ordinarily be considered to be a dorsal pathway process (Ueno et al., 2011).

Finally, left FAT connectivity was also found to be associated with several different language processing scores. While damage to this tract has been correlated with verbal fluency in PPA (Catani et al., 2013), these results indicate a similar relationship in post-stroke aphasia, supporting previous findings (Basilakos et al., 2014) and claims that this tract plays a role in language processing. It would appear from our results that the FAT may support semantic processing in particular, with positive correlations between its connectivity and AAT Comprehension, as well as RWT semantic fluency scores. This may explain its positive correlations with both more general naming scores - pure and interfered naming. In fact, these are some of the most positive correlations found in the study, suggesting the FAT may support semantic processing much more than previously believed. However, it also exhibited a postive correlation with RWT Phonological fluency scores, and so a role for it in phonological processing should also be examined in greater detail in future research.

Conclusions

In conclusion, these findings show some support for a dual pathway model of language processing (Saur et al., 2008; Ueno et al., 2011; Yang et al., 2017). They have also allowed us to explore some of the questions produced by this model, although further, more detailed analyses are needed to make assertions regarding these questions which are more than tentative. The findings also offer some support for previous findings regarding the role of the FAT, fornix and cotricospinal tract in language (Basilakos et al., 2014; Catani et al., 2013; Kemerdere et al., 2016, Solca et al., 2013).

Chapter 4: A comparison of different approaches to behavioural treatment of chronic post-stroke aphasia: methodological considerations

Christopher Heath[a]*, Paul Conroy[a], Andrew Stewart[a], Stefanie Bruehl[a-d]

[a]School of Biological Sciences, University of Manchester, UK; [b]St Mauritius Rehabilitation Centre, Meerbusch, Germany; [c]Clinical and Cognitive Neurosciences, Department of Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany; [d]Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine University, Duesseldorf, Germany

*Corresponding author:

Christopher Heath

School of Biological Sciences

University of Manchester

Oxford Road Manchester M13 9PL, UK

christopher.heath-2@postgrad.manchester.ac.uk

Statement of Contribution

Christopher Heath collected the majority of behavioural data in this and the subsequent chapter. Some behavioural data and neuroimaging data were collected by members of the PLORAS and NARU research groups in the course of addition of participants to their respective databases. Neuroimaging data preprocessing was performed by members of the PLORAS research group. Data analysis was performed by Christopher Heath. Verification of behavioural analyses was performed by Andrew Stewart. Paul Conroy, Stefanie Bruehl and Andrew Stewart provided guidance and support throughout the project and repeated feedback on drafts.

Abstract

This article details a novel methodology used to allow direct comparison between six different treatments for chronic post-stroke aphasia and their effects on naming of trained items, untrained items, language profile and functional language and communication.

Background: Treatment of lexical deficits in chronic post-stroke aphasia includes several different theoretical approaches (Crinion & Leff, 2015). Although it is well-known that many approaches to treatment are effective (Abel et al, 2015; van de Sandt-Koenderman et al, 2016), there has been limited comparison between them, so it remains unclear which are most effective for which people with aphasia (PWA). As treatment outcomes often vary greatly between individual PWA (Tippet, 2015), determining which approaches are most effective for which lesion and language profile is an important step towards providing an optimal treatment approach for each individual PWA.

Aims: The study aimed to directly compare the effects of six different treatments of chronic post-stroke aphasia. There were a total of 12 research questions, spanning three broader areas: effectiveness of the different treatments in improving trained item naming, how this effectiveness varied as a result of inter-participant factors, and generalisation of treatment effects beyond the treated items.

Methods & Procedures: 18 participants in the chronic phase of post-stroke aphasia were recruited. Participants were right-handed, spoke English as a first language, and aged between 18 and 80, as per the inclusion criteria. Participants underwent 3 separate prompt-based treatment approaches, each of which consisted of 2 treatments which share a similar theoretical approach to facilitating recovery. Each approach lasted 3 weeks, with a week break between each approach. The approaches applied: Model oriented prompts (split into treatments of semantic versus phonological aspects of language production), alternative modality prompts (melody versus gesture treatments), and executive demand (speeded naming versus naming in the presence of a distractor, i.e. interference naming). Treatments were applied using powerpoint presentations featuring recorded audio or visual prompts, so could include a mixture of in-person and teletherapy. A battery of language testing was carried out before and after the therapy regimen, and interim assessments following each approach were used to determine the effects of each treatment on language abilities.

Keywords: Aphasia; Treatment; Neuroimaging; Stroke; Recovery; Therapy

Introduction

Aphasia is a language disorder which can be acquired in several different ways (Gorno-Tempini et al., 2011; Glosser & Deser, 1991; Patterson et al., 2006), one of the most common of which is following damage to the language regions as a result of stroke. It may affect up to 250,000 people across the UK (Gerenmayeh et al., 2014), and has been found to have an effect on quality of life comparable to life-threatening diseases such as cancer (Lam & Wodchis, 2010). In the chronic phase (over six months post-stroke), there is little evidence that spontaneous recovery continues to have a significant effect, and recovery of language abilities for people with aphasia (PWA) is most commonly facilitated via behavioural treatment in the form of Speech and Language Therapy (SLT).

Literature in recent years has explored a wide range of different approaches to the behavioural treatment of post-stroke aphasia. There are many disparate behavioural treatments for aphasia, including Intensive Language Action Therapy (ILAT) and related therapies (Mohr et al., 2016; Nenert et al., 2017), audio-visual speech entrainment (Fridriksson et al., 2012), and behavioural treatment in combination with neurostimulation (Hara et al., 2017; Hara et al., 2015). However, most can be categorised by the broad approach they take to facilitating recovery. Currently, the optimal way to facilitate recovery from aphasia is unclear, both generally and in regards to specific PWA (Heath et al., in preparation). Therefore, several different approaches to facilitating recovery have developed, each using a different strategy based on existing research and theories regarding aphasia recovery. While there are a number of these different approaches, a large number of treatments can be categorised into one of three approaches which use a neuroscientific rationale as their basis.

One of the most researched of these is the model-oriented approach. This focuses on targeting specific areas of language ability for improvement. Language processing, in its totality, involves a network of different cognitive mechanisms working together to retrieve, combine and physically produce its semantic and phonological components (Ueno et al., 2011). Impairments in one of these left hemisphere-based processes can cause word finding difficulties (Howard & Gatehouse, 2006; Schwartz et al., 2006). Targeting specific components of this network which may be affected, and so reducing the language deficits found in PWA, can therefore be effective in facilitating improvement (Abel et al., 2015; Abel et al., 2016). Treatments using this approach generally target either the semantic or phonological components of language, or both separately, either via mechanisms like cuing hierarchies, in which different levels of semantic or phonological information are provided to aid the PWA in naming a picture (Abel et al., 2015; Abel et al., 2016), or via exploration and encouraging the production of semantic or phonological information regarding a specific word or object, as in Semantic Feature Analysis (SFA) (Coelho et al., 2000) or Phonological Components Analysis (PCA) (Leonard et al., 2015). These cues might include closure sentences (sentences ending in the target word) or a definition for semantic information, or the first phoneme or syllable of the word for phonological information (Abel et al., 2015). Overall, this approach has been shown to be effective across numerous studies. Abel et al. (2014, 2015) used separate phonological and semantic cueing hierarchies in 14 PWA, with both semantic and phonologically-trained item sets improving significantly more than the control set. Leonard et al. (2015) compared two variations of PCA in five PWA, all of whom improved significantly in trained item naming. Significant item naming improvements were also found for both PCA and SFA in a study performing a total of 12 sessions of PCA and SFA in eight PWA. Several different varieties of model-oriented treatments have therefore been shown to improve item naming in previous research. However, these studies did not perform comparisons between the model-oriented approach and other specific approaches to treatment.

Conversely to the model-oriented approach, some treatments could be categorised by their focus on supra-segmental features of communication (such as stress, rhythm and prosody) using methods such as music or gesture. Generally, this alternative modality approach is taken with the aim of increasing activation in the right hemisphere homologues of damaged left hemisphere language regions, with the goal of encouraging these areas to compensate for the damage to their left hemisphere homologues with increased activation in the right hemisphere during language processes (Maher & Raymer, 2004; Cocquyt et al., 2017; Gili et al., 2017). It is therefore juxtaposed with the model-oriented approach, which purports to target core language components exclusively, and so is much more focused on activation in the preserved left hemisphere language regions. Treatments that fall into this category include Melodic Intonation Therapy (MIT), which focuses on using melody and rhythm to learn phrases (Norton et al., 2009; van de Sandt-Koenderman et al., 2016). In this treatment, specific phrases are generally ‘intoned’ (spoken rhythmically with syllables varying between different pitches) by the therapist and PWA together, while tapping the rhythm with their left hand. The therapist then tries to gradually reduce their input with the aim of allowing the PWA to eventually produce the intoned phrases independently (van de Sandt-Koenderman et al., 2016). Intention Therapy, which uses specific hand movements in an attempt to facilitate right hemisphere activation (Crosson et al., 2005; Crosson et al., 2009), and treatments involving object use, pantomime and gesture, which focus on providing information about an item or situation via this modality (Kroenke et al., 2013; Pierce et al., 2017). These treatments generally focus on assisting language production by using actions (either involving the actual objects or a pantomime) to provide additional contextual and semantic information. For example, Kroenke et al (2013) compared participant learning of pseudowords for manipulable objects either through presenting the object picture along with either only the pseudoword aurally, or with the pseudoword aurally and a corresponding gesture providing information on use of the pseudoword (which participants were encouraged to reproduce. Similarly to the model-oriented approach, multiple studies have demonstrated the efficacy of alternative modality treatments in improving measures of language production. Gili et al. (2017) found significant improvement in 10 PWAs' language production abilities in both a pantomime observation therapy and an action observation therapy, while Wan et al. (2014) found significant improvement in production of Correct Information Units (CIUs) per minute in 11 PWA after 75 sessions of MIT therapy. Finally, Peck et al. (2004) also found significant improvement in naming of trained and untrained items in three PWA following intention therapy. However, again, none of these studies directly compared alternative modality approach treatments with any others, instead mostly performing comparisons to a control condition where comparisons occurred.

A less well researched approach currently is executive demand. This approach usually involves picture naming with additional constraints which place demand on the PWA's executive functioning. For example, speeded naming therapy (Conroy et al., 2018; Hodgson & Lambon Ralph, 2008) usually involves picture naming to a deadline, which gets progressively faster, 'speeding up' the PWA's response time and encouraging faster/ more effective executive functioning. PWA are presented with a picture and try to produce the target word before a beep coming a set time after item presentation (for example, 3 seconds). This beep will speed up over the course of treatment at set intervals, occurring sooner after item presentation, with the aim of increasing speed of the PWA's production of the target word (Conroy et al, 2018). Another treatment which could be categorised as using an executive demand approach is interfered naming, in which the PWA names pictures in the presence of a nearly simultaneous spoken distracting word (which may be related to the target word or randomly selected) with the aim of facilitating recovery of linguistic-executive abilities

(Abel & Willmes, 2016; Bruehl, Willmes and Binkofski, 2021a, 2021b). This also places a large amount of demand on their executive functioning in order to name with additional constraints. The approach that considers increasing demands often involves the goal of improving overall connected speech (Conroy et al., 2018), as factors like speed in language production and the ability to ignore or disregard distracting words or objects are often seen as important in improving connected speech ability as opposed to naming ability only. Research regarding the efficacy of this approach is currently limited. Conroy et al. (2018) found significant improvements in use of items trained using a speeded approach in connected speech, while Bruehl, Willmes and BInkofski (2021a) found significantly improved pure and interefered naming scores in PWA following an interfered naming treatment. While these early findings are promising, they again offer little in regards to comparisons with other approaches to treatment.

While it is clear that a wide variety of approaches, and even more individual treatments, are available to aid PWA in recovery of language abilities, even during the chronic phase of the disorder, some gaps in the existing research prevent a clear direction forwards for treatment in terms of which approaches to pursue in greater depth. A review of recent literature on the subject by Heath et al (in preparation) found that, while many treatments had been studied either in comparison to a control treatment, to 'standard' SLT, or to no treatment, the literature comparing different treatments was very limited. Without a standardised method of comparison between different treatments, knowing which treatments and approaches are most effective, and for which language and lesion profiles, is extremely difficult. In this context, optimisation of treatment based on PWA language and lesion profile, or even knowing the most effective treatments overall, cannot be achieved effectively. Simply knowing which treatments are effective in comparison to controls does not provide as much information regarding comparative effectiveness as a direct comparison of treatments would. Direct comparison of treatments, while controlling for extraneous factors such as intervention volume, intensity and SLT input, could provide a large amount of reliable information regarding comparative effectiveness. Taking the key elements from each treatment and integrating them into a standardised treatment structure based around improving an objective measure such as item naming could allow for control of these extraneous variables. It could also still allow comparison of the key elements employed by each treatment. Gaining information on comparative success of these treatments, which the above would aim to achieve, would provide invaluable steps towards understanding how treatments compare in facilitation of language production improvement generally, as well as within more specific PWA lesion and language profiles. The eventual aim of this type of direct comparison would therefore be an advancement in providing each PWA with the treatment most optimal for facilitation of their language improvement.

This study therefore aimed to address shortcomings in the current research literature regarding behavioural treatment of chronic post-stroke aphasia. A direct comparison of multiple approaches, as well as multiple treatments for each approach, was planned in order to gain novel information regarding the comparative effectiveness of each treatment, and how this effectiveness varied based on PWA language and lesion profiles. In order to provide this direct comparison, the treatments were also required to be standardised in terms of following a similar structure and using the same time allowances. It is hoped that this will aid selection of treatments that are time effective for PWA and SLTs providing treatment. It is also hoped that the information gained in this study is a first step towards a more intricate understanding of treatment and approach effectiveness, and towards optimisation of treatment for individual PWA based on their specific circumstances and requirements.

This study aimed to investigate the effects of 3 different, broad approaches to therapy for aphasia, each of which was split into 2 treatments. The key research questions this study aimed to explore were split into three broad categories:

(1) Treatment Effectiveness

- Did treatments show significant therapy effects in naming accuracy for trained items?
- Which, if any, treatments showed stronger therapy effects in naming accuracy than another?
- Were therapy effects maintained through to follow-up?
- Were therapy effects affected by the order in which treatments were presented?

(2) Inter-participant variation

- Did inter-participant demographic factors modulate the effectiveness of treatment?
- Did inter-participant language factors modulate the effectiveness of treatment?
- Did inter-participant variations in lesion volume and location modulate the effectiveness of treatment?

· Did lesion location modulate the effectiveness of different treatments?

(3) Generalisation beyond treated items

· Did significant therapy effects in naming accuracy generalise to untrained words?

· Did significant therapy effects in naming accuracy generalise to measures of semantic or phonological processing (e.g., changes in error pattens)?

· Did significant therapy effects in naming accuracy generalise to wider measures of language and cognitive functions (e.g., comprehension, executive skills)?

· Did significant therapy effects in naming accuracy generalise to functional language and communication measures?

Considering these research questions, our hypotheses mainly focused on examining the effects of each approach and treatment. We hypothesised that there would be variation in effectiveness between different approaches and treatments, and that different approaches and treatments would display differing levels of effectiveness depending on PWA lesion and language profile. Broadly, we expected the alternative modality, right-hemisphere focused approach to show greater effectiveness for PWA with larger left hemisphere lesions and more severe aphasia than those with mild deficits, based on previous suggestions that increased right hemisphere activation may be associated with more improvement in those PWA with severe deficits (Abel et al., 2014; Crosson et al., 2007) and more left hemisphere damage (Crosson et al., 2007; Heiss & Thiel, 2006). Based on the same theories, we expected the model-oriented approach focusing on aspects of language (which is likely to facilitate activation in left hemisphere language regions rather than the right hemisphere (Maher & Raymer, 2004; Howard & Gatehouse, 2006; Schwartz et al., 2006) to be most effective when used with PWA with mild-moderate aphasia severity rather than those with severe aphasia. Finally, we expected executive demand approaches to have greatest effectiveness with PWA with mild aphasia and limited lesions as opposed to more severe aphasia, due to the heavy demand it places on executive function (Conroy et al., 2018); it was expected that those PWA who could handle these demands as well as the demands placed on remaining language networks the best would gain the most benefit from this approach. As research directly comparing approaches is extremely sparse, we hypothesise as to if any treatments or approaches might prove to be more effective overall.

We therefore predicted that improvements in picture naming scores on trained items from immediately before to immediately after each approach would vary between approaches for each PWA. We predicted that improvements in naming pictures trained in the alternative modality approach would be greatest for those PWA with the lowest initial picture naming scores and most extensive lesions in the left hemisphere. We predicted that improvements in naming pictures trained in the model-oriented approach would be greatest for those PWA with moderate-high initial picture naming scores and limited damage to left hemisphere language regions, while improvements in naming pictures trained in the executive demand approaches would be greatest for those participants with high initial naming scores and low lesion volume.

In relation to our research questions, our hypotheses were as follows:

(1) Treatment Effectiveness

- All treatments would show significant effects in naming accuracy for trained items.
- No treatments would show overall stronger effects in naming accuracy than the others; however, there would be variation in individual responses to each treatment.
- Therapy effects would be maintained through to followup.
- Therapy effects would not be significantly affected by the order of treatment presentation.

(2) Inter-participant variation

- Inter-participant demographic factors, language factors, lesion volume and location would all modulate effectiveness of treatment.
- Lesion location would modulate the effectiveness of different treatments.

(3) Generalisation beyond treated items

- Significant therapy effects in naming accuracy would generalise to untrained words, measures of semantic and phonological processing, wider measures of language and cognitive functions, and functional language and communication measures. However, these generalisations would all be relatively minor and improvement would not reach significant levels for the majority of measures.

Methods

The study aimed to compare the effects six different treatments had, across a range of aphasia subtypes and levels of severity, on single-word picture naming. Several measures of language, ranging from single-word picture naming ability, to sentence-level utterances, to extended connected speech, were also assessed to gain a greater understanding of change in language ability over the course of the study. Neuroimaging data was also used to evaluate if the extent and location of damage in the brain and disruption of white matter tracts was related to variable responses to the different approaches. The study was approved by a University of Manchester ethics committee.

Therapy approaches and treatments

All PWA underwent three separate prompt-based approaches: model-oriented, alternative modality, and executive demand, each of which consists of two treatments which share a similar theoretical approach to facilitating recovery. They also underwent behavioural testing before and after each approach. Each prompt-based approach lasted three weeks and included two 75-minute therapy sessions each week. These sessions were split into two segments, one training each treatment, with up to a 10-minute break between them to limit PWA's fatigue, and the order of treatment segments was alternated each session.

In-person sessions varied in frequency between once, twice, or three times per approach, depending on preference of PWA (with a minimum default of once per approach). The remaining sessions were performed using teletherapy, via telephone or videoconferencing software. PWA were also asked to perform non-directed home training in addition to this, at a minimum (and default) of one hour a week, but increasing up to a maximum of five hours per week if PWA wished to do additional training. However, PWA were encouraged to keep their volume of home training consistent across treatments. Amount of time for home training in each treatment was automatically registered in cases where this option was supported by the device and software used for treatment by the PWA. This allowed PWA to self-regulate training both treatments in each approach equally. There were breaks between each approach, with a default length of one week but increasing to a maximum of three weeks depending on PWA scheduling and preference. Frequency of in-person therapy sessions, amount of undirected home training, and length of break between therapies were kept as consistent as possible within PWA. The order of item presentation within each treatment was randomised every other time the item set was presented. All training was performed on a laptop using presentations in which each slide contained a given prompt level for each item (so training for each item was composed of seven slides, one for each prompt level). Each slide contained an image of the target item and a recorded audio or video (in the case of the gesture treatment) prompt, which would play at the relevant time for the given treatment and progression level. This allowed a very standardised approach to treatment, with the researcher monitoring and therefore being able to direct the treatment and ensure it was being performed correctly.

Stratified group allocation was performed based on baseline naming performance, as well as syndrome categorisation (fluent/non-fluent) if possible. The order of approaches for these groups, and the overall structure of the study, can be found in Figure 4.2.

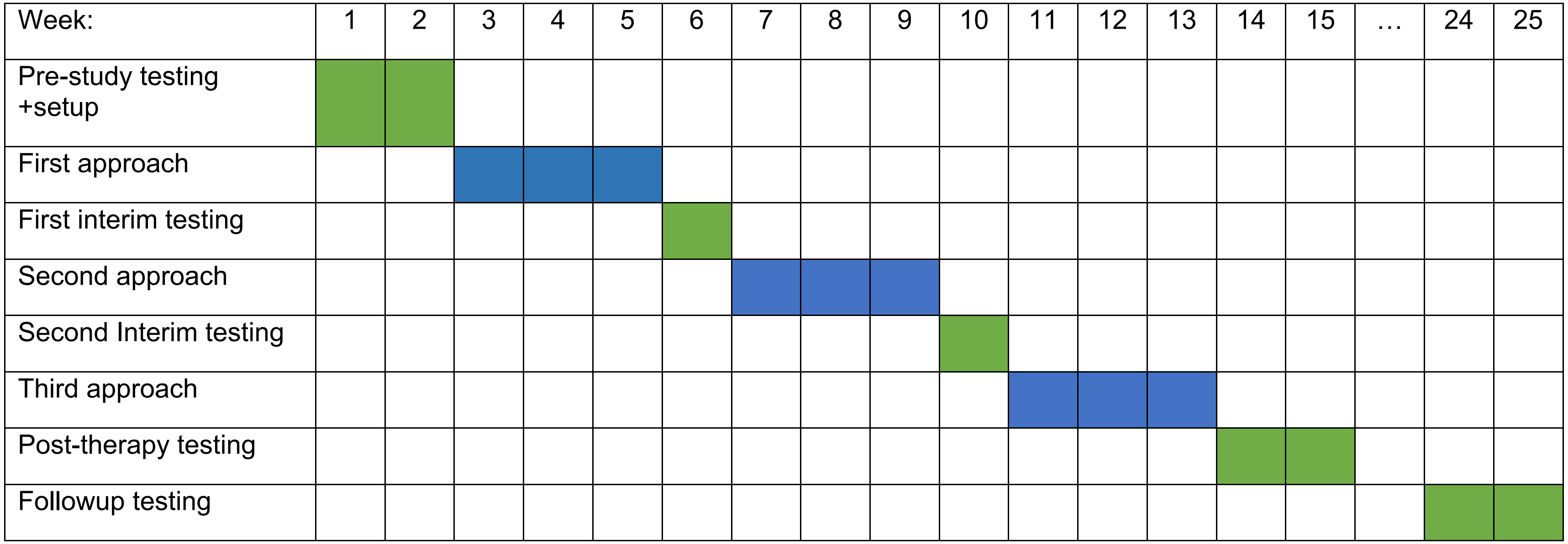

Figure 4.1. Gantt chart to show overall timeline of study for each participant week by week.

	Pre-therapy	Block 1	Interim	Block 2	Interim	Block 3	Post-therapy
Group 1	Baseline testing	Model-oriented	Post-Block 1 assessment	Alternative Modality	Post-Block 2 assessment	Executive Demand	Post-Block 3 assessment
Group 2	Baseline testing	Alternative Modality	Post-Block 1 assessment	Executive Demand	Post-Block 2 assessment	Model-oriented	Post-Block 3 assessment
Group 3	Baseline testing	Executive Demand	Post-Block 1 assessment	Model-oriented	Post-Block 2 assessment	Alternative Modality	Post-Block 3 assessment

Figure 4.2. Study structure and therapies for PWA order groups 1, 2, and 3.

Details of approaches

Model-oriented

The model-oriented approach targets specific stages or aspects of processing. For word-finding difficulties, this usually involves targeting the semantic and phonological aspects of language processing. In this approach, we did so using prompts focused on each separately, and so was split into semantic and phonological treatments. The semantic treatment used prompts which provide semantic information about the pictured item to facilitate PWA's word retrieval. This included superordinates- the semantic category to which the target belongs, coordinates- other items in the semantic category, definitions of the target item, and closure sentences- sentences ending with the target item, which provide semantic information and context. The phonological treatment used prompts providing phonological information about the target word to help PWA produce it. These included the first syllable or onset (first phoneme) of the target item, the number of syllables, or a rhyming nonsense word providing the final syllable of the target item. The initial decreasing prompt order of these treatments are shown in Table 3.1. This order would be reversed (e.g., moving from prompt 7 to prompt 1) in the case of a PWA moving to an increasing prompt hierarchy, and selected by the PWA for a self-selected prompt structure. However, the prompts themselves would remain consistent throughout.

Table 4.1. Model-oriented approach prompt levels and examples for the target item "Cat".

Model-oriented approach			
Semantic Treatment		**Phonological Treatment**	
Prompt level	**Prompt example**	**Prompt level**	**Prompt example**
1: Spontaneously		1: Spontaneously	
2: Auditory target comprehension	"That's a cat."	2: Overt repetition	"That's a cat."
3: Closure sentence	"On the sofa there purrs a?"	3: First syllable	"That's a /ca/."
4: Definition	"That's an animal which catches mice."	4: Onset	"That's a /c/."
5: Superordinate	"That's an animal."	5: Number of syllables	"That's a (1x knocking)."
6: Coordinate	"That's like a dog."	6: Rhyme	"That rhymes with yat."
7: Spontaneously		7: Spontaneously	

Alternative Modality

The alternative modality approach is focused on prompts which involve using supra-segmental features of communication to facilitate production of the target items. The prosody treatment does this via melody and rhythm in the form of word intonation (a form of speech which uses consistent rhythm and pitch intervals between syllables to emphasise the natural cadence of utterances), rhythmic humming and rhythmic tapping. At the level it provides the most assistance in producing the target item, therapist and PWA intone the word together. Helpfulness is decreased by reducing therapist involvement in PWA intonation and spoken production of the target word, and moving from intonation to rhythmic, melodic humming of the emphasised cadence of the phrase. Intonation always alternated between two pitches, a minor third apart. The higher pitch was used for the emphasized syllables of each intoned sentence, and the 'tune' was kept the same for each sentence across both intonation and humming. Each sentence was structured as either "That's a…(target item)" or an appropriate variation (e.g. "Those are…(target item)". Participants began with intonation of the sentence in unison with the researcher, then repeated the intonation. At prompt level 4 they were required to name the item after listening to the researcher intone the relevant sentence. They were then required to name the item after repeating the hummed tune of the relevant sentence, then, for the final prompt level, after listening to this hummed tune.

The gesture treatment uses gestures to provide the PWA with additional information about the size, shape, use and context of the target item. The complexity and amount of information conveyed by the gesture, as well as PWA involvement in the gesture (whether they copy it directly or just observe it) decreases to decrease prompt helpfulness. All gestures aimed to convey semantic information regarding the target item. In most cases, the gesture therefore involved use of the target item (e.g. smoking a cigarette), but in some cases involved information about an object's size and shape (for example, using the arm to show the curved slope of a slide). Gestures were split into longer complex gestures for prompt levels 3 and 4, which provided more information and context, and shorter simple gestures for prompt levels 5 and 6, which provided less information and context and were usually shorter versions of the complex gestures with certain elements excluded (e.g. the complex gesture for 'cigarette' involved both lighting and smoking a cigarette, while the simple gesture only included the smoking element, excluding the 'lighting' part of the gesture).

Again, the prompts for these approaches were carried out in a descending order (from prompt level one to seven), and participants performing well on this order could progress to increasing, and finally self-selected, prompt order.Table 4.2. Alternative modality approach prompt levels and examples for the target item “Cat”.

Alternative Modality approach			
Prosody Treatment		**Gesture Treatment**	
Prompt level	**Prompt example**	**Prompt level**	**Prompt example**
1: Spontaneously		1: Spontaneously	
2: Produce word after intonation in union.	"Sing the word with me. *intoned* That's a cat. Ready? That's a *unison intoned* a cat"	2: Overt repetition	"That's a cat.”
3: Intonation (repeated) with left hand tapping	Please copy the melody and rhythm I use when speaking as closely as possible when repeating. *intoned* That's a cat."	3: Performance of complex motor action	"Please copy my gesture while trying to name the item. That's a *perform complex gesture*.”
4: Intonation (observed) with left hand tapping	" *Intoned* That's a cat."	4: Observation of complex motor action	"That's a *perform complex gesture*.”
5: Humming (repeated) with left hand tapping	Please copy the melody and rhythm I use when humming as closely as possible when repeating. *hum*."	5: Performance of basic motor action	"Please copy my gesture while trying to name the item. That's a *perform basic gesture*.”
6: Humming (observed) with left hand tapping	" *hum*.”	6: Observation of basic motor action	"That's a *perform basic gesture*.”
7: Spontaneously		7: Spontaneously	

Executive Demand

The executive demand approach focuses on the wider cognitive processes underpinning language production, specifically information processing and executive control. The speeded treatment aimed to do this by focusing on both speed and accuracy of naming, both of which are required for effective connected speech (Conroy et al., 2018). A naming-to-deadline paradigm (Hodgson & Lambon Ralph, 2008) was used to do this, where following the prompt, the PWA aimed to name the item before a beep. Over the course of the therapy, the time between the prompt and the deadline beep decreased in several stages, providing continual speed pressure for the PWA. The interfered treatment aimed to do this by presenting distractors at certain times relative to picture presentation (Abel & Willmes, 2016). These distractors could share a phonological or semantic component with the target word, or be a random word. This treatment is based on the finding that, depending on time of presentation relative to the picture, distractors can facilitate or inhibit successful picture naming (Abel et al., 2009).

The first week of treatment for both speeded and interfered treatments focused on priming of the items and naming with prompts. This aimed to ensure a basic level of comprehension and familiarity with the items to ensure the later weeks with more executive demand can be completed. Priming involved identifying the correct item picture from a selection of four, after the item had been named by the researcher. This selection included a semantic, phonological and random distractor along with the target item. This was done once each session for each trained item. After priming was completed, the remainder of the sessions in the first week are spent on confrontation naming using a short increasing prompt hierarchy (see Table 4.3). In the speeded condition, a three second deadline was also applied to this naming before prompt exposure to allow PWA to adjust to speeding without a large executive demand. Practically, this was performed with a beep which sounded three seconds after item presentation on each slide, which participants aimed to 'beat' by naming the item before it sounded.

In weeks two and three, the Interfered treatment focused on executive function by performing naming in the presence of a distractor word which is related to the target name. Distractors may be phonological (e.g., for target word 'cat', distractor is 'bat'), associative situational (e.g., for target word 'cat', distractor is 'mouse'), associative part-whole (e.g., for target word 'cat', distractor is 'whiskers'), or unrelated (no relationship). Distractors were presented 200ms before picture exposure. While the items do not undergo a prompt hierarchy on every presentation, the increasing prompt hierarchy shown in Table 4.3 is used as assistance when PWA are otherwise unable to correctly name the item.

In weeks two and three, the speeded treatment targets executive function by inclusion of naming to a deadline. In week two, PWA attempted to name the target before a beep occurring two seconds after picture presentation. In week three, this beep reduced to one second after picture presentation. Speeded naming, therefore, attempted to gradually increase PWA speed of naming, starting with a fairly comfortable three second deadline in the first week, and increasing speed to eventually match that of elderly neurotypical PWA (Conroy et al, 2018). During weeks two and three, as in the interfered naming condition, an increasing prompt hierarchy was also used for assistance only for items PWA were unable to correctly name, up to the point that they were correctly able to name the object (Table 4.3).

Table 4.3. Executive demand approach prompt levels and examples for the target item "Cat".

Executive Demand Approach			
Speeded Treatment		**Interfered Treatment**	
Prompt level	**Prompt example**	**Prompt level**	**Prompt example**
1: Spontaneously (with deadline)		1: Spontaneously (with distractor)	
2: First syllable	"That's a /ca/."	2: First syllable	"That's a /ca/."
3: Closure sentence	"On the sofa there purrs a?"	3: Closure sentence	"On the sofa there purrs a?"
4: Overt repetition	"That's a cat."	4: Overt repetition	"That's a cat."

Stimuli

Each approach was trained using a subset of the 280-item picture set used for picture-naming assessment. This 280-item set was split into seven sets of 40 items: two sets for each approach (one for each treatment), plus one control set. The sets were matched on frequency and word length as measured by number of syllables using the Match program (van Casteren & Davis, 2007).

For each PWA, 30 items were selected from each of these sets to form performance-adapted sets, totalling 210 items (180 trained items, in addition to 30 control items). These items were selected based on difficulty, which was determined by combining the frequency and length in syllables for each PWA based on their initial naming score (e.g. PWA with low naming scores were allocated an easier set of words- PWA with an initial average score of below 94 were given the easiest set, those between 94 and 188 used the medium set, and those with a score of 188 or above used the hardest set) to help reduce potential floor or ceiling effects.

All items were presented on a laptop, with recorded prompts being played on each presentation of each item. All PWA initially followed a decreasing prompt hierarchy for the model-oriented and alternative modality approaches, attempting to name each picture, first spontaneously with no prompt, then once following each of five prompts, presented in decreasing order of helpfulness. They then attempted to name the picture again with no prompt assistance. The prompt hierarchy continued regardless of PWA ability to produce the word to both ensure an equal number of prompts were performed in each hierarchy and create a strong association between the word and information provided in the prompts. Depending on PWA progress and confidence, they progressed to an increasing prompt hierarchy in the second or third week, or remained using the decreasing prompt hierarchy for all three weeks of each approach.

In the increasing prompt hierarchy, the PWA attempted to name each item with the help of the same five prompts (as well as spontaneously before and after these prompts), but the prompts are presented in the opposite order, from least to most helpful. Depending on PWA progress and confidence, they progressed again to a PWA-selected prompt hierarchy in the third week, or remained using the increasing prompt hierarchy for a given week.

In the PWA-selected prompt hierarchy, PWA self-selected their preferred order of prompts using cue cards. They then attempted to name a suitable prompt for that level of helpfulness (a prompt was only presented to them if they were unable to do this), before attempting to name the target item. This methodology aimed to build stronger connections between the target words and the information or modality which can facilitate their production, as in Semantic Feature Analysis (Boyle & Coelho, 1995) and Phonological Components Analysis (Leonard, Rochon, & Laird, 2008).

For the executive demand approach, a shorter increasing prompt hierarchy was used. In the first week this hierarchy followed the same methodology as the prompts for the other approaches (hierarchy continues regardless of PWA ability to produce the word) in order to promote familiarity with the word. However, in weeks two and three, the prompt hierarchy was only used when PWA were unable to successfully produce the word spontaneously, as the focus was placed on the speeded and interfered elements for this approach. The possibilities for participant progression over each 3-week approach are summarised in Figure 4.3.

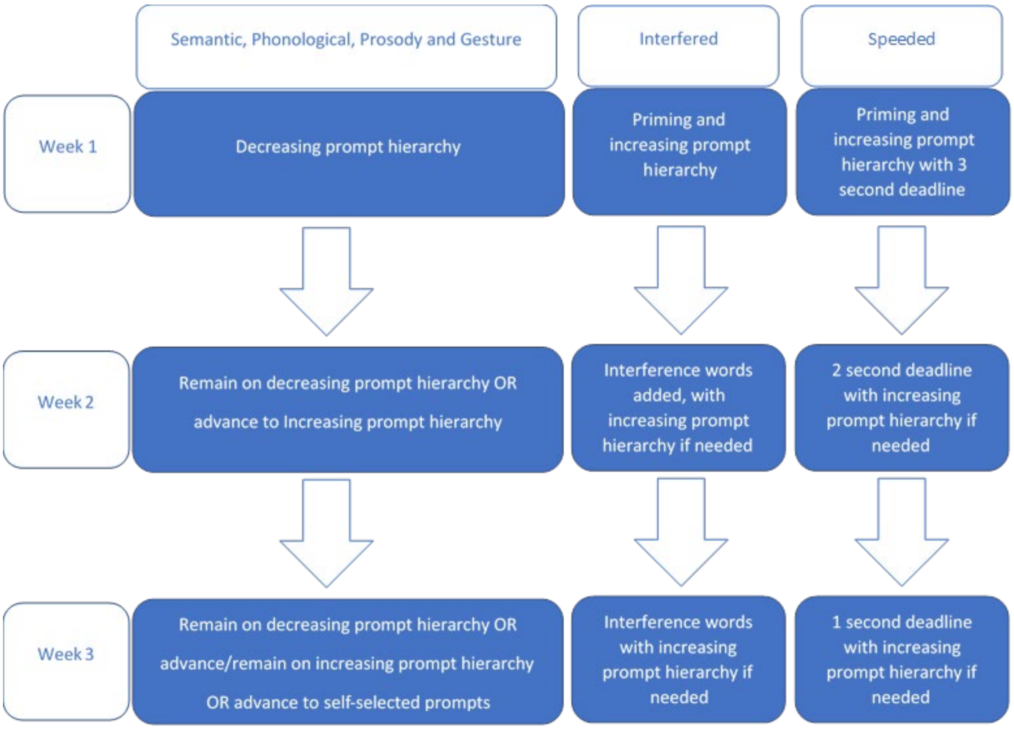

Figure 4.3. Progression of each treatment week by week. Decreasing prompt hierarchy: Beginning on the easiest prompt and progressing to the hardest. Increasing prompt hierarchy: Beginning on the hardest prompt and progressing to the easiest. Self-selected prompts: Participant-selected prompts and/or order of prompt exposure. See text for further details.

Assessments

Testing was carried out at several different time points. Firstly, initial testing was used to ensure that PWA fit the criteria for the study. Baseline testing was then carried out before therapy began, with a follow-up block of post-therapy testing after completion of the final block of therapy. Less extensive testing was also carried out in the breaks between each block (Post-Block 1 and 2 assessments in Figure 4.2). All assessments were conducted in person by the researcher.

Initial testing

Several initial tests were carried out before PWA inclusion in the study to ensure that they are suitable to participate. The Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983; Goodglass, Kaplan, & Barresi, 2000), BNT, PALPA 9 Repetition subtest, and CBU Category Comprehension Test were carried out to ensure that PWA's language abilities fit the criteria for the study and provide a basic profile of their deficit.

280-item picture naming

Pre-therapy baseline testing included confrontation naming testing with a 280-item set of black and white pictures of common objects, which included all of the items that would be trained during the treatments for each PWA. The 280-item picture set was performed twice at baseline, with at least a week between assessments, in order to determine a baseline score typical of PWA's usual abilities. This set was selected from the International Picture-Naming Project (IPNP) database (Szekely et al., 2004). The IPNP database includes 520 drawings of common objects in black-and-white, 174 of which are from the Snodgrass and Vanderwart (1980) set. A 280-item picture set with a variety of lengths and frequencies was selected from this database.

Responses to the items in this category were split into three categories: Incorrect, correct, and correct with uncertainties including self-corrections. Items that did not elicit the exact name or that were named correctly with delay (more than 10 seconds after presentation) were counted as incorrect. Items that were correctly named without delay but after self-corrections and apparent hesitations during word production were counted as correct with uncertainties. Items that were correctly named without delay and uncertainty were counted as correct. Any use of meaningful gestures was also noted, but did not contribute to overall score.

In addition to the item names provided by the IPNP database, additional names were selected as acceptable substitutes due to being common synonyms for or variations on the given items. These synonyms were selected based on pilot control data in which six healthy native English speakers performed confrontation naming with the 280-item picture set. After discarding of clear semantic errors not due to lack of clarity in the selected pictures, the remaining answers provided by the healthy controls were included as acceptable substitute answers. This increased the range of acceptable answers to include commonly used synonyms (e.g., “spade” in addition to the formally correct response of ’Shovel’ for the picture of a shovel), variations (e.g., ”cricket” in addition to the correct response of ’grasshopper’), and some errors caused by lack of clarity in the available pictures rather than any language deficit (e.g., “lime” in addition to the correct response of ‘lemon’- in black and white these items are not clearly distinguishable). Picture naming was always audio recorded.

Other baseline tests

In addition to this, PWA underwent the Amsterdam-Nijmegen Everyday Language Test (ANELT)(Blomert, Kean, Koster, & Schokker, 1994) and Scenario Test (van der Meulen, van de Sandt-Koenderman, Duivenvoorden, & Ribbers, 2010), as well as three composite picture description tasks: Cookie Theft from the BDAE, Park Scene from the Western Aphasia Battery (WAB) (Kersetz, 1982), and Living Room Scene from the Comprehensive Aphasia Test (CAT) (Swinburn, Porter, & Howard, 2005), in order to provide more general measures of communication ability. The remainder of the CAT was also performed. The Mini Mental State Examination (MMSE)(Folstein, Folstein, & McHugh, 1975) was performed before and after the study to ensure PWA cognition was not significantly affected by any neurodegenerative illnesses.

Interim and post-therapy assessments

Interim assessments were carried out between each block of therapy, in the week after completion of the block. These assessments included only testing of confrontation naming via the 280-item picture naming set.

Post-therapy, another round of testing was carried out. This included all the tests performed at baseline- the 280-item picture naming set (performed only once this time), ANELT, Scenario Test, Cookie Test, Park Scene, Living Room Scene, CAT, and MMSE. The BDAE was also performed to capture any broader changes in language ability.

Follow-up testing

A follow-up assessment was performed in order to evaluate the longer-term effects of each approach, featuring all the tests performed at the post-block 3 assessment. T

Within-block assessments

The first and final session of each therapy block was also recorded. This was to allow potential analysis of the immediate effects of each cue on ability to name the target item and speed of naming. This could potentially be used to investigate the helpfulness of each cue in naming, and how this helpfulness changes over the course of each therapy as PWA language ability improves.

Methodology strengths and limitations

This methodology was designed with the aim of allowing direct comparison of multiple disparate approaches to treatment of aphasia. The study design included six treatments, split into three approaches, to allow for direct comparison of the model oriented approach with both the alternative modality and executive function approaches, as well as between the alternative modality and executive function approaches. As discussed earlier, both the model oriented and alternative modality approaches have been demonstrated to be effective in improving language production measures in

PWA. However, previous research has generally only compared each treatment in a vacuum- treatment is compared to a control or to no treatment. It is therefore currently unclear how the treatments perform against each other- which treatment would produce the optimum benefits for any given PWA. This methodology was designed in order to allow this comparison. Performing multiple studies employing a simpler design comparing only two approaches each would not have allowed for this. This design also allowed for more optimal use of an often-limited participant pool, as multiple studies with a simpler design would have taken up more participant time and likely resulted in participant characteristics being more spread more thinly across different studies, resulting in fewer participants in each study as opposed to using all available and willing participants in a single study.

However, this methodology also resulted in limitations to the study which must also be considered. The use of three approaches without a control or dummy block meant that the order groups were not deployed in a fully Latin square design. While the methodology accounted for each approach's position in the order (e.g. each approach was delivered first in one order group, second in another, and third in still another), each approach's individual effect on the next was not (e.g. the alternative modality approach directly followed the model oriented approach in two order groups, while the model oriented approach only followed the alternative modality approach in a single group, and not directly). This meant that order effects were not fully accounted for within the study design, and may subsequently have had an influence on the results of the study. Arguably another limitation of the study methodology was the lack of post-treatment scans. which could have potentially provided evidence for a direct relationship between improvements in language ability as a result of treatment and changes in brain structure. While pre-treatment scans allow for correlative analysis between brain structure and subsequent language improvement, they do not provide any information by themselves on if and how treatment may result in changes in brain structure.

Disclosure of interest

The authors report no conflict of interest.

Chapter 5: A comparison of different approaches to behavioural treatment of chronic post-stroke aphasia: Results and Discussion

Christopher Heath[a]*, Paul Conroy[a], Andrew Stewart[a], Stefanie Bruehl[a-d]

[a]School of Biological Sciences, University of Manchester, UK; [b]St Mauritius Rehabilitation Centre, Meerbusch, Germany; [c]Clinical and Cognitive Neurosciences, Department of Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany; [d]Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich-Heine University, Duesseldorf, Germany

*Corresponding author:

Christopher Heath

School of Biological Sciences

University of Manchester

Oxford Road Manchester M13 9PL, UK

christopher.heath-2@postgrad.manchester.ac.uk

Statement of Contribution

Christopher Heath collected the majority of behavioural data in this and the subsequent chapter. Some behavioural data and neuroimaging data were collected by members of the PLORAS and NARU research groups in the course of addition of participants to their respective databases. Neuroimaging data preprocessing was performed by members of the PLORAS research group. Data analysis was performed by Christopher Heath. Verification of behavioural analyses was performed by Andrew Stewart. Paul Conroy, Stefanie Bruehl and Andrew Stewart provided guidance and support throughout the project and repeated feedback on drafts.

Abstract

This article details the findings of a direct comparison between six different treatments for post-stroke aphasia. These span three different approaches to treatment: Model Oriented (semantic and phonologically focused treatments), Alternative Modality (Prosody and Gesture-based treatments), and Executive Demand (Speeded and Interfered naming). The background and methodological information for this study is covered in Heath et al (in preparation). Cumulative-link mixed models (CLMMs) were used in combination with more basic statistics to compare treatment performance while accounting for a range of inter-participant factors. Results are broadly split into three main categories: Treatment effectiveness, inter-participant variation, and generalisation beyond treated items. In terms of treatment effectiveness, the two executive demand treatments were found to perform significantly better than the model oriented or alternative modality treatments in terms of improving naming of trained items. While effects were found relating to a number of inter-participant factors across differing demographic, language and lesion profiles, Speeded and Interfered naming generally seemed to have the best performance, although the effect of Interfered naming was more resilient to inter-participant factor variation, making it potentially more beneficial in patient cases with poorer prognosis for recovery. Findings for generalisation to other language and cognitive functions were mixed, although there was some evidence to suggest that treatment length is more important in generalisation than previously considered, due to its effect improving generalisation to untrained words.

Methods

Participants

The PWA initially recruited for the study were all in the chronic phase of post-stroke aphasia, and were recruited for aphasia-related research from aphasia speech and language therapy services and support groups across North-West and South-East England over the last several years. The majority of those PWA from the North-West of England were referred from the University of Manchester Neuroscience and Aphasia Research Unit (NARU) database, while those from the South-East of England were referred from the Predicting Language Outcome and Recovery After Stroke (PLORAS) project at University College London. All PWA were right-handed (pre-stroke), native English speakers aged between 18 and 80, who suffered from aphasia following a single left hemisphere stroke at least six months previously. Any PWA with significant age-related cognitive deficits or neurodegenerative disorders which may have an effect on language or cognition were excluded. A range of different severities and subtypes of aphasia were included in the study. In order to ensure that PWA had the minimum picture naming, repetition and comprehension skills necessary to effectively complete the therapy, PWA who scored below 5% on the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983), 25% on the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) 9 Immediate Word Repetition subtest (Kay, Lesser, & Coltheart, 1996), or 50% on the CBU Category Comprehension Test (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000) during a pre-study testing phase were excluded. Upon an initial 280-items picture naming test of the study material, PWA were also excluded if they could correctly name at least 242 of these items spontaneously, in order to avoid a ceiling effect for PWA who were already near mastery level for picture naming (i.e., at least 90% spontaneously correct with a confidence level of 95% according to the binomial model).

Study Design

Study design is fully described in Chapter 4. 20 PWA were tested at 2 timepoints pre-therapy, completing the initial tests in the first stage, then the 280-item picture set and the other baseline tests in the second stage if their initial scores made them eligible. Following each 3-week approach, they had an interim test in their 1-3 week break before the next approach. After all approaches were completed, an immediate set of post-therapy testing was completed. 18 PWA reached post-therapy testing, with the remaining 2 dropping out mid-study. Further follow-up testing was performed for 15 of the 18 PWA, on average two-four months after completion of the study.

For the final three PWA, exceptions to the standard study design were made in terms of number of sessions required to be performed in person. Each underwent one-two approaches and subsequent testing, entirely via videoconferencing software in order to ensure their safety during the COVID-19 pandemic.

PWA were split into three different order groups of five-seven using stratified group allocation, in order to account for order effects. Stratified group allocation was performed based on baseline naming performance, as well as syndrome categorisation (fluent/non-fluent) if possible.

Scans and scan pre-processing

Pre-therapy T1 structural scans were acquired from the NARU and PLORAS participant databases. Several stages of pre-processing were performed in SPM8 (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007) using the ALI toolbox (Seghier et al., 2008) before analyses took place. Scans underwent unified segmentation, spatial smoothing, abnormality detection and lesion definition in order to produce binary lesion images, as well as fuzzy and standard normalised brain images.

Results Section 1: Treatment Effectiveness

Q1.1: Do any treatments show a significant therapy-related change in naming of trained items?

Firstly, it was required to confirm which treatments were effective in facilitating naming improvement. A series of Wilcoxon tests were therefore conducted to determine which treatments were associated with a significant improvement in the naming of the items they were used to train. These tests compared participant naming scores for trained items immediately pre and immediately post each treatment. A significant difference from pre- to post-therapy was found for the trained items of all treatments: Phonological ($p < .001$, mean score increased from 42 pre-therapy to 50.17 post-therapy), Semantic ($p < .001$, mean score increased from 39.72 pre-therapy to 48.89 post-therapy), Gesture ($p < .001$, mean score increased from 41.17 pre-therapy to 49.28 post-therapy), Prosody ($p < .001$, mean score increased from 40.11 pre-therapy to 48.5 post-therapy), Interfered ($p < .001$, mean score increased from 39.11 pre-therapy to 50.89 post-therapy), and Speeded ($p < .001$, mean score increased from 35 pre-therapy to 51.39 post-therapy). Total possible score in every case was 60, with a maximum of 2 points from each of the 30 trained items.

Figure 5.1. Participant scores for items trained using the Phonological treatment pre- and post- therapy.

.

Figure 5.2. Participant scores for items trained using the Semantic treatment pre- and post- therapy.

Figure 5.3. Participant scores for items trained using the Gesture treatment pre- and post- therapy.

Figure 5.4. Participant scores for items trained using the Prosody treatment pre- and post- therapy.

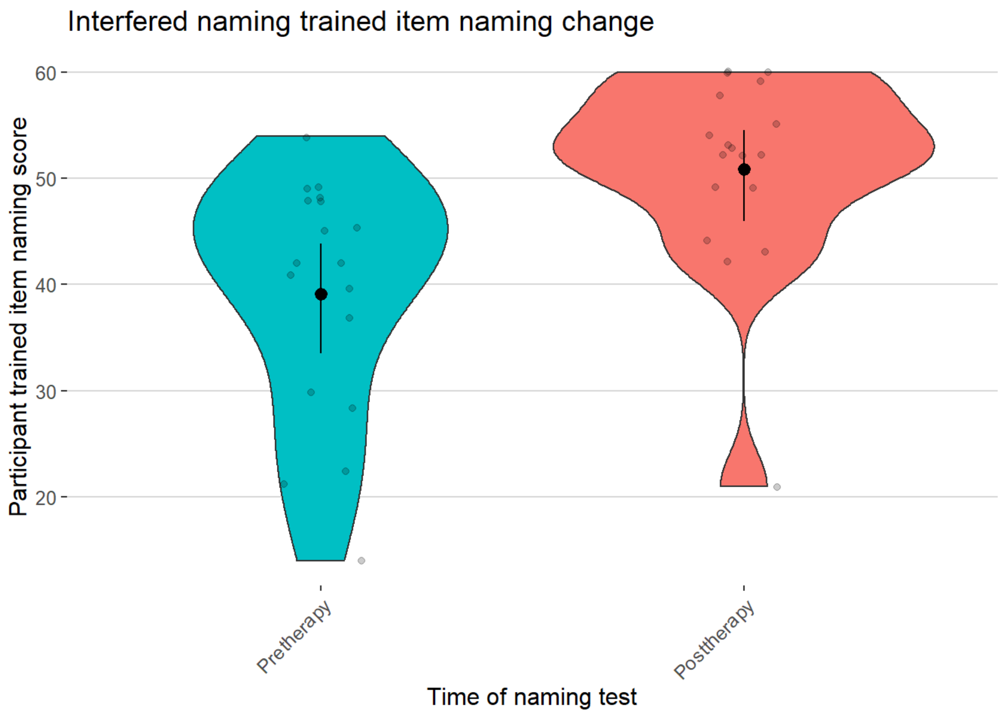

Figure 5.5. Participant scores for items trained using the Interfered treatment pre- and post- therapy

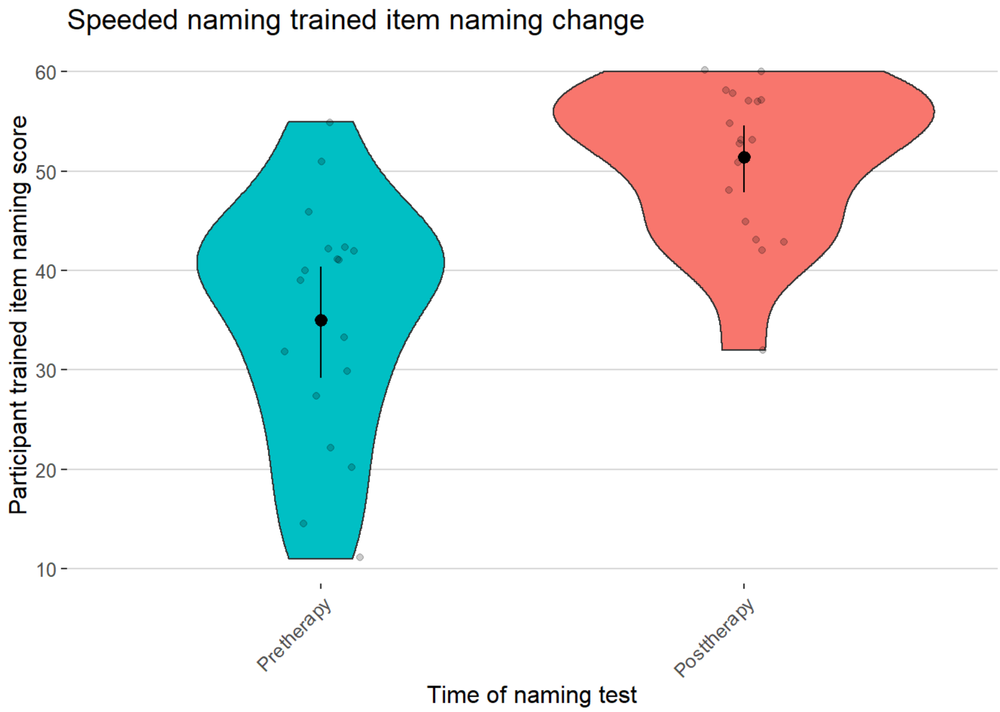

Figure 5.6. Participant scores for items trained using the Speeded treatment pre- and post- therapy.

Table 5.1. Raw scores for each participant (out of a total of 60) before and after (pre and post) each treatment (Pros = Prosody, Phon = Phonology, Sem = Semantic, Spe = Speeded, Int = Interfered, Ges = Gesture).

Participant	Pros Pre	Pros Post	Phon Pre	Phon Post	Sem Pre	Sem Post	Spe Pre	Spe Post	Int Pre	Int Post	Ges Pre	Ges Post	Control average Pre	Control average Post
AB	34	48	38	45	43	47	30	53	30	49	44	43	35.83	35.33
AL	52	57	56	60	56	57	51	60	49	60	52.5	56	52.17	52.33
CD	40	55	34	46	33	50	32	57	42	54	32	49	38.5	38.67
CH	53	56	52	58	47	55	42.5	58	48	55	53	56	49	51.33
CM	30	53	35	43	28	41	14.5	45	28.5	53	32	54	43.83	44.67
DA	11	22	26	40	18	43	20	42	22.5	42	18	28	20.5	19
DF	41	50	48	54	41	55	33.5	48	37	49	48	53	39.67	37.33
DM	52	59	45	48	47	50	39	53	49	52	49	53	45.33	47.67
Ebo	47	58	53	57	51	52	41	55	45.5	58	49	59	47.67	49.33
JE	53.5	57	55	59	53	57	55	60	54	60	52	60	53.67	55.67
JP	41	56	46	55	35	54	27.5	57	39.5	52	40	60	38.5	41
MH	5.5	14	9	24	15	24	11	32	14	21	13.5	26	10.83	12.67
NB	45	54	38	51	43	50	41	51	45	44	38	50	43.67	43.33
NC	50	53	49	59	45.5	54	46	57	48	60	54	55	47	49.67
PD	40	40	41	46	40	47	42	43	42	53	45	45	41.67	41.33
PT	30	35	29.5	46	26	40	22	43	21	43	24	32	28.17	30
Shi	53	57	53	60	50	54	42	58	48	59	53	56	52.33	52.67
WT	44	49	48.5	52	43.5	50	40	53	41	52	44	52	42.17	44.33
Mean	**40.11**	**48.5**	**42**	**50.17**	**39.72**	**48.89**	**35**	**51.39**	**39.11**	**50.89**	**41.17**	**49.28**	**40.58**	**41.46**

Q1.2: Which, if any, treatments show a stronger therapy-related change in naming than others?

Having established that all the treatments were shown to have been associated with a significant improvement in the naming of trained items, we next investigated whether treatments differed significantly in the degree of effectiveness. A comparison was therefore performed to determine if there was a significant difference between the improvements facilitated by the different treatments. Likelihood ratio tests were used to compare the goodness of fit of experimental and null cumulative-link mixed models (CLMMs) in order to determine if there was a significant effect of treatment on item improvement scores, with variation between people with aphasia (PWA) accounted for as a random effect. CLMMs allow us to perform analyses based on the properties of each individual item used, meaning that the effects of many different factors in driving the overall results can be assessed. Analysis at the item level also allows for more robust analyses compared to using a single score for each participant. For each CLMM, a null model is compared to an experimental model (which includes the factors we are interested in understanding the impact of). As well as showing whether there is a statistically significant difference between the null and experimental models, this comparison provides an AIC value- a measure of how well the experimental model predicts the dependent variable scores in comparison to the null model (with lower AIC scores indicating a better fit). This means that between several models for a set of data, AIC values show the quality of each model relative to the others- how good each model is at predicting the results found in the study, and therefore the level of influence of the factors included in each model in determining those results. If a model with one additional factor than another also has a dramatically lower AIC value, this indicates the large influence this factor had on the results.

When quantifying the differences between models, model AIC scores are compared to the lowest AIC value of the available models, i.e. to the AIC value of the best fitting model. When assessing the relative merits of models in the set, models with an AIC value ≤2 larger than the one for the best model are seen as having substantial support. Models with an AIC value 4 to 7 larger than the most optimal value are seen as having considerably less support. Finally, models with an AIC value 10 or more larger than the optimal value are seen as having essentially no support relative to the best model's AIC value (Burnham and Anderson, 2004). So when interpreting the relative effectiveness of models, we can judge AIC values within 2 of each other to be reasonably similar, values 4-7 apart to have a notable difference, and values 10 or more apart to show a large difference between the models. This allows us to compare the AIC values between multiple models, allowing us to gain a better understanding of which factors and interactions are driving our results, and build a model which accounts for as much of the variation in our results as possible. We can then perform comparisons within and between the different factors in these models to gain a better understanding of how they interact and their associations with the dependent variable outcomes. All CLMMs were performed using all item scores for each participant, as opposed to using average overall participant scores for each treatment.

Pairwise comparisons from the above likelihood ratio test were then performed to identify which specific treatments drove any effect of treatment on improvement. The likelihood ratio test found a difference in the experimental and null models (AIC values: null = 8465.1, experimental = 8454.1, df = 5, $p < .001$). The difference of >10 in AIC values suggests a large difference between the models. Subsequent pairwise comparisons found significant differences between the Executive Demand treatments and the treatments in the other two approaches. The Interfered treatment differed significantly from the Phonological ($z = 2.422$, $p = .015$), Semantic ($z = 2.563$, $p = .0104$), Gesture ($z = 2.545$, $p = .0109$), and Prosody ($z = 2.589$, $p = .01$) treatments. The Speeded treatment differed significantly from the Phonological ($z = 3.453$, $p < .001$), Semantic ($z = 3.548$, $p < .001$), Gesture ($z = 4.39$, $p < .001$), and Prosody ($z = 4.675$, $p < .001$) treatments. These differences are evident when noting the mean score improvement per item (each of the 30 items score a maximum of 2) in each treatment: Speeded: 1.152, Interfered: 0.945, Phonological: 0.657, Semantic: 0.643, Gesture: 0.601, and Prosody: 0.589.

The difference between experimental and null models was also persistent in likelihood ratio testing when using a model based on the post-therapy score, with the pre-therapy score included as a covariate (AIC values: null = 4904.2, experimental = 4900.2, df = 5, $p = .016$), showing that the effect of therapy was significant even when accounting for variation between PWA and variation in initial item scores.

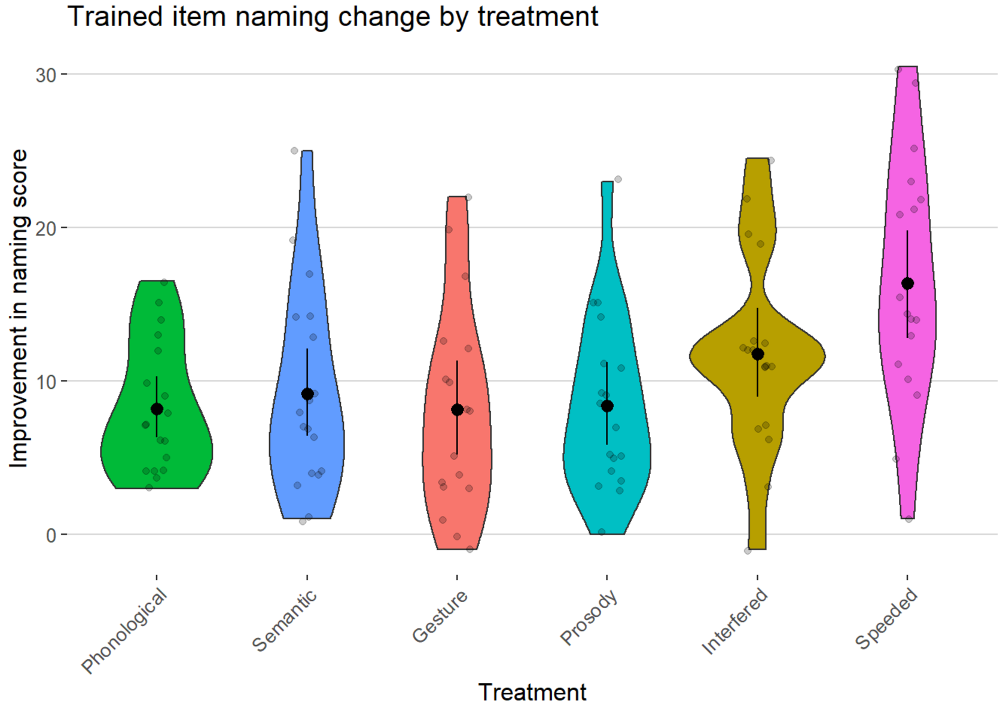

Figure 5.7. Improvement in participant trained item naming scores for each treatment.

In addition to these analyses, a series of ANOVAs were performed using overall participant scores for each treatment at the pre- and post-treatment timepoints (see Table 5.1 for raw data). Initially, a 6x2 ANOVA was performed to analyse the effects of treatment type (Semantic, Phonological, Gesture, Prosody, Interfered, or Speeded) and timepoint (pre- and post-treatment) on participant treatment score (out of a total of 60). A significant effect of timepoint ($p < .001$), and a significant interaction between treatment and timepoint ($F(5,85) = 9.343$, $p < .001$).

A series of 2x2 ANOVAs were then performed to compare each individual pair of treatment types, in order to identify the source of this significant interaction. All comparisons showed significant effects of timepoint. The only comparisons containing significant interactions were Speeded and Semantic ($F(1,17) = 21.180$, $p < .001$), Speeded and Phonological ($F(1,17) = 24.010$, $p < .001$) Speeded and Prosody ($F(1,17) = 53.581$, $p < .001$), Speeded and Gesture ($F(1,17) = 32.006$, $p < .001$), Speeded and Interfered ($F(1,17) = 8.254$, $p = .011$), Interfered and Prosody ($F(1,17) = 5.231$, $p = .035$), and Interfered and Phonological ($F(1,17) = 5.059$, $p = .038$) (Table 5.2).

Table 5.2. Interactions between Treatment and Timepoint effects for each 2x2 ANOVA treatment comparison.

Treatment comparison	*F value*	*Degrees of Freedom*	*p value*
Semantic/Phonological	0.679	1,17	.421
Semantic/Gesture	0.484	1,17	.496
Semantic/Prosody	0.246	1,17	0.627
Semantic/Speeded	21.180	1,17	< .001*
Semantic/Interfered	2.599	1,17	.125
Phonological/Gesture	0.001	1,17	.971
Phonological/Prosody	0.022	1,17	.884
Phonological/Speeded	24.010	1,17	< .001*
Phonological/Interfered	5.059	1,17	.038*
Gesture/Prosody	0.067	1,17	.799
Gesture/Speeded	32.006	1,17	< .001*
Gesture/Interfered	3.491	1,17	.079
Prosody/Speeded	53.581	1,17	< .001*
Prosody/Interfered	5.231	1,17	.035*
Speeded/Interfered	8.254	1,17	.011*

Q1.3: Were therapy effects maintained through to followup?

In order to investigate the maintenance of therapy effects through to the follow-up test, we compared changes in item scores for both the whole dataset and each treatment at multiple timepoints. Firstly, in order to evaluate whether or not item scores were maintained effectively from immediately post-therapy to the delayed followup, Wilcoxon signed rank tests were performed to determine if there was a significant difference between scores at these timepoints. Significant differences between immediate and delayed post-therapy scores were found for the items in the Phonological ($p = .045$, mean 50.17 immediately post-therapy and 45.6 at follow-up), Semantic ($p = .002$, mean 48.89 immediately post-therapy and 42.07 at follow-up), Gesture ($p = .012$, mean 49.28 immediately post-therapy and 43.87 at follow-up) Speeded ($p < .001$, mean 51.39 immediately post-therapy and 39.93 at follow-up), and Interfered ($p = .003$, mean 50.89 immediately post-therapy and 43.8 at follow-up) treatments, suggesting that there was significant loss of naming gains in the months after study completion for these treatments. Overall item scores were also significantly different at follow-up ($p < .001$, mean 45.31 immediately post-therapy and 42.71 at follow-up). However, differences between item scores only approached significance for the Prosody treatment, with an average difference between scores of only 4.7 ($p = .053$, mean 48.5 immediately post-therapy and 43.8 at follow-up), suggesting that improvements for this treatment was maintained more successfully than the other treatments, albeit from more modest immediately post-therapy levels.

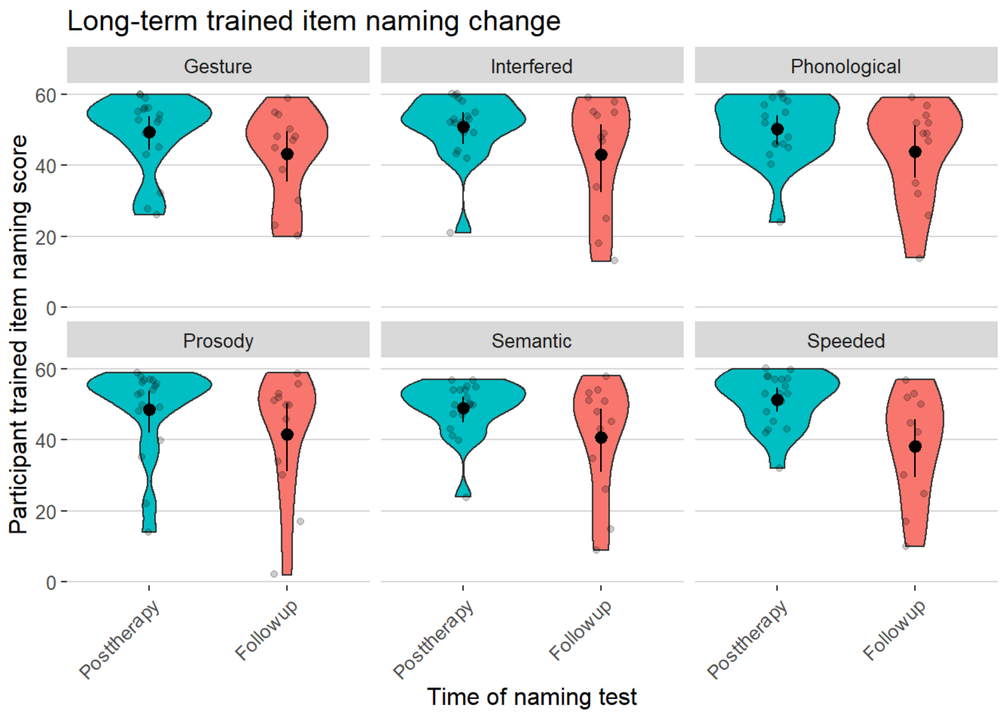

Figure 5.8. Scores by PWA at the immediate and delayed posttherapy timepoints.

As several treatments did not show consistent maintenance of naming gains from immediate to delayed post-therapy testing, further Wilcoxon signed rank tests were performed to determine if this drop in scores was large enough to cancel out all therapy effects; i.e., whether there was still a significant difference between pre-therapy and post-therapy delayed item scores. Overall, and for each treatment, item scores were still significantly different between pre-therapy and delayed post-therapy tests. Overall: $p < .001$, with a mean score of 38.94 pre-therapy and 42.71 at follow-up; Phonological: $p = .012$, with a mean score of 42 pre-therapy and 45.6 at follow-up; Semantic: $p = .041$, with a mean score of 39.72 pre-therapy and 42.07 at follow-up; Gesture: $p = .036$, with a mean score of 41.17 pre-therapy and 43.87 at follow-up; Prosody: $p = .006$, with a mean score of 40.11 pre-therapy and 43.8 at follow-up; Speeded: $p < .001$, with a mean score of 35 pre-therapy and 39.93 at follow-up; Interfered: p = .003, with a mean score of 39.11 pre-therapy and 43.8 at follow-up. This suggests that, while item scores were significantly lower for the Phonological, Semantic, Gesture, Speeded and Interfered treatments between the immediate and delayed posttherapy tests, both timepoints still showed a significant improvement in items using these treatments from pre-therapy. However, overall, the Prosody treatment performed the most successfully at maintaining therapy gains in the months following completion of the study.

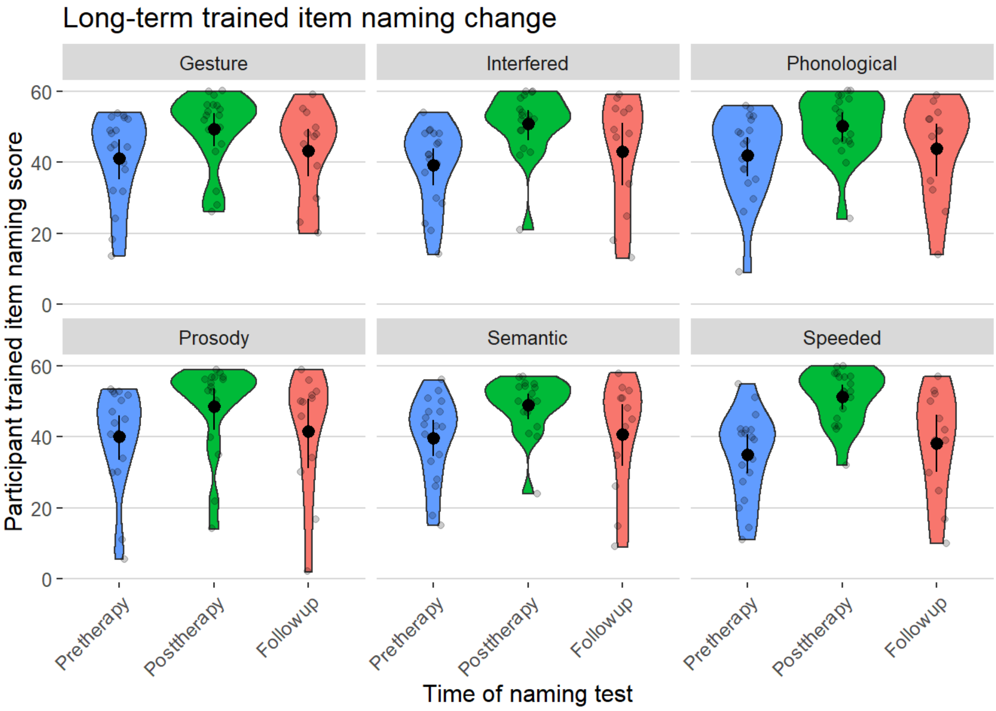

Figure 5.9. Immediate and delayed improvement in PWA scores for each treatment.

Q1.4: Are these therapy-related changes affected by the order the treatments are presented in (are there any order effects?)

In order to determine if the improvements facilitated by the treatments were affected by the order they were presented in, item score improvements were compared by order group. A Friedman test was performed on the item scores from the 3 different order groups. An order effect was found, with a significant difference being present between the groups (Friedman chi squared = 9.29, p = .009).

To investigate the impact of this order effect on the data, another cumulative-link mixed model was created with Order included as a fixed factor (accounting for its effect). Likelihood ratio testing found a significant difference in this experimental model and the null model used in Q1.2, with an experimental model AIC value close to that found in the original Q1.2 experimental model (AIC values: null = 8465.1, experimental = 8456.3, df = 5, p = .002). Similarly, a direct comparison of the two experimental models showed similar AIC values and a high p value (AIC values: Q1.2 model = 8454.1, Q1.4 model = 8456.3, df = 2, p = .413). While there was a significant difference in improvement by item between order groups, this had little impact on the effect of the differing treatments on improvement and is likely to vary little between treatments. A difference in AIC value of slightly over 2 does not meet the threshold of notable difference between models (a difference of at least 4 points) (Burnham and Anderson, 2004).

Including Order as a fixed effect altered the item means of each treatment only slightly- Speeded moved from 1.152 to 1.148, Interfered moved from 0.945 to 0.939, Phonological from 0.657 to 0.649, Semantic from 0.643 to 0.637, Gesture from 0.601 to 0.594, and Prosody from 0.589 to 0.582.

Table 5.3. Statistics for each CLMM used in Q1 and their likelihood ratio tests to the null model.

Factors included	AIC value	df	P-value
null	8465.1		
Treatment	8454.1	5	<.001 ***
Treatment and Order	8456.3	7	.002**

Results Section 2: Inter-participant variation

The questions in this section of the results are all focused on the amount and type of inter-participant variation between the treatments. What we are most interested in assessing in this section is which factors affect participant response to each treatment, and whether or how the optimal treatment for recovery may vary based on different inter-participant factors. Initially, alluvial plots were created to attempt to gauge the degree of inter-participant variation in improvement prompted by the different treatments and approaches.

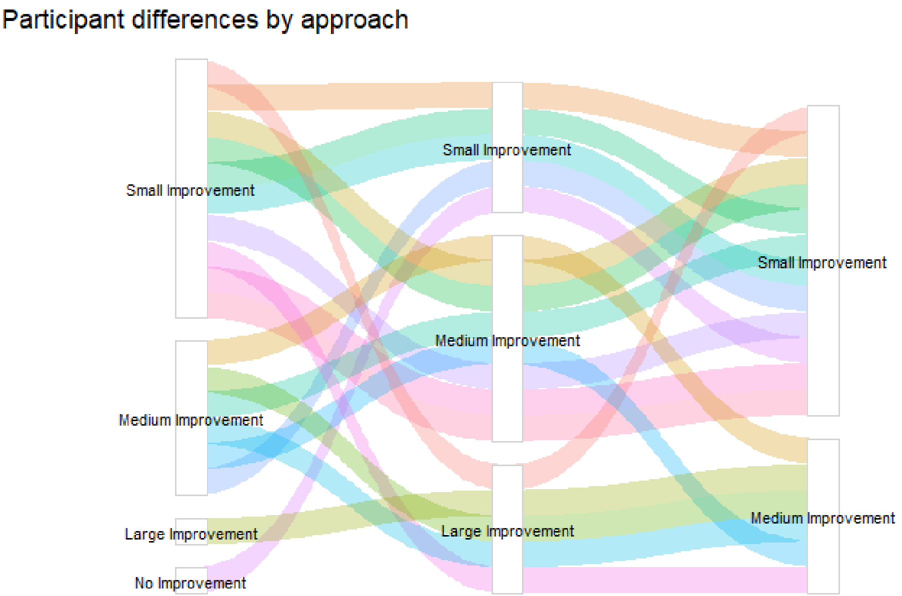

Figure 5.10. Alluvial plot showing broad degree of improvement in naming trained items by each participant across the different treatments.

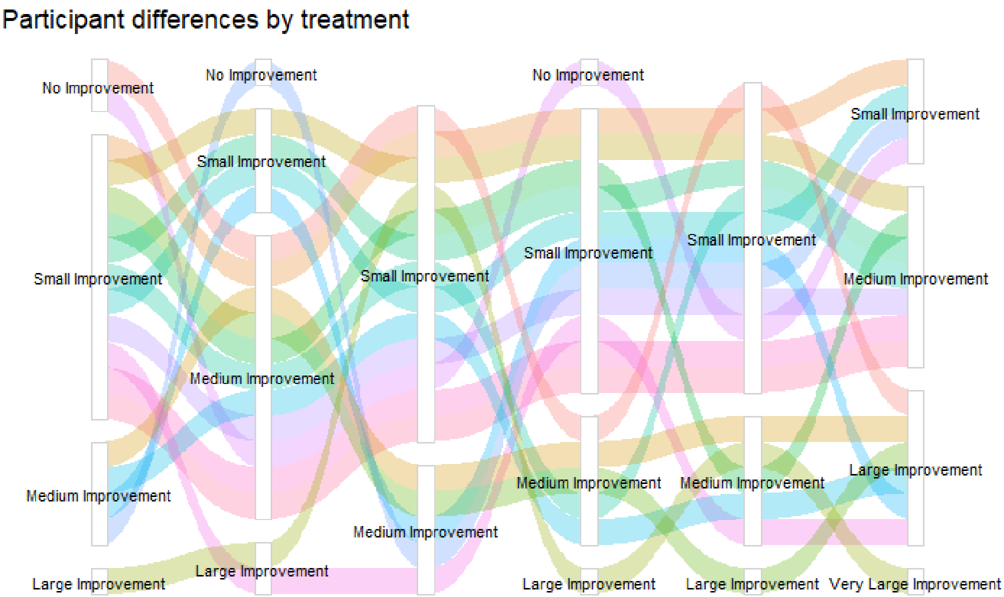

Figure 5.11. Alluvial plot showing broad degree of improvement in naming trained items by each participant across the different treatments.

2.1. Did inter-participant demographic factors modulate the effectiveness of treatment?

This question aimed to investigate whether previous findings – i.e., that factors such as PWA age, education level and time post-stroke have an impact on treatment effectiveness- were supported in this study, and to what extent these factors affected improvement. We hypothesised that age, education level and time post-stroke would all modulate the effectiveness of the different treatments (with a benefit for younger age, higher education level and earlier time post-stroke). We began by performing Spearman correlations between each of these three factors and immediate participant improvement in trained items for each treatment used. None of the factors exhibited strong correlations with overall participant improvement (Age: $r = 0.271$, Months poststroke: $r = 0.119$, Years in education: $r = 0.07$). Although initial visualisations and analyses did not indicate strong overall relationships between improvement and age, time in education or time poststroke, our analyses were focused on how these factors affected treatment effects differently as opposed to any overall trends.

2.2. Did inter-participant language factors modulate the effectiveness of treatment?

In order to determine whether or not inter-participant language factors modulated the effectiveness of treatment, participant naming error patterns prior to treatment were used to determine semantic and phonological ability ratings based on Dell's semantic-phonological model (Foygel & Dell, 2000; Schwartz et al, 2006). These measures of semantic and phonological ability for each participant were then correlated with immediate participant improvement in trained items for each treatment used. Spearman correlations were performed on improvement in the Semantic (Dell s $r = -0.501$, Dell p $r = -0.355$), Phonological (Dell s $r = -0.062$, Dell p $r = -0.592$), Gesture (Dell s $r = -0.17$, Dell p $r = -0.542$), Prosody (Dell s $r = -0.411$, Dell p $r = -0.385$), Interfered (Dell s $r = -0.328$, Dell p $r = -0.464$) and Speeded (Dell s $r =-0.478$, Dell p $r =-0.396$) trained items. This demonstrated a level of variance in how closely associated improvement in the naming was with participant initial semantic and phonological ability between the different treatments, suggesting that some treatments may show increased effectiveness for participants with a particular initial language deficit profile. Correlation with change in the control items was also examined (Dell s r = 0.199, Dell p r = 0.484). The difference in correlation coefficient between the treated items and control items suggested that the correlations for treated items were not simply due to a ceiling effect (e.g., that the participants with more severe deficits simply had more 'room' to improve in naming over the course of the study). The most notably varied treatments were Gesture and Phonological, both of which had a weak correlation between improvement and semantic ability, but a much stronger association with initial phonological ability.

In order to explore the relationships between language profile and improvement on each treatment further, Spearman correlations were also performed between improvement and numbers of specific error types pre-therapy- specifically, semantic and nonword errors. Each treatment was again correlated- Semantic (Semantic error r = 0.762, Nonword error r = 0.479), Phonological (Semantic error r = 0.351, Nonword error r = 0.628), Gesture (Semantic error r = 0.304, Nonword error r = 0.519), Prosody (Semantic error r = 0.487, Nonword error r = 0.386), Interfered (Semantic error r = 0.601, Nonword error r = 0.431), Speeded (Semantic error r = 0.652, Nonword error r = 0.451), and also the Control items (Semantic error r = -0.22, Nonword error r = -0.546). The overall stronger correlations revealed some clearer variations in response by treatment than obtained using the Dell model. In addition to the varied responses shown by Gesture and Phonological treatment shown in the Dell model analysis, using specific error types also revealed a stronger variance in Semantic and Speeded treatment improvement, which were strongly correlated with Semantic errors and less strongly correlated with Nonword errors. Prosody and Interfered improvement also showed a greater correlation to semantic errors and lesser correlation with nonword errors, but were overall more uniform than the other treatments.

Table 5.4. Spearman's correlation coefficients between measures of language ability and improvement in different treatments.

Treatment	Language factor	Correlation
Phonological	Dell s	-0.062
Phonological	Dell p	-0.592*
Phonological	Semantic errors	0.351
Phonological	Nonword errors	0.628**
Semantic	Dell s	-0.5*
Semantic	Dell p	-0.355
Semantic	Semantic errors	0.762**
Semantic	Nonword errors	0.479*
Gesture	Dell s	-0.17
Gesture	Dell p	-0.542*
Gesture	Semantic errors	0.304
Gesture	Nonword errors	0.519*
Prosody	Dell s	-0.411*
Prosody	Dell p	-0.385
Prosody	Semantic errors	0.487*
Prosody	Nonword errors	0.386
Interfered	Dell s	-0.328
Interfered	Dell p	-0.464*
Interfered	Semantic errors	0.601**
Interfered	Nonword errors	0.431*
Speeded	Dell s	-0.478*
Speeded	Dell p	-0.396
Speeded	Semantic errors	0.652**
Speeded	Nonword errors	0.451*
Control	Dell s	0.199
Control	Dell p	0.484*
Control	Semantic errors	-0.22
Control	Nonword errors	-0.546*

*Moderate correlation (>0.4). **Strong correlation(>0.6).

2.3: Did inter-participant variations in lesion volume modulate the effectiveness of treatment?

To determine the impact of lesion volume on treatment effectiveness, pre-therapy participant scans underwent preprocessing detailed earlier. Lesion volume was then calculated from the number of voxels included in each participants' binary lesion mask. This provided a lesion volume number for each participant that was used in subsequent analyses. A Spearman correlation showed a moderate association between lesion volume and overall improvement in naming across all treatments (r = 0.37).

Further investigation of Q2.1, 2.2 and 2.3

We were interested in teasing apart the contributions of each factor to the variation in results. While basic correlations could give us an indication of general trends in the data, we hoped to break down the contribution of each factor to the overall variation in treatment, and understand how each factor affected individual variation in improvement across the different treatments. In order to do this, we used likelihood ratio testing for a variety of CLMMs.
We began by investigating the demographic factors. We used a series of CLMMs to compare a null model against 3 separate models which included each of the demographic factors as fixed effects in order to assess how much additional variation in the results was explained by these factors. Likelihood ratio tests comparing a null model to different experimental models found that the models including number of years in education, age, or number of months post-stroke all provided a marginally better fit than treatment alone when compared to the null model (AIC value: 6445.6). Years in education: (AIC value: 6443.0, df: 6, p = 0.023), Months poststroke: (AIC value: 6443.2, df: 6, p = .025), Age: (AIC: 6441.0, df: 6, p = .011).

A model including all 3 demographic factors as fixed factors was also created; though although likelihood ratio testing in comparison to the null model revealed that this did not provide a significant improvement over any of the individual models (AIC value: 6443.3, df: 8, p = .019). However, a subsequent model also including the interactions between the factors (for which likelihood ratio testing was also performed, in comparison to a null model), did not further improve the model fit (AIC value: 6486.3, df: 47, p = .243). This seems to show that, as including both factors improves the fit of the model, but including the interactions between them does not, the factors may have fairly independent effects on improvement, without much overlap or interaction between them.

In order to assess language factors, we then performed individual CLMMs including pre-therapy Dell phonological and semantic scores as fixed factors and compared them to a null model only using treatment. Again, likelihood ratio testing in comparison to the null model found that including these factors improved the fit of the model, with Phonological score in particular providing a better fit than any previous model: Semantic (AIC value: 6442.3, df: 6, p = .018), Phonological (AIC value: 6435.4, df: 6, p = .001). As in section 2.1, a model including both factors was then produced and compared to the null model - this was a significant improvement over the null, and a better fit than the models based only on Phonological or Semantic Dell scores (AIC value: 6431.0, df: 7, p < .001).The AIC value was more than 4 lower than any other model, suggesting that this model was notably better than any of the others The model including interactions between the two factors failed to improve on this model any further (again when likelihood ratio tests in comparison to the null model were performed) (AIC value: 6454.7, df: 23, p = .033). This suggests that, while both pre-therapy semantic and phonological ability separately impact improvement as a result of treatment, the interaction between the two does not add any additional explanatory power.

A likelihood ratio test was then performed comparing a model including lesion volume as a fixed factor with the null model. Surprisingly, including lesion volume did not improve on model 2 (AIC value: 6441.5, df: 6, p = .013).

A model was also created containing all factors from Q2.1,2.2 and 2.3 – likelihood ratio testing showing this model to be an improvement, but not the overall best fit model (AIC value: 6436.8, df: 11, p = .001), performing similarly to the phonological score model, with a difference of less than 2 between them. The lack of improved fit provided by the inclusion of multiple additional factors in these models suggests an overlap in explanatory power between several of the factors; for example, the fact that the model including all the factors was not a better fit than the model including only language factors suggests that Dell s and p may overlap with the demographic factors to a large degree in the variance in the results that they account for. This may suggest that pre-study language ability is largely determined by factors like education and time poststroke, meaning they overlap in terms of the treatment related improvement they explain.

Table 5.5. Statistics for each CLMM used in Q2 and their likelihood ratio tests with the null model. CLMMs are split into those focused on demographic factors (red), those focused on language factors (green) and those focused on the lesion-related factor (blue), with each branch increasing in complexity to identify the model of best fit. Bold rows show models with lowest AIC value.

Factors (in addition to Treatment) included in the model [model reference number in brackets]	AIC value	df	P value
Null model [1]	6445.6		
None (Treatment only) [2]	6441.2	5	0.013 *
Age [3]	6441.0	6	0.011 *
Time poststroke [4]	6443.2	6	0.025 *
Years of education [5]	6443.0	6	0.023 *
Age + Time poststroke + Years of education [6]	6443.3	8	0.019 *
Age + Time poststroke + Years of education + interactions [7]	6486.3	47	0.243
Dell p score [8]	**6435.4**	**6**	**0.001 ****
Dell s score [9]	6442.2	6	0.018 *
Dell s score + Dell p score [10]	**6431.0**	**7**	**<0.001 *****
Dell s score + Dell p score + interactions [11]	6454.7	23	0.033 *
Lesion volume [12]	6441.5	6	0.013*
Age + Time poststroke + Years of education + Dell s score + Dell p score + Lesion size [13]	**6436.8**	**11**	**0.001****
Age + Time poststroke + Years of education + Dell s score + Dell p score + Lesion size + interactions [14]	6510	83	0.08

Coefficient statistics for CLMM 13 (including all the considered inter-participant factors) were also examined in order to determine the directionality and magnitude of the effect each individual inter-participant factor had on PWA improvement (as shown in Table 5.6). Coefficient statistics for the CLMMs including interactions were also examined in order to untease the interactions between the effects of each inter-participant factor and the effects of each treatment.

Table 5.6. Coefficient statistics for each inter-participant factor considered in the penultimate CLMM (including all inter-participant factors).

Factors	Z value	P value
Dell p score	-1.785	0.036*
Dell s score	-2.103	0.074
Age	1.316	0.188
Time poststroke	-0.296	0.767
Years of education	-0.829	0.407
Lesion size	1.059	0.289

As shown in Table 5.6, coefficient statistics largely supported the findings in sections 2.1,2.2 and 2.3. Pre-treatment language scores seemed to have the largest effect, with PWA with lower initial scores again having overall greater improvement. Age and Lesion Size were again found to have some effect on improvement, with older PWA with larger lesions improving more. Finally, time poststroke and years of education were again found to have a relatively minor impact on improvement.

Interaction effects in CLMM 14 were also examined. However, they were mostly comparatively minor. The only interactions between treatments and inter-participant factors which reached significance was between the Interfered treatment and the following factors: time poststroke ($z = -2.292$, $p = .0219$), Age ($z = -2.282$, $p = .023$), Years in education ($z = 2.362$, $p = .018$) Lesion Volume ($z = -2.42$, $p = .015$). This would seem to indicate that the Interfered treatment is affected less than the other treatments by inter-individual variation.

2.4. Did lesion location modulate the effectiveness of different treatments?

In addition to demographic and behavioural factors and overall size of the lesion, we were also interested in how the effects of each treatment varied based on lesion location. We correlated each treatment improvement score with signal intensity values of the T1-weighted images for the 17 participants who had scans performed. This was done using a voxel-based correlational methodology (VBCM: Tyler, Marslen-Wilson, & Stamatakis, 2005), a variant of voxel-lesion symptom mapping (VSLM: Bates et al., 2003). This allowed us to determine the areas of damage associated with differences in improvement for each treatment.

VBCM results were mixed- significant cluster-level associations were only found for improvement in the Prosody treatment at a threshold of $p < .001$, p-FWE-corr < .01. However, at a reduced threshold of p-FEW-corr < .05, cluster-level correlations were also found for improvements in the Phonological and Speeded treatments.

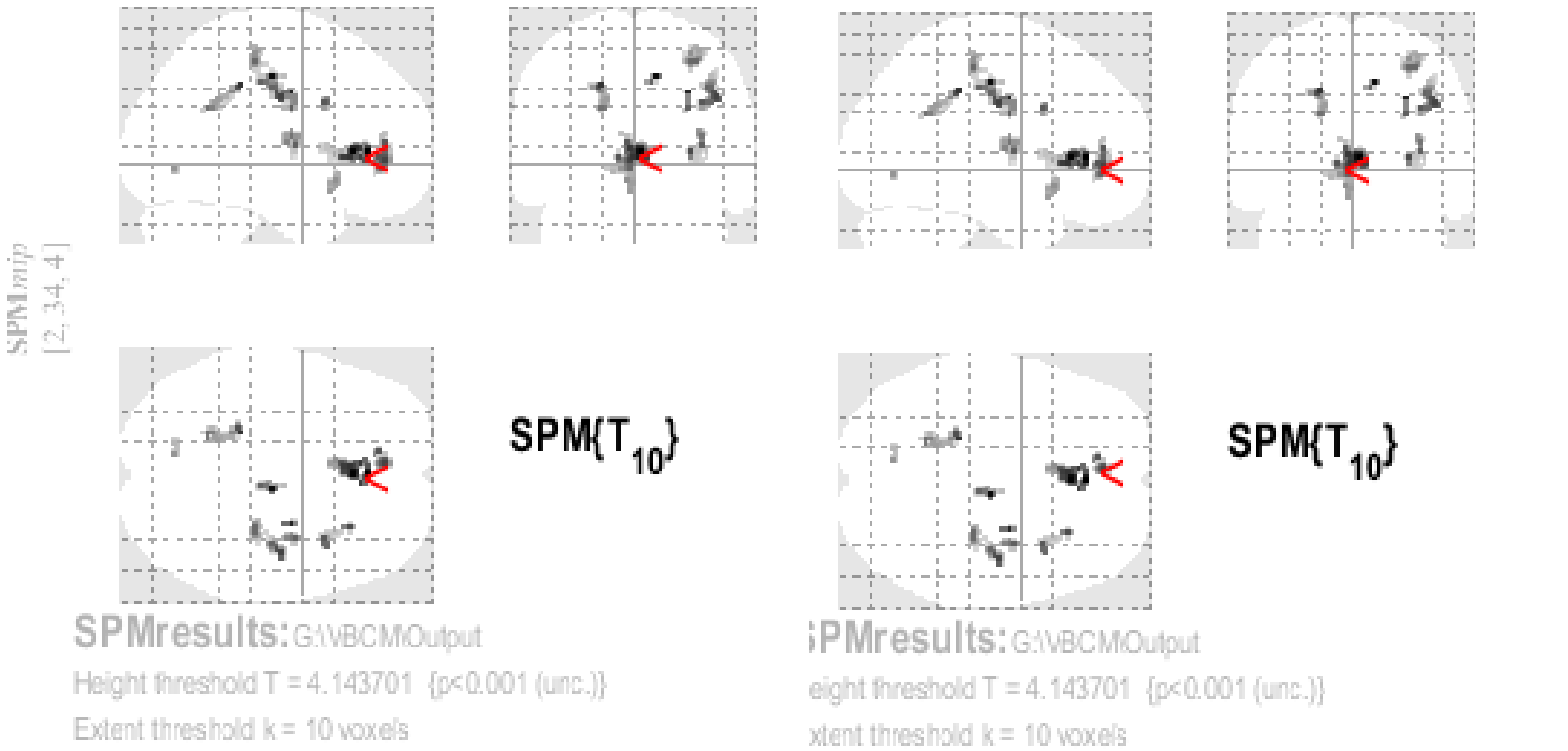

Figure 5.11. VBCM analysis showing correlates to improvement in the Phonological treatment. Clusters significant at reduced threshold are marked by arrow (p-FWE-corr: Left image: .015, Right image: .042).

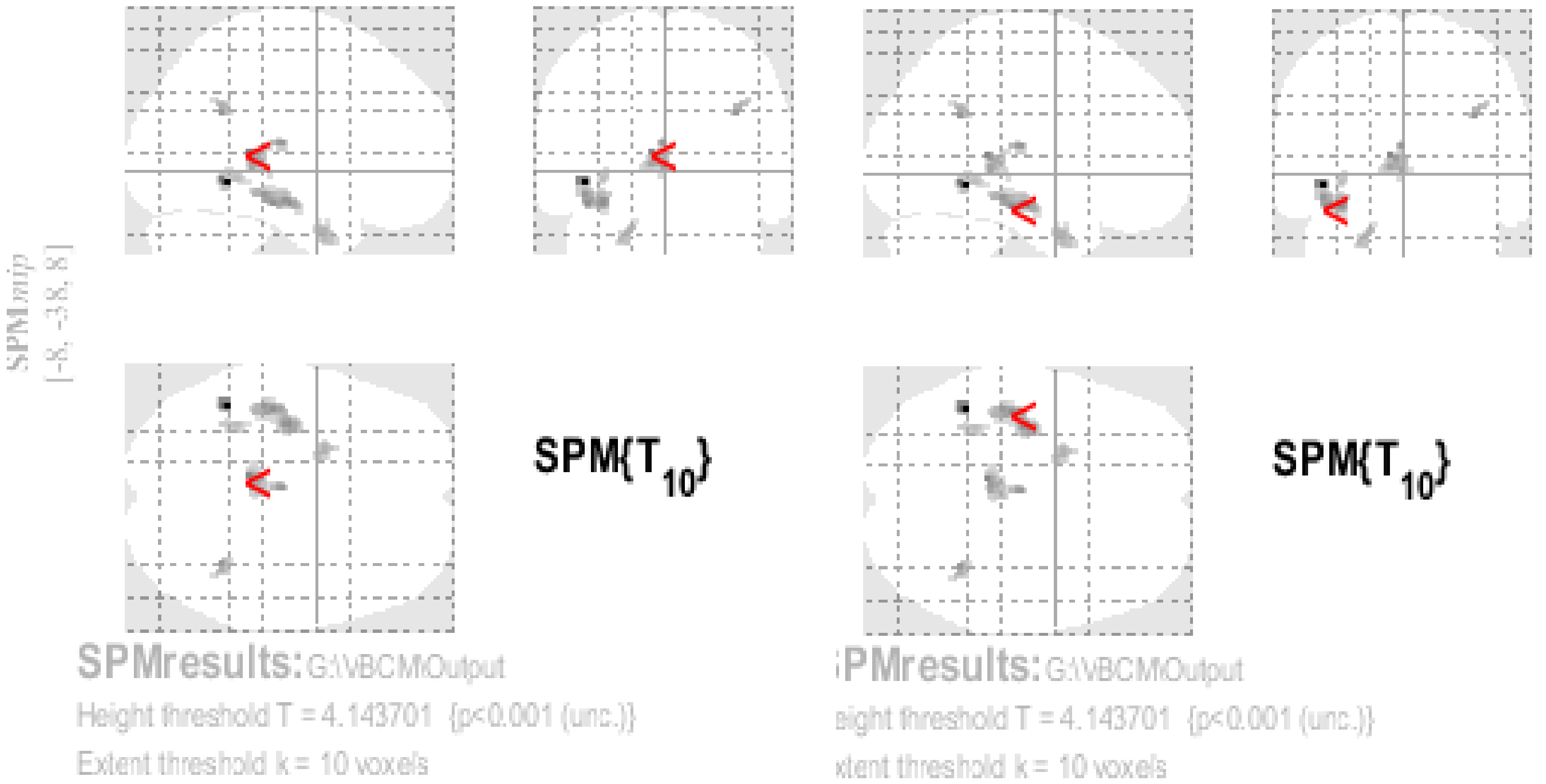

Figure 5.12. VBCM analysis showing correlates to improvement in the Prosody treatment. Significant clusters are marked by arrows (p-FWE-corr: Left image:.006, Right image: .001).

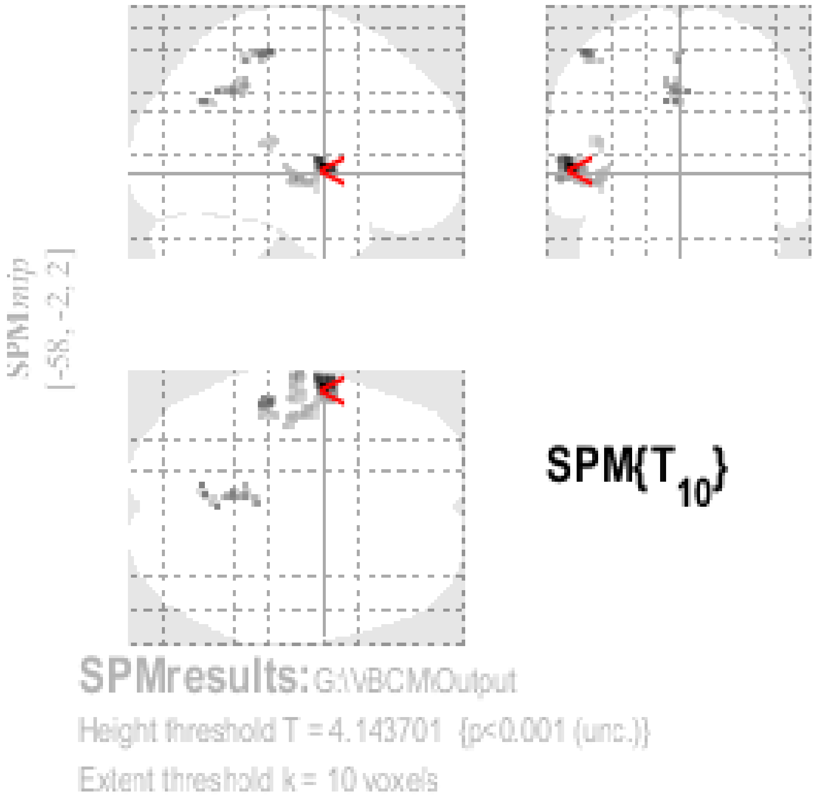

Figure 5.13. VBCM analysis showing correlates to improvement in the Speeded treatment. Cluster significant at reduced threshold is is marked by arrow (p-FWW-corr = 0.039).

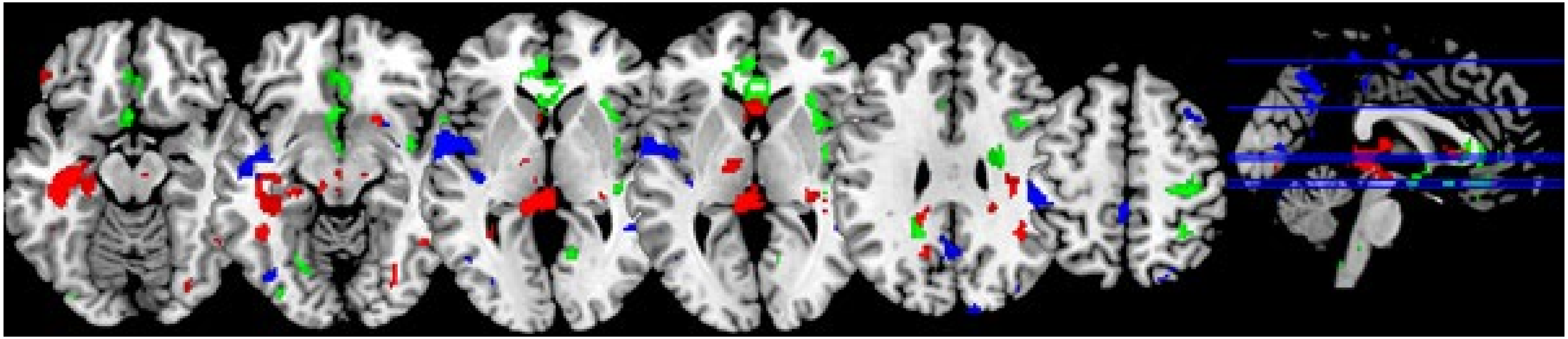

Figure 5.14. Results of VBCM analysis showing neural correlates to improvement in naming for Phonological (Green), Speeded (Blue) and Prosody (Red) treatments. Image thresholded at p < .005.

The strongest neural correlates were those with improvement in naming for Prosody treatment items, which showed clusters in both the posterior Inferior Temporal Gyrus (ITG) and posterior cingulate gyrus (see Figure 5.12). When using a reduced threshold, other neural correlates were found with improvement in naming for Phonological treatment items in the anterior Cingulate Gyrus and Paracingulate Gyrus, as shown in Figure 5.11, and with improvement in naming for Speeded treatment items in the Planum Polare/anterior Superior Temporal Gyrus (STG) (Figure 5.13). While these reduced threshold correlations were not significant at the more stringent threshold, their presence is notable in indicating some potential variation in treatment effectiveness based on the specific areas of lesioning in each PWA. Conversely, no significant correlations were found at either stringency level for improvement in the naming for the Semantic, Interfered, or Gesture treatments, suggesting these treatments may be more consistent in their effects regardless of specific regions of damage in each PWA.

Results Section 3: Generalisation beyond treated items

Q3.1: Did significant therapy effects in naming accuracy generalise to untrained words?

The broadest question being asked in this section was whether or not there was any generalisation in naming improvements to untrained words over the course of the entire study. In order to test this hypothesis, a Wilcoxon signed-rank test was performed on the item scores for the 40 control (untrained) items, comparing them at baseline and after all 3 periods of therapy. A significant difference was found ($p = .006$), demonstrating a significant improvement in participant naming scores over the entire length of the study.

In addition to this, we explored whether any generalisation occurred within the single 3-week bouts of treatment. An additional Wilcoxon signed-rank test was performed, comparing naming scores immediately before and after each period of treatment (e.g., comparing baseline naming scores to naming scores at interim test 1, interim test 1 to interim test 2, and interim test period 2 to post-therapy naming). A significant difference was found when comparing scores before and after the 3-week periods of treatment ($p = .005$).

Similarly, Wilcoxon signed-rank tests were also performed to compare scores before and after each individual approach, finding no significant difference (Model oriented (MO): p = .69, Alternative Modality (AM): p = .185, Executive Demand (ED): p = .236). Pairwise comparisons between the 3 approaches also found no significant differences in terms of improvement in untrained items over the course of each approach (AM – ED: z = 0.181, p = .856, AM – MO: z = 0.694, p = .488, ED – MO: z = 0.479, p = .632). These results would indicate that, over any individual 3-week therapy period with a particular approach, there is not any significant generalisation of naming improvement to untrained items, and this is true for each separate approach, which also did not have significant differences between them in terms of the level of generalisation they display- no one approach seems to be significantly more or less effective at producing generalisation of improvement to untreated items than any other approach.

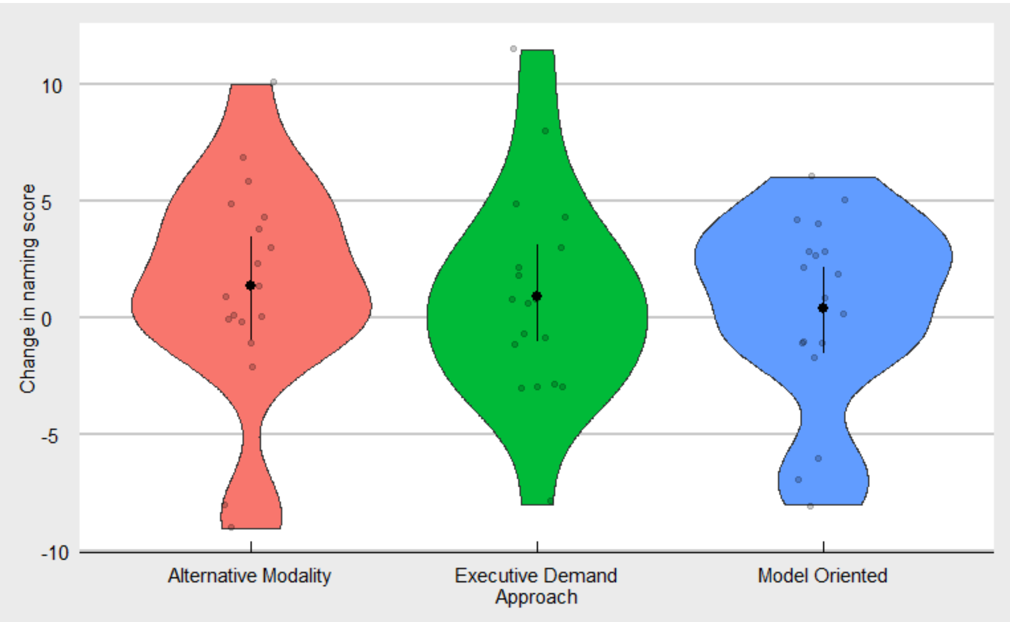

Figure 5.15. Violin plot showing change in naming score (out of a total maximum of 60) over each 3- week approach.

However, there was a significant improvement in untrained items over the whole study, suggesting that generalisation of improvement to untrained items does indeed occur in the long term if treatment continues. The lack of difference between the different approaches in the short term suggests, however, that this generalisation may occur regardless of the approach used.

Q3.2 Did significant therapy effects in naming accuracy generalise to measures of semantic or phonological processing (e.g., changes in error patterns)?

In order to judge whether or not therapy effects generalised to measures of semantic or phonological processing, scores in measures of these stages of processing were compared before and after each of the three approaches. A number of t-tests were performed comparing both Dell s and p scores and Semantic and Nonword errors before and after treatment in each approach- Model Oriented (Semantic and Phonological treatments), Alternative Modality (Gesture and Prosody treatments), and Executive Function (Speeded and Interfered treatments). Significant differences were found in Semantic errors before and after treatment with the Alternative Modality approach ($t = 2.149$, $df = 14$, $p = .05$) and the Executive Function approach ($t = 3.723$, $df = 14$, $p = .002$) (both showing a reduced number of errors post-treatment). When comparing Dell scores, the only significant difference found was in the Dell s score from before to after treatment with the Executive Function approach ($t = -3.00$, $df = 14$, $p = .01$). The strength of significance in the differences in measures of semantic processing for the Executive Function approach demonstrates that generalisation to semantic processing from this approach is present, while there is some evidence for generalisation to semantic processing from the Alternative Modality approach. However, the lack of a significant difference in the Dell s score casts doubt on the degree of generalisation for the latter.

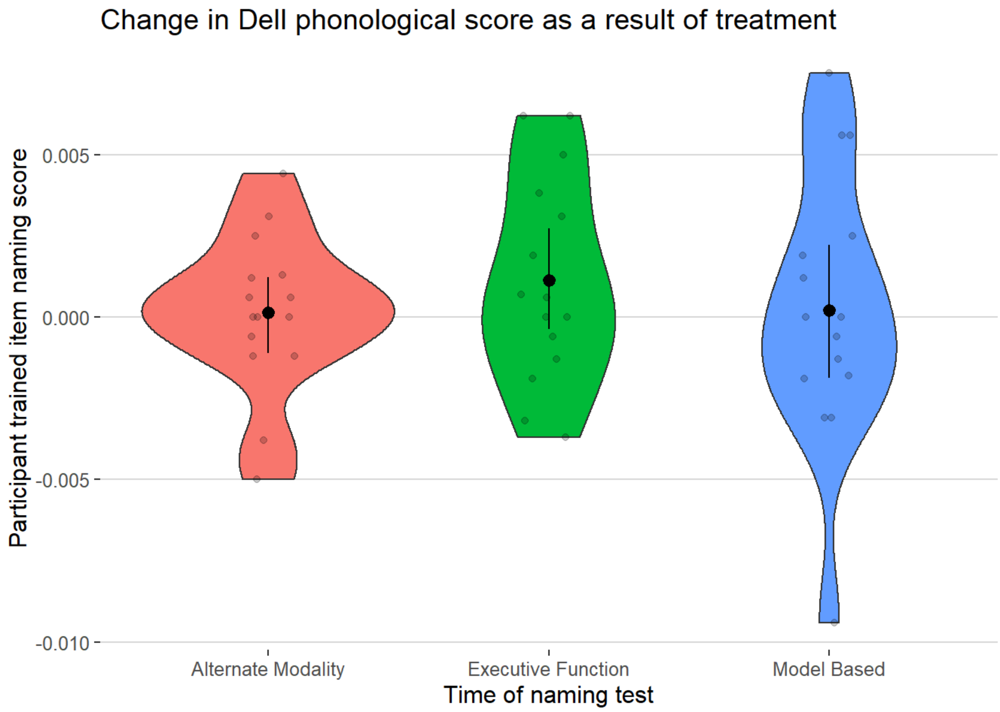

Figure 5.16. Change in Dell p score before and after treatment using each approach.

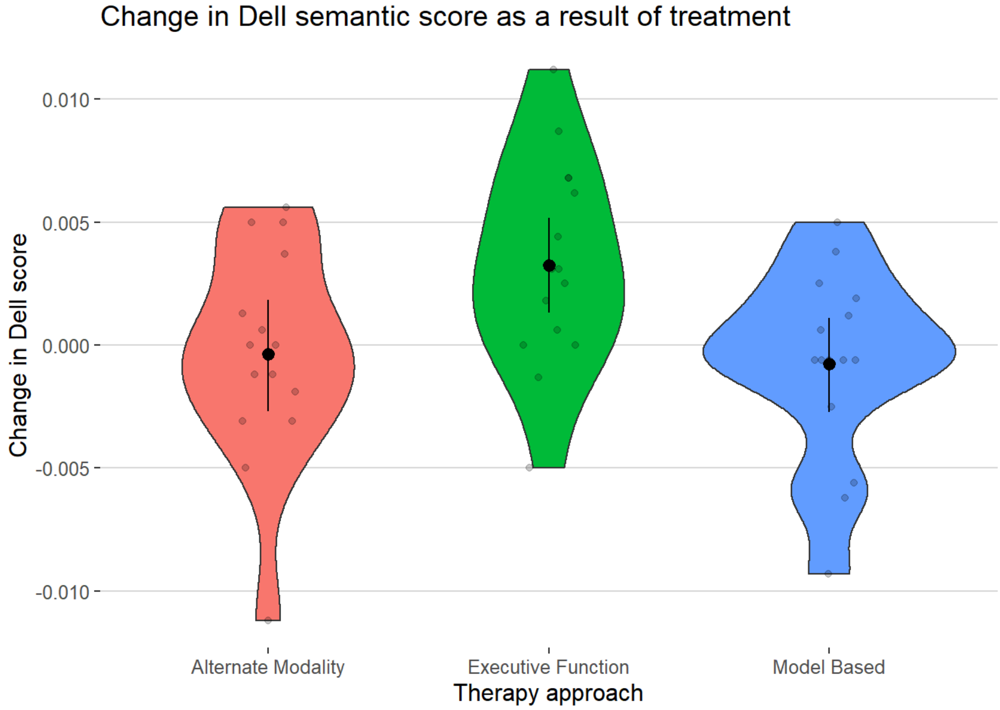

Figure 5.17. Change in Dell s score before and after treatment using each approach

In order to confirm the effect of difference between approaches on the level of generalisation to semantic and phonological processing, experimental and null linear mixed models were compared to determine if there was an overall significant effect of approach on the difference between Dell s and p scores, with variation between PWA accounted for as a random effect. Likelihood ratio testing found a significant difference between the experimental and null models was found (experimental AIC = -732.6, null AIC = -729.01, Chisq = 7.589, df = 2, p = .023). Subsequent pairwise comparisons found significant differences between the Executive Function approach and the Alternative Modality (z = 2.367, p = .033) and Model Oriented (z = 2.503, p = .025) approaches. No significant difference was found between the Alternative Modality and Model Oriented approaches (p = .869). This shows a significant effect of approach on the amount of generalisation to semantic and phonological processing, which is driven by the Executive Function approach having a greater level of generalisation than the other approaches. As this approach also has a significantly greater improvement in naming of treated items, it may be the case that the generalisation proportional to amount of improvement in trained items is similar across approaches. However, the greater improvement caused in the Executive Function approach also results in a greater level of generalisation to semantic processing.

3.3. Did significant therapy effects in naming accuracy generalise to wider measures of language and cognitive functions (e.g., comprehension, executive skills)?

In order to determine the effects of treatment on wider measures of language and cognitive functions, Wilcoxon signed rank tests were performed comparing both grouped and individual Comprehensive Aphasia Test (CAT; Swinburn, Porter and Howard, 2005) subtest T scores before and after treatment (delayed treatment effects- the CAT was performed before any treatments were carried out and after all treatments were completed). CAT T scores were calculated from the raw subtest scores, allowing individual scores to be scaled to allow comparison against PWA as a whole (Conroy and Scowcroft, 2012). In order to investigate possible treatment generalisation effects from the significant gains in lexical retrieval to overall language comprehension and expression, averages of the T scores for several sections of the CAT- each Comprehension total (Auditory Comprehension Total and Visual Comprehension Total) and each Expression total (Expressive Reading, Writing, Repetition and Naming Totals) were calculated, producing scores for overall language comprehension and expression. The delayed treatment effects of these scores were then compared, producing a significant difference for overall comprehension ($p = .007$), but not for expression ($p = .222$).

To investigate the effects at both word and sentence level, delayed treatment effects for the T scores for word and sentence repetition subtests (representing language expression at word and sentence level) and averages for the comprehension auditory and visual word comprehension subtests (representing overall comprehension at the word level) and auditory and visual sentence comprehension subtests (representing overall comprehension at the sentence level) were also compared. Despite a significant result for overall generalisation to comprehension, none of these scores were found to significantly differ between pre- and post-treatment (expression at word level: $p = .528$, expression at sentence level: $p = .166$, comprehension at word level: $p = .851$, comprehension at sentence level: $p = .753$). Finally, a number of individual subtest scores were also compared in order to evaluate any generalisation to more specific language and cognitive functions. Word Fluency scores were compared to evaluate any changes to executive function; gesture repetition scores to evaluate changes in ability to produce gestures as a result of the gesture specific treatment, semantic memory to investigate semantic memory, nonword repetition to investigate phonological processing, and finally, digit string repetition to investigate any changes to short term memory. Again, none of these individual subtests were found to differ significantly from pre- to post-treatment (gesture: $p = .154$, word fluency: $p = .249$, gesture repetition: $p = .167$, semantic memory: $p = .181$, digit string repetition: $p = .281$, nonword repetition: $p = .419$). Overall, there is therefore very limited support for generalisation of the significant therapy effects in naming to wider language and cognitive functions. While there was an overall significant improvement in average comprehension score, this was very small in scale (the mean T score improved from 54.179 to 54.972). This effect was not significant when looking at subtests for word or sentence level comprehension specifically.

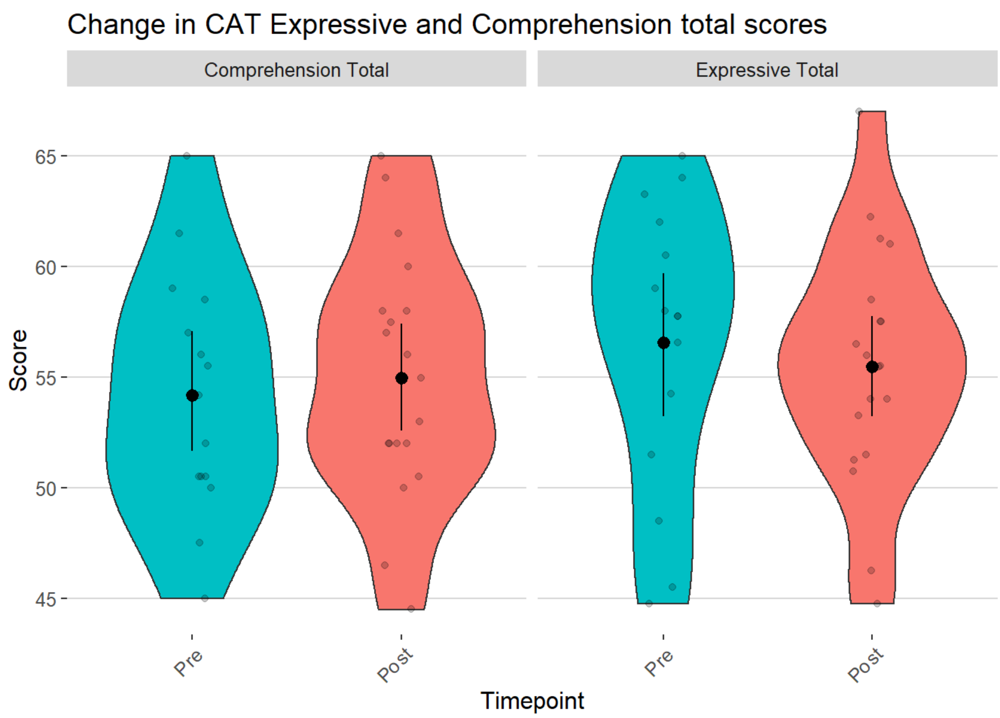

Figure 5.18. A graph to show delayed improvement for PWA T score averages for each comprehension and each expressive language section in the CAT (maximum possible T score: 75)

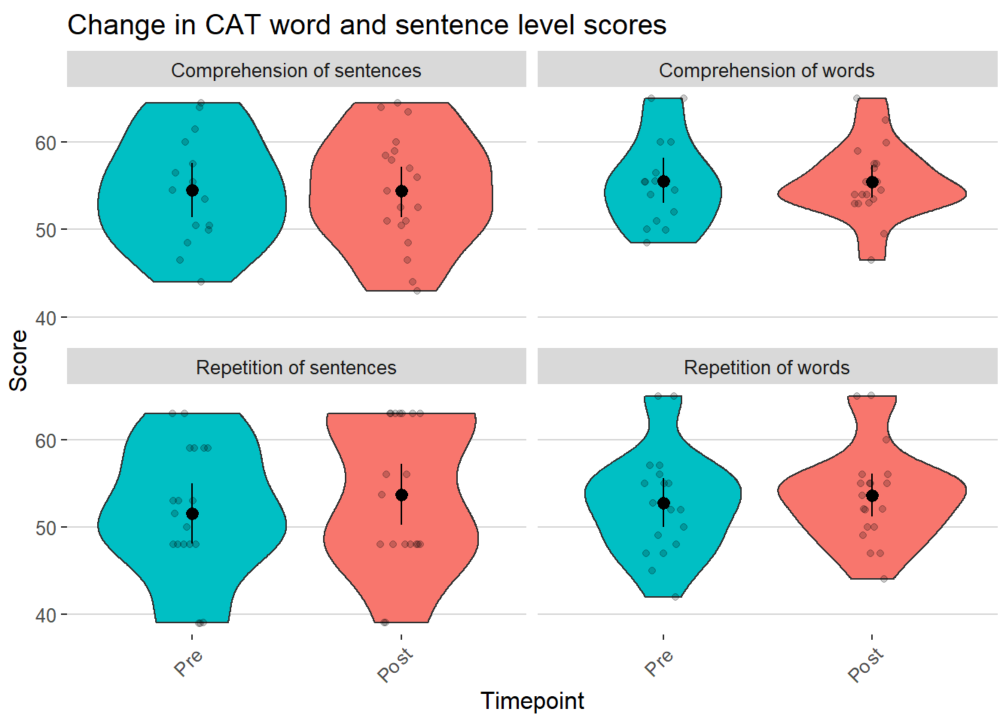

Figure 5.19. A graph to show delayed improvement for PWA T score averages for subtests relating to expression and comprehension at the word and sentence level on the CAT (maximum possible T score: 75)

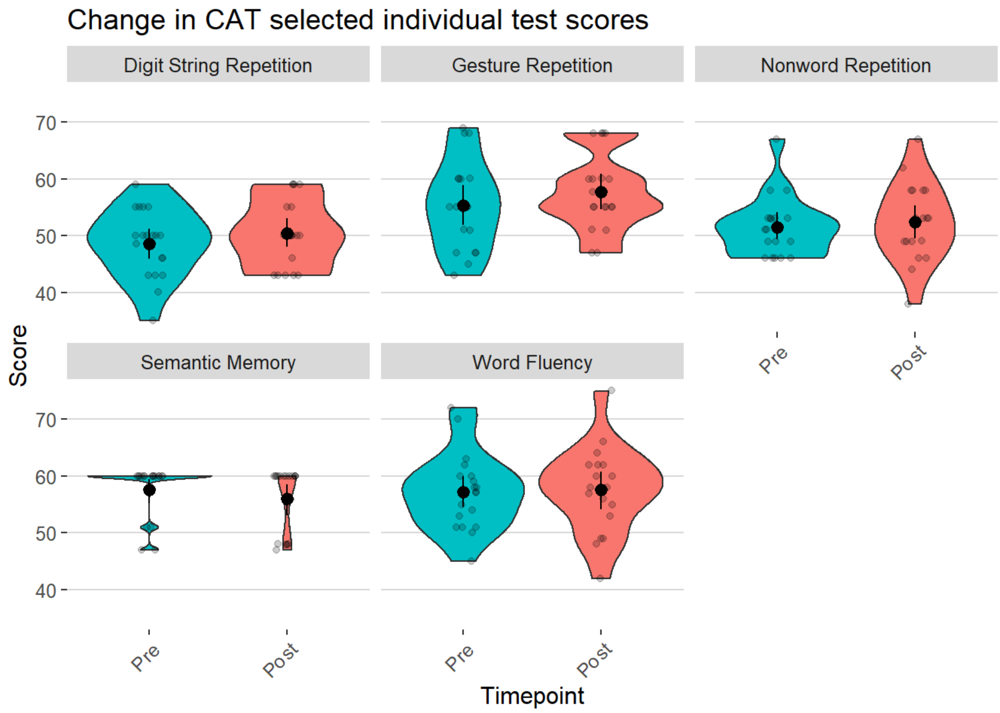

Figure 5.20. A graph to show delayed improvement for PWA T scores on individual subtests in the CAT (maximum possible T score: 75)

In order to discern the driving factors behind the significant difference and lack of significant difference in pre- and post-treatment scores for the Comprehension and Expression total scores respectively, we then divided them into separate modalities, as opposed to by ability at word/sentence/paragraph level. Comprehension was broken into Auditory and Visual comprehension, with Wilcoxon signed rank tests finding a significant difference in Auditory (p = .0208) but not Visual (p = .1087) comprehension, suggesting that improved auditory comprehension is driving the overall improvement of comprehension. Breaking auditory comprehension down further into auditory word, sentence and paragraph comprehension did not reveal any additional significant differences (word: p = .055, sentence: p = .213, paragraph: p = .4685), possibly due to the reduced number of items each of these scores is based on relative to the whole Auditory comprehension score. However, auditory word comprehension approaches the closest to significance and could be considered to have the largest effect on the overall auditory comprehension change.

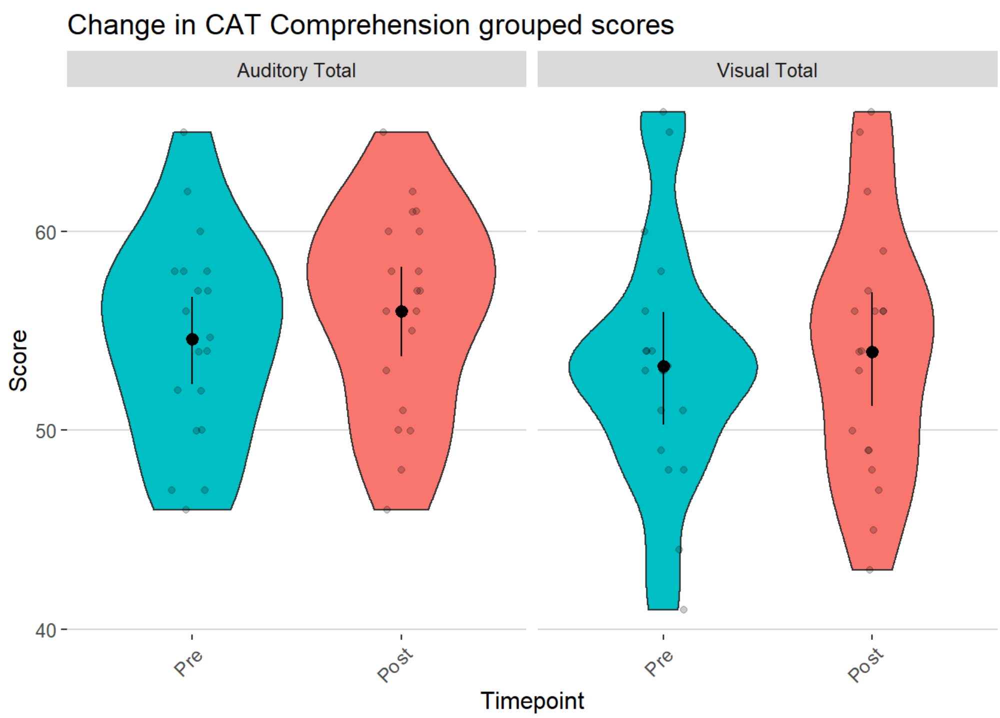

Figure 5.21. A graph to show delayed improvement for PWA T scores on comprehension in the CAT (maximum possible T score: 75)

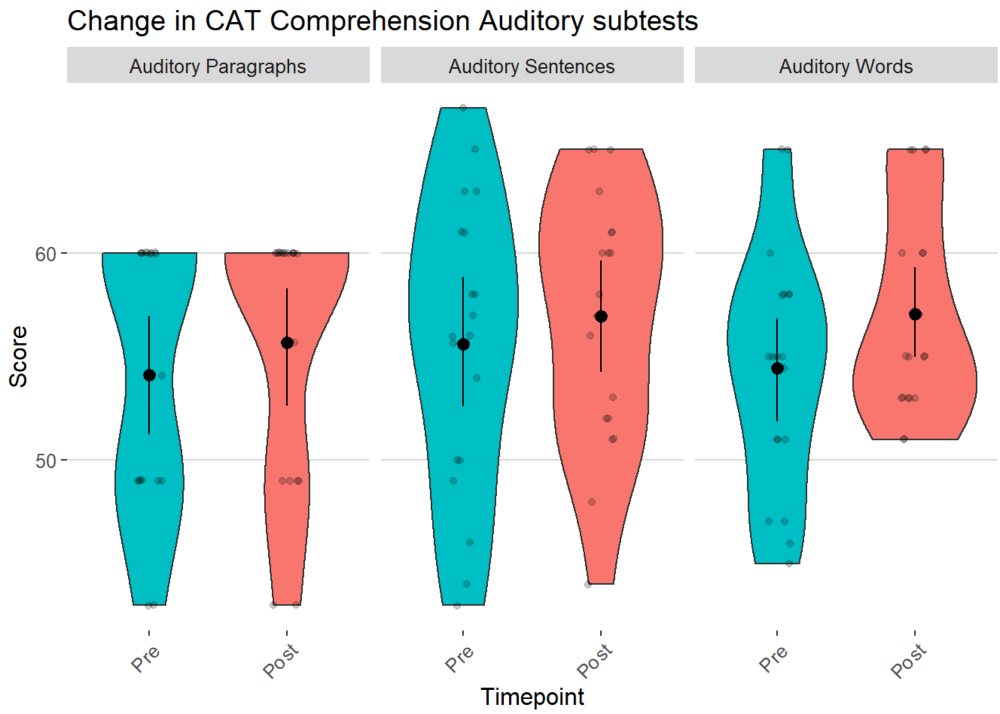

Figure 5.22. A graph to show delayed improvement for PWA T scores on individual auditory comprehension subtests in the CAT (maximum possible T score: 75)

Similarly, Expression was broken into scores for Expressive reading, writing, repetition, and naming. Wilcoxon tests found that reading (p = .187) and writing (p = .887) did not demonstrate significant improvement, however, repetition (p = .049) did, and naming score (p < .001) actually decreased. It therefore appears that repetition was the only aspect of expressive language which experienced any clear generalisation as measured by the CAT from the improvement to named items.

Figure 5.23. A graph to show delayed improvement for PWA T scores on expressive language in the CAT (maximum possible T score: 75)

Table 5.7. CAT T scores used in analyses- columns in bold underwent Wilcoxon tests

	Pre-treatment scores											
	Total								Word-level			
	Comprehension			Expression					Comprehension			Expression
PWA	**Auditory total**	**Visual total**	Average	**Repetition**	**Naming**	**Reading**	**Writing**	**Average**	**Auditory**	Visual	**Average**	Word **Repetition**
AB	**52**	**49**	**50.5**	**46**	**52**	**49**	**47**	**48.5**	**51**	46	**48.5**	**48**
AL	**60**	**54**	**57**	**56**	**62**	**66**	**69**	**63.25**	**58**	50	**54**	**55**
CD	**54**	**53**	**53.5**	**50**	**57**	**52**	**57**	**54**	**55**	55	**55**	**49**
CH	**62**	**56**	**59**	**53**	**62**	**59**	**57**	**57.75**	**65**	65	**65**	**55**
CM	**58**	**58**	**58**	**56**	**56**	**60**	**55**	**56.75**	**51**	55	**53**	**65**
DA	**46**	**44**	**45**	**44**	**48**	**47**	**43**	**45.5**	**55**	45	**50**	**45**
DF	**50**	**51**	**50.5**	**50**	**59**	**65**	**57**	**57.75**	**58**	51	**54.5**	**53**
DM	**50**	**51**	**50.5**	**47**	**64**	**51**	**55**	**54.25**	**55**	65	**60**	**55**
Ebo	**58**	**53**	**55.5**	**62**	**58**	**71**	**65**	**64**	**58**	55	**56.5**	**65**
JE	**57**	**66**	**61.5**	**59**	**72**	**71**	**58**	**65**	**55**	65	**60**	**57**
JP	**47**	**53**	**50**	**48**	**61**	**71**	**62**	**60.5**	**45**	59	**52**	**50**
MH	**56**	**48**	**52**	**44**	**45**	**48**	**42**	**44.75**	**58**	53	**55.5**	**42**
NB	**58**	**54**	**56**	**52**	**58**	**53**	**69**	**58**	**60**	51	**55.5**	**47**
NC	**57**	**60**	**58.5**	**55**	**64**	**62**	**55**	**59**	**47**	55	**51**	**52**
PD	**N/A**	**41**	**41**	**50**	**54**	**52**	**45**	**50.25**	**47**	41	**44**	**56**
PT	**47**	**48**	**47.5**	**51**	**50**	**55**	**50**	**51.5**	**51**	49	**50**	**57**
Shi	**65**	**65**	**65**	**55**	**62**	**67**	**64**	**62**	**65**	65	**65**	**52**
WT	**52**	**54**	53	**47**	**55**	**55**	**57**	**53.5**	**46**	50	**48**	**47**
Mean	**54.65**	**53.22**	53.94	**51.39**	**57.72**	**58.56**	**55.94**	**55.9**	**54.31**	54.17	**54.24**	**51.39**

Table 5.8 CAT T scores used in analyses- columns in bold underwent ilcoxon tests

	Pre-treatment scores					Individual Subtests				
	Sentence-level Comprehension			Paragraph-level Comprehension	Expression					
PWA	**Auditory**	Visual	**Average**	**Auditory**	**Sentence Repetition**	**Semantic Memory**	**Word Fluency**	**Gesture Repetition**	**Digit Repetition**	**Nonword Repetition**
AB	**54**	53	**53.5**	**49**	**39**	**60**	**51**	**47**	**50**	**46**
AL	**63**	57	**60**	**49**	**63**	**60**	**62**	**60**	**55**	**49**
CD	**56**	51	**53.5**	**49**	**53**	**60**	**59**	**47**	**50**	**53**
CH	**61**	54	**57.5**	**60**	**53**	**60**	**60**	**68**	**50**	**53**
CM	**63**	59	**61**	**60**	**48**	**60**	**57**	**60**	**50**	**58**
DA	**44**	44	**44**	**43**	**39**	**51**	**51**	**43**	**43**	**49**
DF	**49**	51	**50**	**49**	**48**	**60**	**54**	**68**	**40**	**53**
DM	**50**	47	**48.5**	**49**	**39**	**60**	**70**	**55**	**43**	**53**
Ebo	**58**	53	**55.5**	**60**	**59**	**60**	**55**	**60**	**55**	**67**
JE	**58**	65	**61.5**	**60**	**59**	**60**	**72**	**69**	**55**	**58**
JP	**50**	51	**50.5**	**60**	**50**	**60**	**58**	**51**	**46**	**46**
MH	**56**	45	**50.5**	**60**	**48**	**47**	**51**	**45**	**43**	**46**
NB	**57**	56	**56.5**	**60**	**59**	**60**	**60**	**47**	**50**	**53**
NC	**67**	61	**64**	**49**	**59**	**60**	**63**	**60**	**59**	**49**
PD	**43**	42	**42.5**	**N/A**	**48**	**47**	**50**	**55**	**46**	**46**
PT	**46**	47	**46.5**	**60**	**53**	**60**	**45**	**55**	**35**	**51**
Shi	**65**	64	**64.5**	**60**	**63**	**60**	**58**	**51**	**55**	**51**
WT	**61**	57	**59**	**43**	**48**	**51**	**53**	**55**	**50**	**46**
Mean	**55.61**	53.17	**54.39**	**54.12**	**51.56**	**57.56**	**57.17**	**55.33**	**48.61**	**51.5**

Table 5.9. CAT T scores used in analyses- columns in bold underwent Wilcoxon tests

	Post-treatment scores											
	Total								Word-level			
	Comprehension			Expression					Comprehension			Expression
PWA	**Auditory total**	**Visual total**	**Average**	**Repetition**	**Naming**	**Reading**	**Writing**	**Average**	**Auditory**	Visual	**Average**	**Word Repetition**
AB	**55**	**49**	**52**	**50**	**50**	**55**	**50**	**51.25**	**65**	48	**57.5**	**52**
AL	**58**	**56**	**57**	**58**	**58**	**65**	**64**	**61.25**	**53**	63	**53.5**	**56**
CD	**53**	**48**	**50.5**	**52**	**52**	**54**	**58**	**54**	**65**	48	**55.5**	**49**
CH	**60**	**56**	**58**	**54**	**54**	**65**	**57**	**57.5**	**60**	48	**57.5**	**55**
CM	**56**	**56**	**56**	**52**	**52**	**71**	**55**	**57.5**	**53**	48	**59**	**55**
DA	**48**	**45**	**46.5**	**43**	**43**	**47**	**46**	**44.75**	**53**	39	**49.5**	**44**
DF	**51**	**53**	**52**	**47**	**47**	**67**	**55**	**54**	**55**	39	**53**	**50**
DM	**50**	**50**	**50**	**50**	**50**	**53**	**53**	**51.5**	**53**	48	**53**	**55**
Ebo	**60**	**56**	**58**	**57**	**57**	**71**	**64**	**62.25**	**55**	63	**54**	**60**
JE	**62**	**66**	**64**	**64**	**64**	**71**	**69**	**67**	**60**	63	**62.5**	**65**
JP	**50**	**54**	**52**	**50**	**50**	**64**	**58**	**55.5**	**51**	48	**53**	**50**
MH	**57**	**47**	**52**	**47**	**46**	**48**	**44**	**46.25**	**60**	48	**54.5**	**47**
NB	**61**	**59**	**60**	**52**	**52**	**57**	**65**	**56.5**	**65**	63	**60**	**47**
NC	**61**	**62**	**61.5**	**60**	**60**	**58**	**56**	**58.5**	**55**	63	**57**	**65**
PD	**46**	**43**	**44.5**	**53**	**53**	**52**	**45**	**50.75**	**51**	56	**46.5**	**56**
PT	**57**	**49**	**53**	**55**	**55**	**54**	**49**	**53.25**	**55**	56	**54**	**52**
Shi	**65**	**65**	**65**	**58**	**58**	**64**	**64**	**61**	**65**	63	**65**	**55**
WT	**58**	**57**	**57.5**	**52**	**52**	**62**	**58**	**56**	**53**	63	**54**	**52**
Mean	**56**	**53.94**	**54.97**	**53**	**52.94**	**59.89**	**56.11**	**55.49**	**57.06**	53.94	**55.5**	**53.61**

Table 5.10. CAT T scores used in analyses- columns in bold underwent Wilcoxon tests

	Post-treatment scores									
	Sentence-level					Individual subtests				
	Comprehension			Paragraph-level Comprehension **Auditory**	Expression **Sentence Repetition**					
PWA	**Auditory**	Visual	**Average**			**Semantic Memory**	**Word Fluency**	**Gesture Repetition**	**Digit Repetition**	**Nonword Repetition**
AB	**53**	48	**50.5**	**60**	**48**	**48**	**53**	**55**	**43**	**53**
AL	**61**	56	**58.5**	**60**	**63**	**60**	**64**	**60**	**59**	**53**
CD	**52**	50	**51**	**49**	**48**	**60**	**60**	**60**	**50**	**58**
CH	**60**	56	**58**	**60**	**48**	**60**	**66**	**68**	**50**	**58**
CM	**60**	54	**57**	**49**	**48**	**60**	**58**	**60**	**50**	**49**
DA	**48**	38	**43**	**49**	**39**	**47**	**42**	**47**	**43**	**38**
DF	**52**	45	**48.5**	**43**	**39**	**60**	**58**	**68**	**43**	**49**
DM	**51**	42	**46.5**	**43**	**48**	**60**	**62**	**51**	**43**	**58**
Ebo	**65**	47	**56**	**49**	**63**	**60**	**57**	**55**	**50**	**53**
JE	**63**	65	**64**	**60**	**63**	**60**	**75**	**68**	**59**	**62**
JP	**51**	54	**52.5**	**60**	**48**	**60**	**55**	**68**	**50**	**46**
MH	**56**	46	**51**	**60**	**48**	**48**	**49**	**55**	**43**	**44**
NB	**60**	60	**60**	**60**	**63**	**48**	**62**	**55**	**55**	**49**
NC	**65**	62	**63.5**	**60**	**63**	**60**	**62**	**60**	**59**	**53**
PD	**44**	44	**44**	**60**	**56**	**48**	**49**	**47**	**46**	**49**
PT	**58**	47	**52.5**	**60**	**56**	**48**	**48**	**55**	**55**	**67**
Shi	**65**	64	**64.5**	**60**	**63**	**60**	**60**	**51**	**59**	**58**
WT	**61**	57	**59**	**60**	**63**	**60**	**56**	**55**	**50**	**46**
Mean	**56.94**	51.94	**54.44**	**55.67**	**53.72**	**55.94**	**57.56**	**57.67**	**50.39**	**52.39**

3.4. Did significant therapy effects in naming accuracy generalise to functional language and communication measures?

Our final question was concerned with whether or not the significant lexical therapy effects achieved would lead to more generalised gains in the functional communication skills of the participants. Several measures of functional communication were obtained before and after treatment in order to capture any potential improvement. The Amsterdam- Nijmegen everyday language test (ANELT; Blomert, Kean, Koster & Schokker, 1994) and Scenario Test (van der Meulen, van de SandtKoenderman, Duivenvoorden, & Ribbers (2010); Hilari, Galante, Huck, et al. (2018)) were implemented at baseline and post-therapy time points to compare measures of communication ability in everyday situations. Samples of discourse elicited through composite picture description were taken using the 'Cookie Theft Scene' from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass, Kaplan, & Barresi (2001), the 'Park Scene' from the Western Aphasia Battery (WAB; Kersetz, 1982) and the 'Living Room Scene' from the Comprehensive Aphasia Test (CAT; Swinburn, Porter and Howard, 2005) pictures) and were used to examine potential changes in connected speech ability.

Wilcoxon signed rank tests were again used to investigate any potential improvement in the ANELT and Scenario Test, as both of these tests produce one or two total scores, summarising communication ability in everyday situations. The Scenario Test produces a single overall score, while the ANELT produces a score for both Meaningfulness (Ability to communicate meaningful information) and Intelligibility (Clarity/ intelligibility of pronunciation). As expected, there was no overall difference in Intelligibility score pre- and post-treatment. For direct measures of ability to produce informative language in an everyday situation, the tests were split. No significant difference was found between pre- and post- treatment Scenario Test scores ($p = .073$), while a significant difference was found in the ANELT Meaningfulness score ($p = .008$). This provides some limited evidence for some generalisation to functional language measures. However, as shown in Figure 5.24 and Figure 5.25, any improvement to functional communication as a result of this generalisation, while potentially significant, appears to be small in size. The mean Scenario Test score increased by two points from pre- to post-treatment (42.33 to 44.33, from a maximum possible score of 54), while the mean ANELT Meaningfulness score increased also by around two points (24.11 to 26.33, from a maximum possible score of 40). Overall, seven PWA improved by at least four points on the ANELT meaningfulness score over the course of the study, while five PWA did the same for the Scenario Test. The highest improvements on the Scenario Test were participants JP and PD, who improved by 11 points each, while the highest improvements on the ANELT were participants AL and CM, who improved by 7 and 6 points respectively, demonstrating improvements in ability to communicate the correct, meaningful information in a described situation, as shown in the following examples:

CM

Tester: "You are now at the dry cleaners. You have come to pick this up and you get it back like this [present shirt with scorch mark]. What do you say?"
Pre-treatment: "I'm sorry, that was produced before"
Post-treatment: "What, uh, sorry but there's, you left a mistake on the front"

Tester: "The kids on the street are playing football in your front garden. You have asked them before not to do that. You go outside and speak to the boys. What do you say?"
Pre-treatment: "Um…I'm telling them they must keep…keep attention, uh uh uh got a mother at home and she's very old"
Post-treatment:"'Scuse me, please, this is my space"

Tester: "You are in the chemist and you find this [present glove] lying on the floor. What do you say?"
Pre-treatment: "I'll ask and, is this… perhaps somebody dropped them"
Post-treatment: "I can, I want to put, what are these? 'Max protection'? I asked them to take these, remove them"

AL

Tester: "You are now at the dry cleaners. You have come to pick this up and you get it back like this [present shirt with scorch mark]. What do you say?"
Pre-treatment: "Oh, good grief, um, I refuse to have that. I either want a new one or I you you get that mark, Iron mark isn't it? Someone's been ironing it too hot."
Post-treatment: ""Excuse me, mistress, I think you've definitely made a mistake on my shirt, so can you please do it again? Thankyou"

Tester: "You go to the shoemaker with this shoe. [present shoe] There is a lot wrong with this shoe, but for some reason you want him to repair only one thing. You may choose which one. What do you say?"
Pre-treatment: "I would like the heels done on my shoes please"

Post-treatment: "Excuse me mister, can you sew the lining of my- it's not really sew, it's like, glue the lining of my shoe on please?"

Tester: "You have an appointment with the doctor, but something else has come up. You call the surgery and what do you say?"
Pre-treatment: "Hello, I've got an appointment with such a doctor at such a time, and I've decided I need to see him about two things instead of the one thing please"
Post-treatment: "Hello receptionist, I have got an appointment but I can't make it. Can you please arrange another one for me?"

Spearman's correlations were then performed between the changes in ANELT and Scenario Test scores and improvement on trained items for each treatment, in order to gain an indication of which treatments may have been driving this generalisation. The only treatments which had a correlation of moderate strength with these measures were the Semantic and Phonological treatments, both of which were moderately correlated with improvement in the Scenario Test score (Semantic: r = .407, Phonological: r = .45). This may suggest a greater role of these treatments in generating the limited generalisation shown than the other treatments.

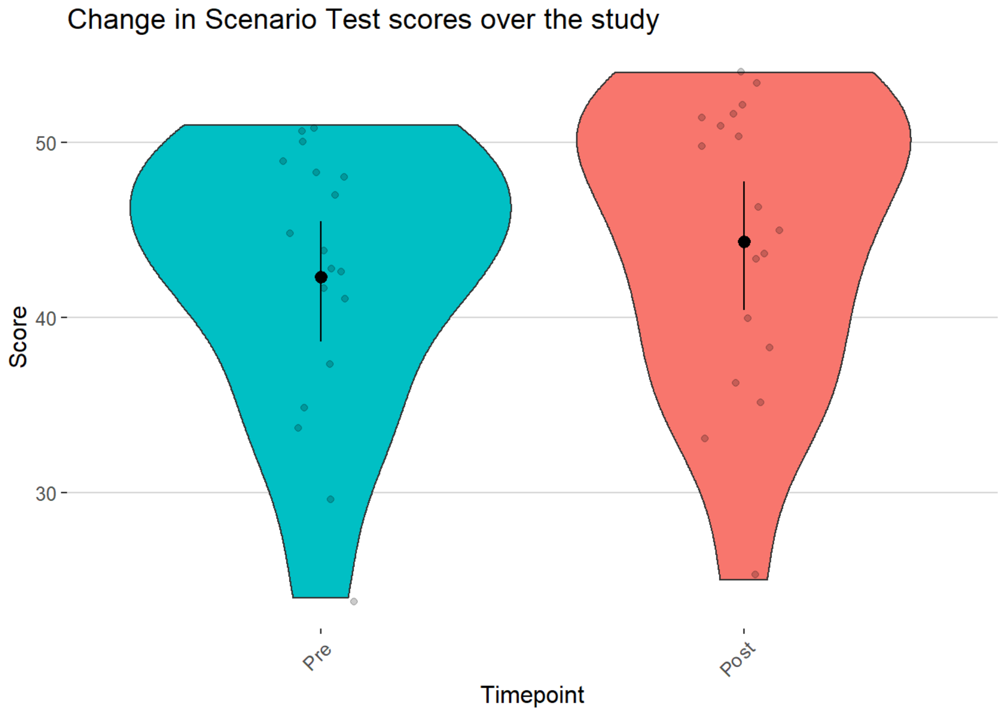

Figure 5.24. Delayed improvement to PWA Scenario Test scores (out of a maximum of 54)

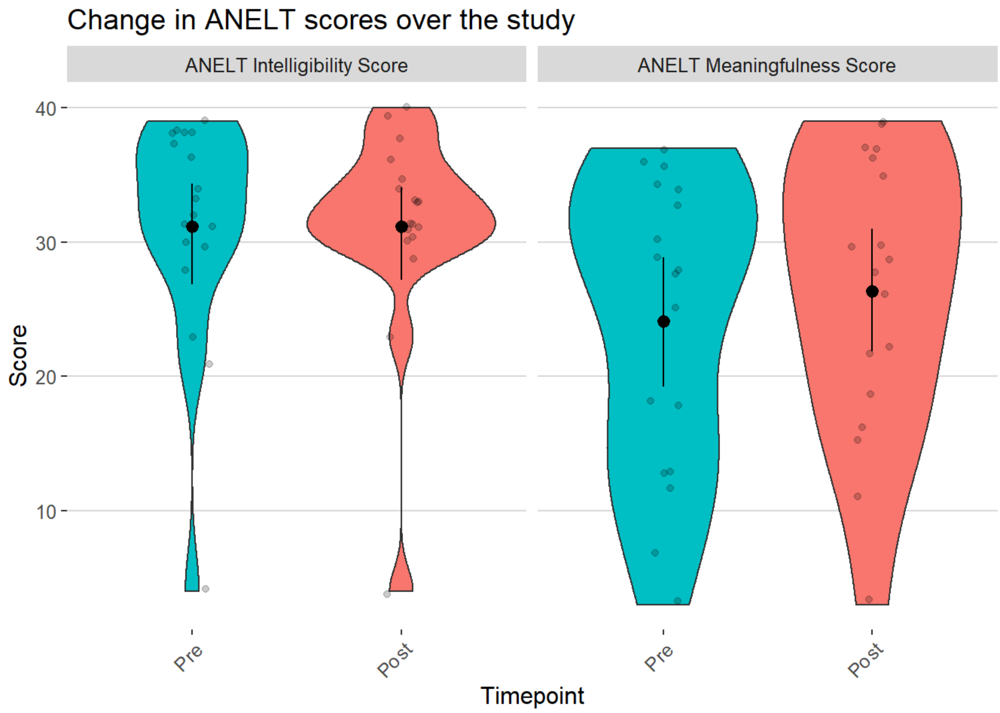

Figure 5.25. PWA ANELT Meaningfulness and Intelligibility scores over the delayed improvement timeframe -pre-treatment to post-treatment (out of a maximum of 40)

Several measures were taken from the Composite Picture Description results – an overall Mean Length of Utterance (MLU), Type to Token Ratio (TTR- proportion of unique words used), and Words Per Minute (WPM) were calculated for each PWA. Changes in these measures were surprisingly uniform, with 14 of 18 participants demonstrating increased WPM and a small subsequent reduction in TTR., indicating that the majority of participants improved in terms of speed of language production, but may have produced slightly more redundant speech as a result. MLU change was more varied, with 13 PWA increasing their MLU over the course of the study, and the remaining 5 decreasing it. Overall, a range of PWA improvement 'profiles' were demonstrated. The majority of PWA (10) improved on WPM and MLU, but suffered a decreased TTR. One PWA dropped on both WPM and TTR, but increased their MLU. Three participants scored a higher WPM, but lower TTR and MLU in their post-study tests, while two PWA improved on both WPM and TTR, but decreased on MLU. Finally, two PWA increased their scores on all three measures. In these participants, the improvement in ability to quickly communicate relevant information was most noticeable, as shown in participant NB's responses to the Living Room Scene:

Pre-treatment:

"Child is sitting with the father on the floor, he's holding a toy car. Uhh the father is a str- a streak no shoes and socks on a cup of coffee on the coffee table. The cat is uhh leaning into the fish bowl and that is staying on the top shelf but he's knocked some blocks, blocks, no books over the father. The uhh secosismund is very good. Ok?

Time taken: 2 minutes 26 seconds

Post-treatment:

"The baby is sitting on the floor with a toy car. The man is getting the feet up on the table and the coffee is there. He's…. he's looked asleep. The cat Is just uhh knocking the books off the top shelf. Uhh The cat is uhh the cat is… just needing into the fishtank. The radio and stereogram is just on the bottom shelf. The plant is on the top shelf."

Tester: "And under there?"

"There is a book under there."

Time taken: 1 minute 51 seconds

Table 5.11. PWA scores pre- and post-treatment for the ANELT and Scenario Test

PWA	Pre-treatment ANELT Meaningfulness	ANELT Intelligibility	Scenario Test	Post-treatment ANELT Meaningfulness	ANELT Intelligibility	Scenario Test
AB	12	30	43	16	30	43
AL	30	38	50	37	40	51
CD	18	23	42	22	31	50
CH	28	36	48	28	33	51
CM	13	28	41	19	31	36
DA	13	21	30	15	23	33
DF	28	37	45	30	31	44
DM	18	30	37	22	29	35
Ebo	34	38	49	36	30	50
JE	36	38	44	39	39	45
JP	33	38	43	30	35	54
MH	3	4	34	3	4	40
NB	34	32	51	35	36	52
NC	36	31	48	39	34	53
PD	29	34	35	26	33	46
PT	7	39	24	11	38	25
Shi	37	31	51	37	33	52
WT	25	33	47	29	31	38
Mean	24.11	31.17	42.33	26.33	31.17	44.33

Table 5.12. Changes in measures of communication for the composite pictures from pre- to post-treatment. Dark green and red denote larger (>1) increases or decreases in each measurement, while light green and red denote smaller (>1) increases or decreases in each measurement.

PWA	WPM Change	TTR Change	MLU Change
AB	9.153696	-0.08	3.125
AL	12.54285	-0.01516	5.555556
CD	9.151807	0.021429	1.8
CH	9.521078	-0.02233	3.687179
CM	-0.12293	-0.02099	6.035714
DA	4.511867	0.065028	-0.38319
DF	15.26881	-0.09341	0.013587
DM	11.70598	-0.10747	1.1
Ebo	21.44997	-0.03445	10
JE	66.95638	-0.01391	-3.2037
JP	11.27542	-0.03387	2.47619
MH	14.58707	-0.01174	6.883929
NB	4.93	0.053333	1.436667
NC	9.976667	-0.1	-1.49667
PD	14.79258	0.04263	5.052885
PT	31.08333	-0.22667	8.886667
Shi	11.54935	-0.01663	-1.22917
WT	2.327742	-0.01663	-3.35714
Mean	14.4812	-0.02666	2.576861

Discussion

Treatment Effectiveness

The first set of questions we aimed to address in the Results focused on the effectiveness of each and all of the six treatments in improving picture naming for both trained and untrained stimuli. We performed analyses examining effectiveness of each individual treatment, as well as comparing the effectiveness of the treatments against each other, in both the short and the long term.

Firstly, in order to determine which treatments were effective in improving picture naming for the items they trained, Wilcoxon tests were used to determine if there was a significant difference between PWAs' pre-therapy (maximum 3 weeks before beginning of treatment) and immediate (within 2 weeks of treatment completion) post-therapy scores. As expected, a significant difference was found for all the treatments at the group level, indicating that they were all effective in improving trained item naming. This is in line with many similar behavioural treatment approaches to picture naming which have consistently and robustly found similar effects demonstrating the effectiveness of all these treatment types in improving PWA naming of trained items (Nickels et al, 1992; Best et al, 2013). Gesture-based treatment is probably the most seldom used of these treatments, but still has some examples showing its effectiveness (Akhavan et al., 2017; Gili et al., 2017). However, these results also have an important secondary function in confirming the validity of the specific methodology used in this study. While all six treatments have been previously examined using various treatment methods, these results demonstrate that they remain effective when using the novel bespoke methods developed for this study. This study used a progressive hierarchy prompt methodology, where participants can progress from decreasing to increasing to self-created prompts over the treatment, which was novel for all the treatments. The Alternative Modality treatments in particular (Gesture and Prosody) were not traditionally prompt hierarchy-based. The methodology also featured a larger proportion of both teletherapy and self-directed practice than previous studies in this field. The fact that all six treatments were still effective while using this novel, standardised methodology demonstrates that the methodology used, as well as the timeframe, intensity and volume of treatment was appropriate in facilitating a significant improvement in trained item naming. This novel methodology was effective in ensuring each treatment facilitated improvement, allowing direct comparison of the different treatments and the inter-participant factors affecting their usefulness.

Question 1.2 was focused on comparison of the different treatments using likelihood ratio tests and subsequent pairwise analyses on CLMMs built on improvement scores for each item (rather than total or mean scores for each PWA), including treatment as a factor , as well as ANOVAs comparing the effect of timepoint and treatment on overall participant scores for each treatment. A significant difference was found between the effects of the different treatments on naming of treated items using likelihood ratios and subsequent pairwise analyses. Specifically, the two executive function treatments- Speeded and Interfered naming- were found to have significantly higher improvement scores than the other four treatments, with Speeded in particular performing very well in comparison to the others. A significant interaction was also found between timepoint and treatment in the 6x2 ANOVA, suggesting a difference in treatment effects. Subsequent 2x2 ANOVAs found that the only comparisons containing significant interactions were Speeded with every other treatment, and Interfered with the Prosody and Phonological treatments, supporting the finding that these two treatments, and Speeded in particular, outperformed the other treatments. Looking at the distribution of overall participant score improvement (see Figure 5.7), this improved performance does not appear to be uniform- all the treatments have some PWA who experience little to no improvement. However, Speeded naming in particular appeared to facilitate much greater improvement in some PWA than others, with seven participants improving by more than 20 points (from a total score of 60), compared to a maximum of two PWA achieving this level of improvement in any of the other treatments. It seems that while all treatments struggle to facilitate substantial improvement for all PWA, the Speeded treatment facilitates a large improvement for more PWA - the PWA that respond well to it seem to respond very well relative to other treatments.

Looking again at Figure 5.7, we can see similar distributions of improvement for the Semantic, Prosody and Gesture treatments, with Semantic having slightly, but not significantly, higher scores, as well as a slightly higher mean. This is reflected in the pairwise comparisons, showing no significant differences between these three treatments. While the Phonological treatment did not significantly differ from these three in overall improvement, Figure 5.7 shows more consistent improvement when looking at PWA totals - in this treatment, all the PWA showed at least a small amount of improvement, but conversely, none showed a very large improvement- no PWA improved by 20 or more points in this treatment. The Interfered naming treatment performed significantly better than the above four treatments, and did not significantly differ from the Speeded condition, although it appears to have an overall slightly lower mean improvement when comparing PWA total improvements. Similarly to the Phonological treatment, it also seems to display more consistent improvement between PWA- while there are some cases of no/very little improvement and very large improvement, a large number of PWA improvement scores are clustered tightly around the mean. The Speeded naming treatment differed significantly from the Phonological, Semantic, Prosody and Gesture treatments, seemingly due to the differences discussed in the above paragraph. Overall, it is clear that the executive function treatments were the generally strongest options when considering improvement of trained items only, while there are some differences in distribution of overall PWA improvement between these two treatments. Interfered naming appears to show more consistency in PWA improvement, while Speeded naming has more PWA showing a very large level of improvement.

Speeded and Interfered naming are both relatively novel treatments in comparison to the more well-established model-oriented semantic and phonologically-based treatments. The fact that both the executive function-based treatments showed significantly better improvement than the other treatments may indicate that the co-targeting of executive functions and language processing skills was what was driving the increased effectiveness of these treatments. While Speeded naming is theoretically based in improving PWA speed of retrieval, and thus overall fluency (Conroy et al., 2018), it may be that improving executive functioning removes a significant limitation in language expression more generally for some PWA, and removal of this limiting factor allows for greater improvement than that facilitated by targeting a more specific aspect of language. Moreover, it may especially improve semantic processing over and above executive processing (Bruehl et al., submitted).

Alternatively, the shorter format used for the Speeded and Interfered treatments may have played a role in their success. The use of speeding and interference in these treatments necessitated a reduced focus on prompts, meaning that, in general, participants may have progressed through the items faster than the other treatments. This potential faster progression would allow for an increased number of repetitions of each item than the other treatments, as treatment volume was controlled by time spent on each treatment per week, rather than number of item repetitions. Increasing the number of repetitions may therefore also be important in maximising PWA improvement in trained items (Fillingham et al., 2005). A more prosaic explanation may have been more motivation and engagement with the executive therapies, where an element of pressure and gaming could lead to enhanced alertness; future research may be able to disentangle a general engagement factor from specific effects of greater difficulty central to the speeded and interfered treatments (look up ref for enhanced engagement for software – maybe a Leff one again!).

We also found a similar pattern when comparing pre-therapy scores to delayed post-therapy scores, which were recorded 2-4 months after treatment completion in order to observe the longer-term effects of the therapies. All six treatments retained a significant level of improvement from pre-treatment scores, and mean PWA score changes seemed to show Speeded and Interfered naming as the most effective overall - a mean improvement of 4.93 points for Speeded and 4.69 points for Interfered naming, in comparison to 3.6 for Phonological, 2.35 for Semantic, 2.7 for Gesture and 3.69 for Prosody. It therefore seems clear that, when comparing pre- and post-therapy naming scores in both the short and long term, the two executive function treatments - Speeded and Interfered naming- were the most effective overall.

We also investigated the difference between immediate and delayed post-therapy scores for each treatment, to get an idea of the relative amounts of decay post-treatment for each of the six treatments. The Speeded, Interfered, Gesture, Semantic and Phonological treatments displayed significantly lower scores at the delayed post-therapy timepoint compared to immediately post-therapy, suggesting that these treatments underwent more substantial decay in treatment gains than the Prosody treatment, which did not show a significant difference in scores between immediate and delayed post-treatment scores. We might therefore conclude that, in the long term, the reduced rate of decay of treatment gains might result in the Prosody treatment being the most beneficial. However, the amount of time between the immediate and delayed post-treatment tests (at least two months) was selected to allow the treatment gains to stabilise- rather than the decay continuing longer-term, we would expect any gains that have been maintained two months after treatment completion to be unlikely to be lost after this point.

It is also unclear what the exact cause of the lack of a significant difference between the immediate and delayed post-treatment scores for the Prosody treatment is. While we might attribute this to this treatment facilitating more robust treatment gains than the others, which are more easily maintained in the long term, this treatment also experienced a significantly lower initial increase in scores (from pre-treatment to immediately post-treatment) than either of the executive function treatments. This does not explain the reduced decay of treatment gains relative to the Gesture, Semantic and Phonological treatments; however, it does make it unclear whether the reduced decay of the Prosody treatment gains relative to the executive function treatments was due to a possible increased robustness of gains facilitated by the alternative modality treatments, or simply due to the initial improvement being lower for the alternative modality treatments, meaning that there were less gains necessary to maintain, making that maintenance more likely.

Therefore, we would argue that, despite the lower amount of decay present in the Prosody treatment, the overall higher improvement in scores at both the immediate and delayed post-therapy timepoints for the Speeded and Interfered treatments would still indicate that they are most effective in improving trained item scores in both the short and long term.

Inter-participant variation

The second set of questions we aimed to address in this study were focused on the differences between PWAs in terms of their responses to each treatment, and the factors that could be driving these differences. Once again, we focused on examining the factors individually before investigating how they combine to influence PWA improvement in different treatments using CLMMs and likelihood ratio tests.

Overall, the inter-participant factors which seemed to have the most impact on improvement in trained items seemed to be previous language ability - phonological, and semantic ability as measured by their respective Dell scores. In terms of demographic factors, Age seemed to play the biggest role in predicting PWA ability to improve - in fact, the model including only age was more beneficial than the model containing all three demographic factors. This suggests that most of the inter-participant variation explained by time poststroke and years in education could also be explained by age only- the factors overlap in terms of what inter-participant variation they explain. Alternatively, years in education and time poststroke may just not have a large effect on PWA improvement- this is supported by the fact that these two factors had the z values closest to zero of all the included. Similarly, the model involving only the Dell s and p scores was more effective in explaining the results than the one including all factors. This information may be helpful in terms of optimisation of future studies- for example, it may be unnecessary to include multiple demographic measures which don't add any additional explanatory power if the age of all participants is known. In future studies, using age only as a predictor may be more efficient in retaining most of the benefits of using all the demographic factors included here, but for less expenditure of participant and researcher time and energy, if factors such as time poststroke and years in education are not readily available. However, the lack of effect of time poststroke and years of education in our findings may also be a result of lack a variation in these factors within our PWA sample- in a more varied sample, any effect of these factors would be easier to untease.

Previous literature in this area has shown a large effect of a variety of inter-participant factors on therapeutic effects (Crinion and Leff, 2015; Plowman et al., 2012; Tippett, 2015). Therefore, being able to optimise treatment for each PWA based on their preexisting attributes is an exciting prospect for the future of speech and language therapy. One of the goals of this set of questions was to determine how and to what extent each treatment was affected differently by different inter-participant effects. Especially considering the dominance of the executive function treatments in the first set of questions, one key area of interest was whether or not we would find enough inter-participant variation between different treatments to justify switching treatments depending on the participant to maximise gains, or if the data pointed towards the most effective solution being executive function treatments being used across the board. Overall, we found little difference between the different models in terms of pairwise comparisons between different treatments, making it hard to justify a large amount of 'personalisation' of treatment. Speeded and Interfered naming were generally found to be the strongest factors in each model, regardless of whether or not they accounted for a range of different inter-participant factors. However, the negative interactions between the Interfered naming improvement effect and several inter-participant factors (Years in education, age, months poststroke, and lesion volume) suggest that this treatment may be more consistent and resilient to inter-participant variation than Speeded naming, possibly making it the stronger choice in PWA with worse initial outlook for improvement, and Speeded naming the stronger choice for PWA with good initial outlook for improvement.

While there was variation at the individual level in terms of which treatment was the most successful for each PWA, we struggled to find enough consistent patterns in the data to justify changing treatments depending on individual demographic or language ability factors, other than Interfered naming appearing to be the treatment most resilient to inter-participant differences, possibly marking it as a stronger option for A with a profile suggesting they are unlikely to progress as much, and a weaker option for PWA with a profile suggesting they can progress greatly. While this would still be an interesting topic for future research, as the presence of individual differences does suggest that, at a certain 'threshold' for different factors, changing to alternate treatments may be beneficial, this threshold is likely quite extreme, as we broadly found the executive function treatments to be most effective regardless of which factors were accounted for.

Generalisation beyond treated items

Finally, a third set of questions aimed at evaluating the generalisation to language abilities outside of the specific trained items for each approach. While improvement in trained items is important for providing a finer grained understanding of the effects of each specific treatment, and how improvement in each treatment differs, it is not the ultimate practical goal of any speech and language treatment. The aim of using speech and language treatments - which is sometimes overlooked in research in this area- is to improve a PWA's language skills and functional communication in order to help them navigate everyday life effectively and thus improve their quality of life. We would expect these skills to broadly correlate to treated item improvement; however, it is important to investigate them separately in order to gain an understanding of which approaches would be most beneficial in practical use by SLTs.

While not all tests were performed at every interim timepoint in order to reduce PWA fatigue and, ultimately, attrition, data for both naming untrained items and semantic and phonological ability (derived from the types of errors produced by PWA during naming of trained and untrained items) were produced at every timepoint, allowing an investigation of generalisation to these measures for each approach. In general, the improved naming of trained items in treatments with the executive function approach was mirrored in the results for semantic and phonological scores. Pairwise comparisons found a significant difference between executive function and the other two approaches when comparing general changes in Dell s and p scores. Tests examining each approach and factor separately, this difference appeared to be driven by semantic ability specifically- a significant improvement in Dell s score was found between pre- and immediately post-treatment scores for executive function treatments, but not either of the other two approaches. The executive function approach therefore seems most well equipped to improve semantic ability in PWA generally, as well as trained items specifically. However, when looking at number of errors as opposed to Dell score, the alternative modality treatments also showed a significant difference in number of semantic errors from pre- to immediately post-treatment. There is therefore some evidence for alternative modality treatments producing a disproportionate generalisation to semantic ability specifically, relative to their effect on trained items. Overall, though, the greater effectiveness of executive function treatments in facilitating improvement in trained items is reflected in improvement in semantic ability, but not phonological. In terms of the Gesture treatment, there is some evidence that different co-speech gestures can aid in increasing semantic processing, as sensorimotor information and memory regarding the item can be processed concurrently with the auditory language information in some regions, facilitating overall semantic processing of the item (Dick et al., 2009; Andres et al., 2008). This could partially explain the evidence of disproportionate generalisation in Gesture, but its cause in Prosody is still unclear.

As mentioned earlier, a significant difference in treatment effectiveness in facilitating trained item naming improvement was found regardless of whether or not a variety of inter-participant factors - including pre-treatment semantic and phonological ability - were accounted for. As the superior performance of the executive function treatments was also found to generalise to semantic ability, this combination of findings may also provide beneficial information regarding PWA suffering from a specific semantic deficit. Executive function treatments would appear to be the most effective option regardless of their degree of semantic impairment, due to both their robust effect regardless of inter-participant differences in language factors, and this effect being reflected in generalisation to semantic ability. Relatively preserved semantic ability has previously been associated with naming gains in interfered naming (Bruehl et al., submitted). However, interfered naming in this study was also found to be more resistant to certain inter-participant demographic and lesion factors. This, combined with the generalisation to semantic ability exhibited by the executive function treatments, may make it a good choice of treatment for PWA with semantic deficits in spite of its associations with naming gains for PWA with relatively preserved semantic ability. PWA with a semantic deficit, and their treating clinicians, could therefore be reasonably confident in assuming that executive function treatments would allow them to address this specific deficit successfully. For PWA with specific phonological deficits, conversely, the picture may not be so clear regarding generalisation; while executive function treatments are the most effective regardless of their level of phonological deficit, the lack of generalisation to phonological ability makes it unclear if executive function treatments would be the most effective option to address this specific deficit.

As mentioned earlier, in addition to semantic and phonological ability scores, separate pre- and immediately post-treatment scores were also obtained for untrained item naming, judging changes in PWAs' picture naming ability more generally. Analysis of these scores revealed unexpected findings. A significant difference was not found for any individual approach (eg. 3 weeks of treatment using a particular approach was not enough to cause a significant improvement in naming of untreated items), or when comparing different approaches on level of improvement (suggesting a similar level of generalisation to untrained items for each approach). However, a significant difference was found when considering the approaches as a whole, if all treatment periods are accounted for or if overall pre- to post-treatment (eg. before and after all 3 approaches) scores are considered. This suggests a possible cumulative effect of treatment over the long term, despite a possible lack of difference in generalisation between approaches. It is therefore possible that generalisation to naming untrained items may occur regardless of treatment used, but instead be more dependent on the duration of treatment - three weeks of any treatment were not enough to create a significant difference, but nine weeks of treatment across all three approaches were. If this is the case, it may be possible that duration of treatment has been relatively under-considered when judging practical treatment effectiveness, compared to factors like intensity and treatment approach.

In addition to the above measures of generalisation of language gains, we also performed several tests pre-therapy and immediately post-therapy. The Comprehensive Aphasia Test (CAT) was one of the batteries used here, with the aim of gaining an understanding of changes to each aspect of general language ability, as well as non-language cognitive changes, over the course of treatment. One of the advantages of using the CAT, is the ability to gain scores for such a wide range of language and non-language abilities; however, the downside to this is the limited statistical power available when comparing individual factors determined by a single subtest, which is often made up of a limited number of items/questions. This may be the reason why we struggled to find any significant differences when comparing pre- and post-treatment scores for individual subtests, despite finding more clear differences when performing comparisons of combined scores at a more macro level. The generalisation results for the CAT scores were unexpected; overall language comprehension was found to significantly improve, but not overall language expression. The results within the improvement to comprehension were more in line with what we might have expected, with a significant difference in auditory comprehension specifically, which appeared to mostly be driven by individual word comprehension improvement. This change in comprehension ability might be in line with our previous results concerning generalisation to semantic ability- improved semantic ability might allow for improved comprehension. As mentioned earlier, some tests, including the CAT, were not performed in between each three-week treatment period, in order to limit PWA fatigue and attrition. However, for a number of reasons, we can still have some confidence in making assumptions regarding the treatments facilitating generalisation in the tests performed only at the beginning and end of treatment as well. This is due to the richness of the data in terms of our ability to understand the differences between treatments in facilitating trained item improvement, as well as their generalisations to untrained item naming improvement and semantic and phonological ability changes. Finally, our understanding of the theoretical basis of each approach, also contributes to this confidence. In this case, assuming that the improved comprehension scores and improved semantic ability are linked, we could therefore assume that the improved comprehension scores were also primarily due to executive function and alternative modality treatments. It also makes sense that auditory comprehension and specifically auditory word comprehension improvements seemed to be driving this change, given that listening to and producing individual words made up a large proportion of the treatment methodology in this study.

The lack of expressive score improvement is harder to explain. The improvement in repetition also makes sense, given that word repetition was also part of the treatment methodology. However, we would have hoped to see generalisation to other measures of language expression ideally, given that language expression is what the treatments are focused on improving. In particular, the reduction in expressive naming scores is particularly surprising, and incongruent with our earlier findings regarding generalisation to untrained words. It could simply be the case that there is less generalisation to general language ability than previous results have indicated. Overall, we found evidence of generalisation to untrained items. However, our evidence of generalisation to untrained tasks, such as those performed in the CAT, was more inconclusive.

The results for tests of functional communication were also mixed. Considering the two measures of sentence-level, situation-based functional communication, while both the ANELT meaningfulness score and the Scenario Test score showed improvement, only the ANELT meaningfulness score change was significant, and overall improvements in functional communication as measured by these tests seem relatively small, regardless of significance. However, when examining individual changes, several PWA improved quite substantially on these measures. The results for compound picture description were similar- while almost all PWA improved in terms of words per minute, whether or not this indicates a notable improvement in everyday communication ability is more unclear. The majority of PWA also had reduced type to token ratios, suggesting that, while they used more words, a reduced proportion of the words used were unique, so the increase in rate of spoken words might not have also led to an increase in information communicated - it may be the case for some participants that they simply used more redundant words in the communication of the same amount of information as pre-treatment. The majority of PWA also improved in mean length of utterance, however. This is an encouraging sign, as an increase in the average length of utterance may indicate that PWAs' utterances also became more complex in structure, allowing the communication of more information within a single utterance. Overall, while the evidence regarding generalisation to functional communication was somewhat unclear, there is enough positive evidence that a notable improvement in everyday communication seems likely for some PWA at least. For example, 3 participants improved in all three measures: words per minute, type/token ratio, and mean length of utterance. At the group level, improvement on measures of functional communication were relatively inconsistent, however.

How our findings relate to each other

In addition to the Results we have studied directly, it is also important to endeavour to understand how these results relate to each other in order to build a larger picture of different treatment effects in different participants in response to different treatments. As we can see from our earlier relating of semantic ability to comprehension, examining the links between the different questions we have investigated can help is infer additional information and build a fuller picture of how each treatment works, including the context of inter-participant factors.

As we were unable to include lesion location in the wider analyses investigating how the different inter-participant factors interact in their effects on improvement. However, it is of particular importance to relate this factor to our other findings and gain a better understanding of the place of lesion location in the larger framework of inter-participant factors. Different clusters were found showing neural correlations to improvement in naming for the Phonological, Prosody and Speeded treatments. Correlations were found between improvement in naming for Phonological treatment items and neural clusters in the left anterior cingulate gyrus (aCG) and paracingulate gyrus. This supports previous findings from Leonard et al. (2015), in which changes in activation were discovered in the left CG during a semantic task following a phonological treatment (PCA), as well as findings by Abel et al. (2015), which found a pre-treatment reduction in activation in the left aCG in PWA relative to controls. Also, Marongolo et al. (2016) reported findings in which tDCS and speech therapy facilitated increased resting state functional connectivity in the left anterior cingulate cortex in addition to word repetition accuracy improvements. The link between improved word repetition accuracy and left aCG activation is of particular note, given that repetition was the only expressive score in the CAT with a significant improvement post-treatment - it may therefore be possible that the Phonological treatment played an important role in this, given its connection to the aCG both in this study and previous studies (Leonard et al., 2015).

The neural clusters linked to improvement in the Prosody treatment were the most strongly correlated, and located in the posterior CG and the posterior ITG. Interestingly, the left or right ITG were found to have increased activation following rTMS to inhibit activity in the left or right hemisphere in combination with speech therapy in Hara et al (2015), a study specifically aimed at exploring activity in language compensation regions in the right hemisphere, which would potentially be targeted by our alternative modality treatments, including our Prosody treatment. This suggests that the ITG may be more relevant in treatments focused on compensation in regions not usually involved in the language network.

The neural cluster associated with improvement in the Speeded treatment was located in the Planum Polare/anterior STG. The STG is associated with high-order auditory processing (Burton et al., 2000; Visser & Lambon Ralph, 2011), which may be the reason why damage to it is more closely associated with a higher-order executive processing-focused treatment such as Speeded naming.

The links between STG damage and Speeded treatment improvement are particularly relevant when examining optimisation- the other inter-participant factors did not demonstrate a great deal of variation between different treatments, leaving the Interfered and Speeded treatments as the overall most effective for the vast majority of cases. However, the link between damage to the STG and Speeded treatment improvement may present a situation in which selecting an appropriate treatment based on participant characteristics is more necessary. Specifically, Speeded may not be the most beneficial treatment in participants who have damage to the left STG. The Speeded treatment is quite less consistent than the Interfered treatment in level of trained item naming improvement success for different participants. This is also demonstrated by the negative interactions between the effect of the Interfered treatment and those of time poststroke, age, years in education, and lesion volume. It is therefore likely that the participants who struggled more with the Speeded condition had STG damage, while those who achieved very large improvements of over 20 points did not. We might therefore expect participants with no STG damage to perform better with Speeded naming treatment. Varying treatment by STG damage level may therefore achieve the best of both treatments- very large improvement for those participants with no STG damage who may perform most effectively with Speeded treatment, while those participants with STG damage would still be likely to achieve a good, close to mean level of improvement via other treatment options. This method of treatment optimisation and personalisation seems promising to explore in future studies.

Another link which has already been discussed to some extent is the relationship between our primary language measure- trained item picture naming- and how exactly its generalisation to our other measures varies depending on inter-participant factors. While we have been able to make some assumptions about the patterns involved in this generalisation, functional communication in particular is hard to decipher. Theoretically, we would expect our executive function treatments, which focus more on practical communication (increasing speed of speech or ability to overcome interruptions and distractors). Given that these are also the most effective treatments for our primary language measure, we could assert with some confidence that they are likely to play a key role in the improvement in functional communication for those participants who improved in this regard. We could also assume that the inter-participant factors which had the largest impact for trained item naming improvement - Age and pre-treatment Semantic and Phonological ability - likely also have the largest effects when considering generalisation to the other language measures. In terms of lesion location, we might expect participants with STG damage to struggle with functional communication in particular, due to its associations with higher-order auditory processing. However, all of the neural clusters found (STG, aCG, paracingulate gyrus, and posterior ITG) could also affect generalisation due to their involvement in language processing. Based on our findings regarding generalisation to untrained items by treatment, it was evident that model oriented and executive function treatments performed proportionally to their trained item equivalents, while alternative modality treatments possibly proportionally outperform their trained item changes when generalising to untrained items. However, there is no clear reason to expect significant variation between our findings for trained items and their generalisation to other language measures, either in terms of variation by treatment or by inter-participant factors.

Disclosure of interest

The authors report no conflict of interest.

Chapter 6: General Discussion

This chapter is split into three main sections. Firstly, I will summarise the key findings across the four empirical previous chapters within this thesis. Then, the key implications of these findings will be discussed, as well as some of the benefits and limitations which should be taken into account when considering these, in order to give an idea of the overall weight and need for future confirmation and exploration of each of these findings. Finally, the directions for future research in terms of confirming and building on the work in this thesis will be considered.

Summary of findings

The focus of this thesis was the evaluation of different behavioural treatments for aphasia and an investigation of the factors- neural, demographic, and cognitive - which affect these treatments. **Chapter 2** contains a systematic literature review which aimed to encapsulate the current state of the literature in this area at the time of writing, and identify areas which would benefit from exploration through further research. This would help direct the rest of the thesis to address these areas of sparsity. Two key areas were identified as requiring more exploration. Firstly, the inclusion of diffusion-weighted scan analysis and assessment of white matter tracts and their relationship to treatment-facilitated improvement warrants investigation. While many studies included structural MRI, fMRI, or rs-fMRI scan analysis, or used neurostimulation methods such as rTMS or tDCS to test theories regarding the neural mechanisms occurring during treatment, there were comparatively few studies focusing on diffusion-weighted scan analysis, despite evidence that structural white matter change is possible even over the length of a treatment study (Schlaug, Marchina, & Norton, 2009; Wan et al., 2014; van Hees et al., 2014). We therefore focused on diffusion-weighted scan analysis in **Chapter 3**. The second key area identified in **Chapter 2** was the lack of direct comparison between different treatments and approaches to treatment. The majority of studies investigating treatments focus on a single treatment, comparing it to a control, which may be no treatment (Efstratiadou, 2018), a generic treatment (Grechuta et al., 2019), or a similar treatment to the experimental one with the theorised active factor removed (Fridriksson et al., 2009). Some studies also investigated a single approach, including two closely-related treatments such as semantic and phonologically focused treatments (Abel et al., 2014). However, direct comparison of disparate treatments using different theoretical approaches was not found. There are many benefits to comparing aphasia treatments to a control: it allows a direct measure of treatment efficacy, and clear, in-depth investigation of how improvement facilitated by the treatment is affected, and affects, neural, language and cognitive factors. However, direct comparison of different treatments is required in order to investigate which treatments are most effective for which PWA, and for which purposes. For example, alternative modality treatments are generally focused on activation of right hemisphere homologues of left hemisphere language regions (Schlaug et al., 2009; van de Sandt-Koenderman et al., 2016; Wan et al., 2014). They may therefore be hypothesised to have a stronger effect on participants with severe left hemisphere damage, in which right hemisphere activation could be more beneficial (Perani et al., 2003; Vitali et al., 2007). Ultimately, we hope that the optimal treatment could be delivered to any given PWA based simply on information about them provided before treatment begins. However, in order to progress towards this goal, direct comparison of different treatments and assessment of how effective they are at improving different measures of language in the same set of participants is required. This is what we aimed to accomplish in **Chapters 4** and **5**, in which a methodology was developed and then used to directly compare 6 different treatments. In addition to these findings, **Chapter 2** also aimed to evaluate and develop the model of mechanisms underlying hemispheric activation increase and reduction relative to both healthy controls and in PWA pre-treatment (in comparison to post-treatment) proposed in Abel et al. (2015), based on evidence from the studies included in the review. An adapted model was developed based on the reasons for region activation changes provided in the studies covered in the review. Some mechanisms, which were not accounted for in any of the studies, were removed, while others were divided to create more clear, distinct categories of mechanisms in cases where many studies fit into a single mechanism category despite slight differences in reasoning regarding changes in activation. This adapted model should therefore provide a clearer, more accurate and practical summary of potential mechanisms underlying changes in activation in PWA. This can be further tested and modified in future research.

Chapter 3 explored the relationships between structural damage and treatment-related improvement, as well as semantic and phonological ability. Correlations identified the Frontal Aslant Tract as one particularly influential tract for both semantic ability and overall naming, as well as phonological ability to some degree. Additionally, Tract-Based Spatial Statistics (TBSS) analyses identified differences in connectivity in a number of right hemisphere homologues between controls and semantically-improved PWA. This finding demonstrates that some areas involved in semantic processing do not require right hemisphere language homologue preservation and activation for improvement of semantic processing in particular to occur.

Chapters 4 and **5** describe the core empirical work reflected in the thesis. **Chapter 4** explains the methodology developed for use in this study. This separate chapter was required due to the complexity of methodology required to meet the needs of the study. We required a framework in which multiple disparate treatments and approaches to treatment could be compared under similar conditions, with a standardised methodology to avoid factors such as volume, intensity, amount and type of researcher involvement, amount of self-directed practice, etc. from biasing results due to methodological inconsistencies between approaches. In addition to this, a range of item difficulty (based on word length and frequency) and prompt progression hierarchies were necessary to ensure that all PWA in our sample, who was varied in terms of language ability, could be challenged and remain engaged in the treatment, while keeping the task easy enough to also facilitate progress. This was also necessary to avoid ceiling effects as much as possible. The methodology, therefore, included three different item-difficulty levels and three different prompt-progression hierarchies which could be selected to provide an appropriate and engaging challenge for the PWA participating in the study. Designing a methodology that was both standardised and long enough to allow direct comparison between multiple treatments, and included multiple measures of language ability, while also being flexible enough to account for the full range of PWA abilities, scheduling requirements, enthusiasm for practice, and technological resources and abilities regarding teletherapy, was complex. However, this methodology now provides a useful template for future studies aiming to compare multiple treatments in a standardised way.

After the methodology description in **Chapter 4**, **Chapter 5** details the results and a discussion of findings of the treatment study, which included a number of potentially exciting findings. Firstly, the direct comparison of trained item naming demonstrated a significantly increased level of improvement over the course of treatment for the two executive function treatments - Speeded and Interfered naming - when compared to the alternative modality treatments - Gesture and Prosody - and model-oriented treatments - Semantic and Phonological. This increase seemed fairly robust, remaining present regardless of which inter-participant factors were accounted for during analysis. There was also a significant increase to improvement in the executive function treatments for measures of semantic and phonological ability. This seems to indicate a clear advantage in using an executive function treatment to facilitate language improvement in PWA relative to the other approaches. Between the Speeded and Interfered naming treatments, however, there was some variation, with Speeded naming having greater variation in results, while Interfered naming results clustered closer to the mean. This possibly suggests that Speeded naming is more affected by inter-participant factors, and so is most beneficial for PWA whose language, lesion and demographic profiles indicate that they are likely to have a large potential for improvement. Conversely, Interfered naming may provide more consistent results for PWA whose language, lesion and demographic profiles suggest they may have somewhat less potential for improvement. This was supported by the negative interactions between the effects of the Interfered treatment and several inter-participant factors, suggesting that the improvement offered by Interfered naming is less affected by inter-participant variation than the other treatments.

Another possibly notable finding was regarding the generalisation of improvement to untrained items. While none of the approaches facilitated a significant improvement in the short term, in the long term the cumulative effects of all three approaches, totalling nine weeks of treatment, did facilitate a significant improvement. This finding suggests that treatment length may play a bigger role than treatment type in facilitation of generalisation of treatment gains. To what extent this effect is purely due to increased treatment volume, however, is unclear, and other confounding factors also exist which could account for this result. This finding is therefore the one which should arguably be taken with the most caution before it is confirmed or challenged by future research.

Furthermore, a number of inter-participant factors were determined to have an impact on treatment facilitation of improvement. However, the factors which appeared to have the most influence were PWA age, pre-treatment semantic ability and pre-treatment phonological ability. Analysis of lesion location also determined damage to regions in the posterior ITG and posterior cingulate gyrus to correlate with improvement in naming for the Prosody treatment. At a less strenuous threshold, damage to the anterior cingulate and paraginculate gyrus was also found to correlate to improvement in the Phonological treatment, and damage to regions in the Planum Polare and anterior STG correlated with improvement to Speeded treatment item naming.

Finally, the research was somewhat novel for the format and approach taken to data analysis. As the majority of analysis has been performed via R statistics (R Core Team, 2020), hopefully, when this study is published, it will be possible to provide electronic access to the core dataset and allow observers to view all the formatting changes and 'wrangling' of the dataset, as well as the specifics of each analysis performed. This will provide a template for doing so in a therapy study, allowing full transparency regarding data analysis, possibly prompting future studies to do the same, avoiding many of the issues associated with lack of transparency (Munafo et al., 2017) and contributing to the evolution of open science. The use of likelihood ratio tests and CLMMs, in addition to simpler statistical analyses, also has a couple of advantages. Use of both CLMMs and simpler analyses allows the same findings to be confirmed across two different analyses. CLMMs also have the advantage of taking into account each individual data point- the score for each PWA's naming attempt at each word, at each timepoint, as opposed to basing analyses on overall participant scores or averages. This means that, in terms of statistical power, we can get the 'most' out of the data we have collected. As neuroscientific and psychological research is often underpowered (Button et al., 2013; Nord et al., 2017), using the analyses which limit this issue as much as possible is beneficial.

Implications of research

Treatment Comparison

As previously stated, our literature review highlighted the fact that direct comparisons of effectiveness between different treatments and specifically, different approaches to treatment, are currently relatively sparse in the literature for aphasia treatment. Many studies have investigated the effects of studies individually, usually in comparison to a control (Bowen et al., 2012). While these studies are incredibly useful in determining the efficacy of treatments in a number of different measures - for example, semantic and phonological ability, functional communication, and trained item naming (Palmer et al., 2019), direct comparisons to evaluate the most effective treatment for a given group of PWA were previously rare. Our research offers a novel comparison of several different approaches to treatment, standardised across a single methodology which seemed to show a clear difference in effectiveness between the treatments focusing on executive function - Speeded and Interfered naming- and the other treatments. While these treatments have been the focus of a few studies (Conroy et al., 2018; Abel & Willmes, 2016; Bruehl et al., submitted), they are both relatively new to the field and as such have been researched less than some of the more established approaches to aphasia treatment - semantic and phonologically-based treatments in particular have had a large amount of research done investigating their effects (Whitworth et al., 2014). Therefore, the finding that, in general, they seem to outperform these more established treatments is exciting, but also requires further investigation, as there are a couple of different reasons why they might have excelled in our study relative to the other treatments.

When designing our study, we had to choose between controlling for treatment volume using amount of time per treatment, or number of exposures of each item. We chose time-based control of treatment volume. The primary reason for this was that we felt it is more ecologically valid and useful for SLTs, who are more likely to care more about how much benefit each treatment gives per hour of time investment by the SLT than per item repetition, which would be significantly less important to them when calculating how to facilitate the most improvement with their limited time with each PWA. However, this means that number of repetitions could not be controlled for- we could either stop each treatment session after a certain amount of time, or a certain number of repetitions of each item, but not both.

The alternative explanation is that the targeting of executive function in particular is the cause of these treatments' increased performance. Conroy et al (2018) posited two hypotheses for the performance of their Speeded naming treatment. The first is that precise representations are required for the language system to convert the required semantic information into motor-speech representations (Lupker et al., 1997). Computational models have shown that refining these representations through learning improves model performance and efficiency (Plaut et al., 1996). It is suggested that the additional pressure of speed as well as accuracy creates supports more refinement of these representations into more precise forms, allowing improved naming accuracy as well as speed (Conroy et al, 2018). While this hypothesis would explain the increased effect of Speeded naming, it does less to explain the increased effect of Interfered naming, unless we assume that challenging these representations via related interference words also allows their development into more precise forms. It does, however, appear that interfered naming targets language and interference control as well as lexical semantics (Bruehl et al., submitted).

The second hypothesis, however, focuses more on a cognitive-executive mechanism (Lambon-Ralph et al., 2010; Geranmayeh, Brownsett, & Wise, 2014) and so may offer a more complete explanation for the improved performance of the executive function approach as a whole. Engaging PWAs' executive and attentional skills to a greater degree (as was attempted in this approach) could allow increased learning and information retention. If PWA did find these treatments to be more engaging and motivating than others, this could have resulted in an increase in dopamine release (Fiorillo, 2013; Sharp et al., 2016), which is associated with improved learning and treatment effects (Berthier and Pulvermuller, 2011; Gill and Leff, 2012). Conroy et al. (2018) found their Speeded treatment to be particularly engaging and motivating- we could assume that the more challenging, 'gamified' elements of these executive function treatments do lead to increased motivation and engagement for PWA. Conroy et al. (2018) also found that degree of treatment maintenance related to patients' cognitive-executive skills, also suggesting a possible general effect of improved cognitive-executive ability in language performance and retaining treatment gains. This hypothesis more readily accounts for the improved performance of both Speeded and Interfered naming over the other treatments, and so may be more supported by our findings than the former hypothesis. A third possibility is that the success of these treatments is because they address multiple aspects of PWA cognition relating to language simultaneously; both executive functioning and word-finding. In addition to the executive functioning benefits provided by the treatments, word access in naming is also challenged- in interfered naming, it is impeded by the distracting word, while in speeded naming it is challenged by the increased speed requirement. This challenge improves word selection, including semantic control and word processing. These treatments may therefore have the benefits of improved executive control, combined with the benefits of challenging word access also improving word selection directly, leading to improvement in picture naming from two different 'sources'.

Generalisation

When considering measures of treatment success, the majority of studies in the area of treatment of aphasia focus primarily on improvement in naming of trained items (Fridriksson, 2010; Marcotte et al., 2012; Marcotte et al., 2013). In addition to this, other studies may include various additional measures of both speech and language and improvement to participant quality of life. These include semantic and phonological ability scores (Abel et al., 2014; 2015), functional communication measures (Schlaug et al., 2009), or qualitative ratings of changes to PWA communication ability or quality of life by PWA, carers or family members . Batteries such as the CAT (Swinburn, Porter, & Howard, 2005) or BDAE (Goodglass, Kaplan, & Weintraub, 2001) may also be used to gain an overall understanding of changes to a range of cognitive factors in addition to language changes. Again, however, direct comparison of generalisation of trained item gains to other language or non-language measures has been unusual in previous research This means that our research has produced some potentially exciting findings via comparison of this generalisation both between different treatments or approaches, and different time periods.

The most potentially exciting of our findings regarding generalisation is the somewhat incidental finding that a significant difference in untrained item naming was not found in the short term for any treatment, but was present over the whole length of the study. This provides some evidence that the length of treatment is a significant factor when considering extent of generalisation. It may, in fact, be more important to generalisation to untrained items than the specific approach to treatment used. This is because pairwise comparisons found no significant differences in improvement in untrained items between any of the three approaches to treatment. Generalisation of improvement to untrained items may therefore occur in the long term regardless of the approach used.

This finding highlights a key advantage of the methodology used in our therapy study. It was designed in a way that allowed collection of a range of information across multiple timepoints. This means that it provides information on the effects of treatment both over different lengths of time- for example, generalisation to untrained items being present in the long term but not the short term, as well as comparison of trained items both immediately post-treatment, and several weeks after study completion. This means that treatments have been compared both across multiple timepoints and using multiple measures of treatment efficacy, providing a detailed picture of each approach to treatment.

However, this highlights a potential limitation of the study with regard to comparison of short term versus long term generalisation. This finding was incidentally noticed due to the amount of data collected. However, the study was not designed around answering this question and as such, there are certainly confounding factors which could explain this finding. For example, the same untrained words were tested at each timepoint throughout the study. The 'long-term generalisation' we are seeing in these words could therefore instead be the result of a practice effect after PWA had multiple exposures to the untrained items over the course of the study. However, these exposures were also only once each 4-6 weeks. We would therefore expect that any improvement caused by these multiple exposures is likely not to be too substantial and may have added to the observed effect rather than explaining it completely. Further investigation into the effects of length of treatment on untrained item naming performance, and generalisation more generally, is therefore required before too much significance is placed on our findings regarding the effects of treatment length on generalisation.

Another potential issue with a methodology of this complexity, which included multiple assessment timepoints to be able to compare short and long-term effects of treatment, is that the length of the study could potentially have impacted participant attrition rate. While each treatment period was only three weeks long- consistent with previous prompt-based research, which generally employs a treatment period of 2-4 weeks (Abel et al., 2014; 2015), three treatment periods were needed in order to assess every treatment. When including pre, post, and interim assessments and breaks of between one and three weeks between each treatment period, the total length of the study could reach 22 weeks - potentially increasing the risk of participant attrition. Participant attrition can, of course, lead to issues with sample bias (Ahern & Le Brocque, 2005) as well as requiring more participants to be included and so increasing data collection time for the researcher. However, only two participants withdrew from the study, one of which was after a single session, so this problem was largely avoided, possibly negated by our attempts to ensure the methodology was varied, motivating, and flexible enough to be adapted to participant ability levels and maintain interest. While some participants did not receive a follow-up assessment, all but one of these participants' follow-up assessments were scheduled after the UK government lockdown due to the COVID-19 pandemic. This of course disrupted our ability to collect data for PWA not familiar with videoconferencing technology such as Skype. It therefore seems unlikely that the lack of follow-up data for these participants was due to any facets of our methodology.

Semantic and Phonological ability change

It is well established in the literature that PWAs' language and cognitive abilities pre-therapy can have an effect on their responses to treatment and scope for improvement of language (Lambon Ralph et al., 2010). More specifically, previous studies on treatments focusing on targeting semantics or phonology found differing effects depending on whether PWA had a primarily semantic or primarily phonological deficit (Abel et al., 2015). We aimed to build on these findings by identifying the primary white matter tracts linked to semantic and phonological ability both before treatment, and which tracts predicted successful improvement of semantic or phonological abilities.

A number of white matter tracts were linked to both semantic and phonological ability and ability improvement in the diffusion study. As discussed earlier, one particularly novel finding was the seeming role of the left FAT in semantic, and also to some extent, phonological processing. We also highlighted the importance of the right fornix, corticospinal tract, Inferior Longitudinal Fasciculus (ILF) and anterior arcuate in treatment-related improvement in semantic ability. Overall, semantic ability seemed to be affected by the integrity of more different tracts, but offered more clarity regarding which tracts preservation of integrity was important in in order to allow for treatment-related semantic improvement.

In addition to this, our analysis of structural MRI scans in the therapy study allowed identification of the paracingulate gyrus and anterior cingulate gyrus as specific regions linked to improvement facilitated by the phonological treatment. In terms of which treatments had the greatest effect on semantic and phonological ability, we found a significantly better performance from the executive function treatments in improving semantic, but not phonological, ability. We also found some evidence for generalisation to semantic ability for the alternate modality treatments. Although we found less direct influence of semantic and phonological treatments on language profile than in previous studies, these results did help expand our picture of the relationship between semantic and phonological ability and how these factors affect, and are affected by, treatments and treatment-related improvement. We identified more interactions between semantic and phonological ability and non-model-oriented treatments -a previously sparsely researched area- and added to our understanding of how these abilities relate to lesions in both the grey matter and white matter of the brain.

As previously discussed, research into aphasia therapy has involved a wide range of different methodologies. In addition to prompt-based picture naming methodologies focused on semantics and phonology (Abel et al., 2014; Abel et al., 2015; Fridriksson, 2010; Kiran et al., 2015; Leonard et al., 2015; Marcotte et al., 2012; Marcotte et al., 2013; Menke et al., 2009; van Hees et al., 2014), other treatment methodologies may focus more on recreation of whole sentences, as is the case in prosody-focused Melodic Intonation Therapy (MIT) studies (Schlaug et al., 2009; van de Sandt-Koenderman et al., 2016; Wan et al., 2014), or descriptions of situations (as in pantomime or gesture-based methodologies (Gili et al., 2017). Methodologies can also vary greatly in intensity and volume - for example, ILAT studies focus specifically on a high level of treatment intensity (Breier et al., 2006; McKinnon et al., 2017; Meinzer et al., 2008; Meinzer et al., 2009; Mohr et al., 2016; Nenert et al., 2017; Pulvermuller et al., 2005; Richter et al., 2008) - or be used in combination with brain stimulation methodologies such as rTMS or tDCS (Cherney et al., 2010; Hara et al., 2017; Hara et al., 2015; Marangolo et al., 2016). Studies also vary greatly in the amount of self-directed practice encouraged or required of participants (Antonucci, 2009; Stark & Warburton, 2018). Generally, outside the picture naming prompt-based studies often used in research into semantic and phonological treatments and research into individual, structured treatments such as ILAT or MIT, aphasia treatment research is largely varied in terms of the parameters and specific methodologies used. Of course, this variation has advantages - understanding how aphasia treatment can function in a range of conditions and exploring many disparate approaches to behavioural aphasia treatment. By comparison, we developed a methodology to allow direct comparison of approaches within the same conditions. This methodology did include flexibility- difficulty of the words and prompt progression used could be varied, and also allowed for some PWA self-directed treatment, as well as teletherapy.

This novel methodological structure is beneficial in that it provides a standardised ‘template’ for direct comparison of different treatments, which complements the variety of methods present in the current literature. The flexibility of the methodology allows adaptation of different treatment approaches to allow direct comparison within the same conditions: volume, intensity, treatment and prompt structure, level of therapist involvement. For example, treatments targeting the right hemisphere, outside of treatments focused on MIT, varied greatly in terms of how they were approached in research (Duncan and Small, 2018; Gili et al., 2017; Peck et al., 2004). However, both prosody and gesture fit well into our novel methodology, allowing comparison to traditionally prompt-based treatments such as semantic and phonologically-focused treatments under similar conditions.

This study also confirmed the effectiveness of teletherapy in aphasia research specifically. Generally, the majority of research has previously been performed in person by a researcher (Dial et al., 2019). There is evidence to suggest that teletherapy does little to hinder treatment effectiveness when compared to in-person treatment for participants with a variety of communicative impairments (Brennan, Georgeadis, Baron, & Barker, 2004) or for PWA more specifically Agostini et al., 2014; Dial et al., 2019). However, our research confirmed that performing a large proportion of treatment via teletherapy does not prevent a significant improvement in trained item naming in aphasia treatment. The treatments were set up on PWAs' home computers, allowing both practice at home and teletherapy via video-calling software or telephone.

The built-in flexibility of the methodology in our research also allowed PWA to practice at an appropriate level for them- one that was both challenging and engaging, but not so hard as to be discouraging. Prompt-based treatments generally use one of three forms of prompt progression. Decreasing prompt hierarchies begin with the most 'helpful' prompt, providing less informative prompts as the hierarchy progresses (Fillingham et al., 2003). Increasing hierarchies do the reverse, starting with the least helpful, informative prompt and progressing to the most helpful if the PWA continues to struggle with naming (Conroy et al., 2009). Finally, some treatments, such as SFA or PCA, include participant-generated prompts (Coelho et al., 2000; Leonard et al., 2015). These prompt progressions offer a range of challenge levels for naming, but are generally used individually, in separate studies. Similarly, the majority of prompt-based research also uses a single set of trained items. Allowing flexibility between PWA in both difficulty of item set (as measured by average word length and frequency), and prompt progression, hopefully provided all participants with an appropriate challenge. This was particularly important considering the range of language ability included in our participant sample, and hopefully increased PWA enjoyment and engagement with the treatment, possibly increasing treatment effectiveness (Conroy et al., 2018) and participant retention over the course of the study.

Another implication of demonstrating the effectiveness of a methodology which could largely be self-directed by PWA if necessary are the future possibilities for development of the treatment into self-directed software to be used by PWA mostly, or in some cases completely, independently, or as a supplement to in-person treatment. A number of applications have been used in research which offer forms of self-directed treatment for PWA (Stark and Warburton, 2018). Development of our methodology into something similar would provide PWA with an evidence-supported treatment option, which could include a range of approaches and prompt structure and item difficulty levels to suit PWA preference. This option is aided by the fact that the treatment is already coded through Visual Basic in Microsoft Powerpoint, providing a basis for the code that would be necessary if the methodology was to be developed into a complete application or other software.

Optimisation of treatment based on inter-participant variables

There are many PWA-related factors which we know have an effect on treatment effectiveness. Many previous studies have highlighted the effect that non-stroke related factors such as PWA age and level of education have on PWA reaction to treatment (Plowman et al., 2012; Tippett, 2015). Stroke-related factors can also have a large effect - time since stroke can make a large difference to likelihood and extent of PWA improvement in language ability, with or without treatment. During the chronic phase, negligible spontaneous recovery is expected (Berthier & Pulvermuller, 2011). However, treatment-related improvement is still possible, and this factor alone means that amount of time post-stroke may affect degree of treatment effect in PWA, especially in a sample like ours, which includes many PWA who have volunteered for a number of assessments and treatments via participation in research studies. Such participants may be closer to their improvement 'ceiling' depending on time post-stroke in comparison to a 'newer', less long-term chronic and less highly 'treated' general population of PWA. Pre-study language profile can also have an effect - (Abel et al., 2014) found differences in the effects of semantic and phonologically-focused treatments depending on initial PWA semantic and phonological ability scores. Finally, of course, PWA lesion size and location can also greatly affect the amount of improvement possible via treatment (Perani et al., 2003; Vitali et al., 2007). However, the exact effects these factors may have and how to optimise treatment around them are still being explored - for example, the involvement of right-hemisphere homologues to left-hemisphere language regions in recovery are still under debate. For example, the effect of right hemisphere activation may be considered to be time-dependent - an aspect of recovery before activation is returned to left hemisphere regions (Tippett et al., 2014). Other hypotheses suggest it is more lesion-dependent, and only effective when required to compensate for severely damaged left hemisphere language regions (Crosson et al., 2007; Heiss & Thiel, 2006). Our research has provided additional information regarding how the effects of these inter-participant variables may differ between different participants and treatments.

One of the benefits of our research has been the amount of data collected on each participant. This has allowed us to explore the links between naming improvement, demographic factors, language profile, cognitive and lesion-related factors, control item naming, and several different levels of functional communication. It is unusual to be able to examine the inputs and effects of, and on, so many different factors. However, our participant sample has, to some extent, limited some of the nuance we had hoped to gain in our insights into the effects of different inter-participant factors. In some respects, the sample lacks diversity. The use of a participant database means that most of the PWA were older and at least several years post-stroke. The sample was also fairly limited in terms of time spent in education- the majority of PWA left education at 16. The sample therefore has limited diversity in several regards. This made it harder to fully evaluate inter-participant differences and tease apart how these factors might affect language measure improvement between different treatments.

Our sample was also more limited than initially hoped in terms of neuroimaging scans. Ideally, both pre- and post-treatment structural MRI scans, as well as DWI scans, would have been obtained in order to gain a greater insight into how PWAs' brain structure and connectivity were affected by treatment, in a similar way to the insight we gained into changes in participants' language and cognitive profiles as a result of treatment via behavioural tests. Unfortunately, these post-treatment scans could not be obtained, limiting the insights we could gain into the relationships between lesions, treatment, language profiles and improvement in trained items to those pertaining to pre-study lesions.

While our sample in some ways limited the robustness of the conclusions we were able to draw regarding inter-participant factors, the use of likelihood ratio tests with CLMMs and LMMs allowed each item to be accounted for in analyses, rather than analysis being performed using participant averages. This meant we were able to get the most out of a dataset which was both rich in terms of variety of language measures, but also lacked diversity in some respects when considering inter-participant factors. This analysis allowed us to identify Age and Dell s and p (Semantic and Phonological) scores as the most impactful inter-participant factors in the study, although again this may be partially due to the lack of sample diversity regarding certain demographic factors meaning they produced less impact in this case than they would in a more varied sample. It also allowed us to compare the performance of each treatment when taking different inter-participant factors into account- interestingly, while participant performance did vary depending on inter-participant factors, the executive function treatments (Speeded and Interfered naming) appeared to be the most effective in all cases. While this finding may also require further exploration, this may suggest that, overall, use of an executive function treatment may be most beneficial at a group level in most or all cases, possibly negating the need for a large degree of treatment optimisation depending on variation between participants.

The analyses in our therapy study, excluding the analysis of lesion location, were all performed via the R data analysis software (R Core Team, 2020). In addition to allowing us to perform CLMMs and LMMs comparatively simply, getting the most out of our dataset and allowing us to explore one of the nuances in the data, using R provides benefits in terms of data transparency. Lack of data reproducibility has been found to be a prevalent issue across a number of fields (Altman, 1994; Ioannidis, 2005), throwing many previously established findings into doubt. This issue is considered by many to be a long-term issue (Nosek et al., 2015), and can be considered to be partially driven by lack of transparency in terms of study data and data analysis (Nord et al., 2017). Our use of R allows us to upload the code used in our data analysis, meaning that the research community will have the ability to run our data analysis themselves using our code and anonymised dataset, providing full transparency in the steps and specific conditions used in our data analysis. It is hoped that this data transparency and ability for researchers to directly inspect or analyses will increase the chances of successful reproduction of the findings of the study and avoid contributing to the broader scientific issue of lack of data reproducibility affecting many findings across many fields.

Directions for future research

The research conducted in this thesis has explored a previously sparsely researched area of the aphasia treatment literature. Direct comparison of different therapeutic approaches via a standardised methodology is novel to this area, and so, while our research has provided a great amount of exploratory information in comparing different approaches and treatments, it was unable to cover every outstanding question that warrants investigation. A lot of the answers it provides also beg their own follow-up questions, meaning there are possible directions for future research in this area.

Arguably the most notable finding from our research concerned the overall greater success of the executive function treatments - Speeded and Interfered naming - in comparison to the other treatments, both in terms of trained item naming and some aspects of generalisation. In particular, we found some evidence of differences regarding variation in participant improvement and the inter-participant factors affecting these two treatments. For example, the effectiveness of Speeded naming was more directly linked to damage in the Planum Polare and Superior Temporal Gyrus. Given the overall greater success of these two treatments, one fruitful avenue of future research might be to further explore the specific differences between and advantages of each treatment. Our results seem largely to suggest that one of these two treatments is likely to be the most effective in the majority of PWA cases (at least in those reflecting the demographics present in our participant group), and that there are some inter-participant differences in effectiveness between the two treatments. Therefore, greater exploration of the specific inter-participant factors affecting their effectiveness, as well as differences in effectiveness in different measures (for example, different types of generalisation) would help to unpick which treatment might be optimal in any given PWA, with specific demographic information, language deficits, lesions, and goals regarding treatment.

Another interesting direction for future research regarding these two more novel treatments would be to investigate the effects of combining them into a more general ‘executive function’ treatment. As previously stated, while both treatments seem to have slightly differing effects depending on certain inter-participant factors, both performed significantly better than the other treatments in terms of trained item naming improvement. It may therefore be the case that the approach these treatments take - targeting executive function - is the primary cause of their improved performance over treatments using other approaches. Combining the two treatments in some way and thus allowing targeting of executive function in multiple ways could potentially, therefore, be hypothesised to be as effective or more than either individual treatment, possibly combining the advantages of both. A treatment combining both would also likely be more engaging for PWA, with the increased variation in reducing boredom and fatigue in the longer term and possibly allowing for increased volume or intensity without affecting participant attrition rate.

Given the effectiveness of the executive function approach in comparison to the other approaches in terms of trained item naming improvement, exploration of this approach and the different methodologies that could be used to most effectively target executive function in the context of speech should also be a priority for future research. This approach both appears to be the most fruitful area for improvement of naming of trained items, and is still an emerging area- in addition to Speeded and Interfered naming, Cognitive Flexibility in Aphasia Therapy (CFAT) (Spitzer et al., 2020; Spitzer et al., 2021) also focuses on executive function treatment. However, we are unaware of any other existing treatments in which targeting of executive functioning forms a key aspect of treatment effectiveness. Therefore, in addition to further research into Speeded and Interfered naming and how they might most effectively be used, either separately or in tandem, the development of other treatments based theoretically around targeting executive function would also be an interesting area of future study. Given that both the current treatments using this approach have proved very successful, the executive function approach looks promising as a whole, and it seems likely that new treatments using it as an approach could also prove very effective, and might end up offering a more optimal treatment option for some PWA - as we can already see the inter-participant variation in the success of Speeded naming in particular, having more options for PWA with characteristics indicating they may not benefit as much from Speeded naming would only benefit our ability to optimise treatment and provide each PWA the best, most effective and engaging treatment option for them. Given the possibility that the executive function treatments may also be the most effective due to their 'game-like' aspects encouraging participant use and enthusiasm for the treatment and aiding with dopamine levels, and thus, learning, another interesting direction for future research would be the formation of even more 'game-like' treatments. If keeping PWA interest could be the key factor in improvement, it makes sense for future research to consider designing treatments with greater 'gamification' elements (for example: more virtual 'rewards' for correct answers, more aspects of challenge and competition, and more 'story-like' elements) to see if this further improves PWA response to treatment.

Another key area which our research has generated questions as well as providing answers regards the capacity for treatments to provide generalisation to non-trained language abilities, and specifically how generalisation is affected by different treatment lengths. Previous research has mostly focused on treatment type, volume and intensity when considering likelihood of generalisation of any improvements made to non-trained measures of language and communication (Bhogal et al., 2003; van der Meulen et al., 2016). However, our research also seemed to point to length of treatment as a potentially important factor in determining amounts of this type of generalisation - in our study, the cumulative effects of treatment over the full study period (11-15 weeks) seemed to produce a significant improvement in untrained word naming despite this significant improvement not being present during any of the individual three-week treatment periods. However, as discussed earlier, this finding was somewhat incidental and, as such, should be viewed with an appropriate amount of scepticism until future research can offer supporting or conflicting evidence, due to the confounding factors which may have affected this finding specifically. In spite of this, the possibility of cumulative treatment length having a larger than previously thought effect on generalisation of language abilities is an exciting one. Therefore, one or multiple studies fully focused on exploring the effects of treatment length on generalisation would be another interesting area of future study.

Initially, future research into the amount of effect treatment length has on any generalisation of improvement would probably benefit from focusing primarily on investigating the veracity of our current findings - identifying whether our findings were supported or refuted, and, if supported, the consistency of these findings. Does longer-term (9+ weeks) of treatment consistently lead to improved generalisation relative to shorter-term treatment? Is this effect purely due to increased treatment volume or does it remain when total volume is controlled for? As many PWA in the UK typically have access to a very limited volume of treatment (Palmer et al., 2018), spreading this volume over the most effective time period would certainly be worth investigation. Generalisation to functional communication and improved quality of life is, for many, the ultimate aim of language treatment for PWA, so gaining an improved understanding of the conditions which can most effectively facilitate this could be considered to be as important as comparison of different treatments.

If supporting evidence for a consistent effect of treatment length on amount of generalisation to non-trained aspects of language is found, the effects of different treatment periods should also be investigated further. Does generalisation vary over different treatment lengths? If so, is this variation proportional to treatment length, or is there an optimal treatment length which provides the maximum generalisation effect relative to investment in terms of time for treatment? What are the limits of this effect? Is there a length of treatment at which it becomes inefficient to continue treatment in terms of generalisation gains, and what is this length? Clearly, there are a number of questions to consider which would be worth investigation in future research. It may be that our findings are found to have little support, but if length of treatment is more important to generalisation of improvement of language to non-trained items than previously thought, it is an important factor to fully investigate in order to provide optimal recovery for PWA based both on the resources available to them and maximising functional communication and thus quality of life improvement.

As previously discussed, a novel methodology for comparison of different treatments and approaches was used in this research. This methodology has shown itself to be flexible and adaptable to a range of disparate treatment approaches. It therefore has potential to be used in the future as a standardised ‘template’ in treatment comparison. It is adaptable enough to be adjusted to suit comparisons involving treatments not included in our research; for example, Intention treatment (Peck et al., 2004) could quite easily fit into this template by adding the hand movements and actions which define it to the picture naming and prompt structure used in our study for Speeded and Interfered naming. This would allow future research to perform different comparisons of treatments in a standardised way to address their specific questions. As mentioned previously, there has been comparatively little direct comparison of different aphasia treatments thus far in the literature, and it is hoped that this methodology will provide a method of comparison which allows researchers to account for potentially confounding factors such as varying treatment structure, level of SLT involvement, amount of self-directed learning, and overall volume, as measured either by number of item presentations or time spent on each treatment. This template could, in addition to making more novel comparisons of different treatment approaches, also be used in comparisons of a single treatment across different contexts- for example, comparisons of different intensity levels, amount of self-directed learning, treatment volume, or, as discussed earlier, exploring the effects of varying treatment timescales. While these comparisons might be possible to perform within the ‘native’ methodology of any single treatment, this methodology template is flexible enough to make these comparisons easily for many different treatments, and the use of a standardised methodology would allow for more direct comparisons to be made between different treatments and studies. As also discussed earlier, the treatments used in this study, and this methodology more generally, could also be possible candidates for software development. Many PWA currently have limited access to in-person treatment with an SLT, and certainly lack the volume of therapy in this form necessary to make optimal progress (Palmer et al., 2018). The development of treatment software which PWA can use as independently as possible could therefore fill a significant need, both providing treatment for PWA without access to in-person SLTs, and allowing those PWA who do have that access to increase the overall volume of treatment they are receiving, allowing them a better chance at getting closer to optimal improvement. Software development would also allow for integration of the ‘gamified’ aspects of treatment discussed earlier to be relatively seamlessly included in treatment, as the software could be more clearly designed in a ‘game’ style. This encouraging of PWA enthusiasm for treatment and increased dopamine levels could help with both PWA engagement in treatment (and thus higher volume of treatment), and retention of information.

While it is clear that Speeded and Interfered naming outperformed the other treatments in terms of improvement in trained item naming, our findings also supported previous research regarding the effect of PWA demographic, language and lesion profile factors. We found that multiple different inter-participant factors have a large effect on treatment outcome. Our findings seemed to indicate that the executive function treatments remained the most effective treatments regardless of which inter-participant factors were accounted for. However, it would definitely be beneficial to spend more time exploring whether this is always the case. Firstly, the specific effects and extent of effect of each main inter-participant factor have on participants must be explored in greater detail, so that it can be determined at what point (if any) switching to a different treatment may predict the best improvement for a given participant. Our research has identified the factors which seemed to hold the most influence in our study, however it has not reached the point of fully identifying the point different inter-participant factors would have to reach in order for different treatments to predict improved outcomes for any given participant- this is a complex task involving many different factors, and unfortunately the conclusions we could reach in this area were, to some degree, limited by the size of our participant sample.

Another interesting area to explore would be whether or not optimising treatment in this way, where treatment is dependent on individual PWA characteristics, provides a better outcome at a group level than always using executive function treatments. As mentioned earlier, our findings suggested that, at the group level the executive function treatments were the most effective regardless of the inter-participant factors taken into account. However, there were of course cases of individuals performing better in non-executive function treatments. It would therefore be beneficial for future researchers to compare the improvement of a group of PWA being treated using executive function treatments against that of a group of PWA being treated using their 'optimal' treatments (based on inter-participant factors) to see if the 'optimal' group significantly outperforms the purely executive function treatment group at a group level, and if so, the extent of this difference in improvement levels. This would allow an evaluation of the benefits of optimising treatments versus simply using what appear to be the most effective overall treatments. If optimisation does provide an increased benefit, a comparison of the degree of this benefit with the extra resource optimisation would be required to determine which is the most efficient method of providing effective treatment to PWA overall.

Conclusions

There are a number of different approaches to behavioural treatment of aphasia currently in use by SLTs. Thus far, research has explored many of these approaches in greater detail, often comparing them either to a generic control treatment, the same treatment without the theorised active element, or no treatment. While this has granted us a good deal of insight into these treatments, how they facilitate improvement, both cognitively and physically, and the factors involved in their success, our literature review identified some areas of the field that would benefit from greater exploration. Firstly, while there were a number of MRI and fMRI-based neuroimaging studies, there were relatively few diffusion-based neuroimaging studies investigating the relationship between language recovery and white matter tracts. We therefore investigated DTI scans performed before treatment in a number of PWA, identifying the tracts whose preservation correlated with treatment-facilitated improvement in semantic or phonological ability. Secondly, treatments were rarely directly compared with each other - while we knew many treatments were effective in comparison to controls, we had little evidence on what the most effective treatment to use for any given PWA would be. Therefore, a treatment comparison study was performed to allow direct comparison of six treatments, spanning three main approaches, across a range of measures of treatment effectiveness. Several key findings emerged from this study. Firstly, the study methodology was found to be effective in facilitating PWA language improvement, and successfully allowed a direct comparison of several disparate treatments, providing a methodological template for standardised comparison of treatment for future studies. Secondly, generalisation of trained item improvement to multiple different measures of language and cognitive ability was examined and, while generalisation was observed in several different areas, when considering untrained item naming, length of treatment was found to be potentially more influential than treatment type in determining amount of improvement. While this finding requires confirmation and exploration, it opens a potentially exciting line of research regarding how to formulate treatment to gain the most practical benefit in terms of functional communication, and whether generalisation to non-trained language measures can be enhanced by distributing treatment over the correct time frame. We also found a number of inter-participant factors which had an effect on improvement which differed between treatment. However, the most notable finding overall was the success of the more novel executive function treatments - Speeded and Interfered naming - in comparison to the other approaches. We have identified two main hypotheses as to why these treatments were particularly effective. These two treatments both outperformed all the other treatments on our primary measure of improvement- trained item naming- as well as measures of semantic and phonological ability. While this success may be somewhat modulated by inter-participant factors, these treatments overall seem to significantly outperform the others, and require more research as practical treatments for aphasia so they may be used more widely in everyday treatment for PWA.

References

Abel, S., & Willmes, K. (2016). Combined linguistic-executive therapy of word processing in aphasia. *Front. Psychol. Conference Abstract: 54th Annual Academy of Aphasia Meeting*. doi:10.3389/conf.fpsyg.2016.68.00029

Bruehl, S., Willmes, K., & Binkofski, F. (2021). Interfered-Naming Therapy for Aphasia (INTA): a neuroscience-based approach to improve linguistic-executive processing. *Aphasiology*, 1-22.

Bruehl, S., Willmes, K., & Binkofski, F. (2021). Interfered-Naming Therapy for Aphasia (INTA): behavioural and computational effects of a novel linguistic-executive approach. *Aphasiology*, 1-22.

Abel, S., Dressel, K., Bitzer, R., Kümmerer, D., Mader, I., Weiller, C., & Huber, W. (2009). The separation of processing stages in a lexical interference fMRI-paradigm. *NeuroImage, 44*(3), 1113-1124.

Abel, S., Schultz, A., Radermacher, I., Willmes, K., & Huber, W. (2005). Decreasing and increasing cues in naming therapy for aphasia. *Aphasiology, 19*(9), 831-848.

Abel, S., Weiller, C., Huber, W., & Willmes, K. (2014). Neural underpinnings for model-oriented therapy of aphasic word production. *Neuropsychologia*, *57,* 154-165.

Abel, S., Weiller, C., Huber, W., Willmes, K., & Specht, K. (2015). Therapy-induced brain reorganization patterns in aphasia. *Brain*, *138*, 1097-1112.

Agostini, M., Garzon, M., Benavides-Varela, S., De Pellegrin, S., Bencini, G., Rossi, G., . . . Meneghello, F. (2014). Telerehabilitation in poststroke anomia. *BioMed research international, 2014*.

Ahern, K., & Le Brocque, R. (2005). Methodological issues in the effects of attrition: Simple solutions for social scientists. *Field methods, 17*(1), 53-69.

Akhavan, N., Göksun, T., & Nozari, N. (2017). Integrity and function of gestures in aphasia. *Aphasiology,* 1-26.

Al-Janabi, S., Nickels, L. A., Sowman, P. F., Burianová, H., Merrett, D., & Thompson, B. (2014). Augmenting melodic intonation therapy with non-invasive brain stimulation to treat impaired left-hemisphere function: two case studies. *Frontiers in Psychology, 5*, 37.

Altman, D. G. (1994). The scandal of poor medical research.

Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage, 125*, 1063-1078.

Anglade, C., Thiel, A., & Ansaldo, A. I. (2014). The complementary role of the cerebral hemispheres in recovery from aphasia after stroke: A critical review of the literature. *Brain Injury, 28*(2), 138-145.

Antonucci, S. M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology, 23*(7-8), 854-866.

Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology, 24*(3), 363-394.

Aung, W. Y., Mar, S., & Benzinger, T. L. S. (2013). Diffusion tensor MRI as a biomarker in axonal and myelin damage. *Imaging in medicine, 5*(5), 427.

Basilakos, A., Fillmore, P. T., Rorden, C., Guo, D., Bonilha, L., & Fridriksson, J. (2014). Regional white matter damage predicts speech fluency in chronic post-stroke aphasia. *Frontiers in Human Neuroscience, 8*, 845.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature neuroscience, 6*(5), 448.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage, 34*(1), 144-155.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., . . . Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 50*(5), 1077-1088.

Benjamin, M. L., Towler, S., Garcia, A., Park, H., Sudhyadhom, A., Harnish, S., . . . Crosson, B. (2014). A Behavioural Manipulation Engages Right Frontal Cortex During Aphasia Therapy. *Neurorehabilition and Neural Repair, 28*(6), 545-553.

Berthier, M. L., & Pulvermuller, F. (2011). Neuroscience insights improve neurorehabilitation of poststroke aphasia. *Nat. Rev. Neurol., 7*, 86-97.

Berthier, M. L., De-Torres, I., Paredes-Pacheco, J., Roé-Vellvé, N., Thurnhofer-Hemsi, K., Torres-Prioris, M. J., . . . Dávila, G. (2017). Cholinergic Potentiation and Audiovisual Repetition-Imitation Therapy Improve Speech Production and Communication Deficits in a Person with Crossed Aphasia by Inducing Structural Plasticity in White Matter Tracts. *Frontiers in Human Neuroscience, 11*, 304.

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex, 49*(9), 2345-2357.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke, 34*(4), 987-993.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767-2796.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *The Journal of Neuroscience, 17*(1), 353-362.

Blasi, V., Young, A. C., Tansy, A. P., Petersen, S. E., Snyder, A. Z., & Corbetta, M. (2002). Word retrieval learning modulates right frontal cortex in patients with left frontal damage. *Neuron, 36*(1), 159-170.

Blomert, L., Kean, M.L., Koster, Ch. & Schokker, J. (1994) Amsterdam—Nijmegen everyday language test: construction, reliability and validity*, Aphasiology, 8:4*, 381-407, DOI: 10.1080/02687039408248666

Bonilha, L., Gleichgerrcht, E., Nesland, T., Rorden, C., & Fridriksson, J. (2016). Success of anomia treatment in aphasia is associated with preserved architecture of global and left temporal lobe structural networks. *Neurorehabilitation and Neural Repair, 30*(3), 266-279.

Booth, J. R., Lu, D., Burman, D. D., Chou, T.-L., Jin, Z., Peng, D.-L., . . . Liu, L. (2006). Specialization of phonological and semantic processing in Chinese word reading. *Brain research, 1071*(1), 197-207.

Bowen, A., Hesketh, A., Patchick, E., Young, A., Davies, L., Vail, A., ... & Tyrrell, P. (2012). Clinical effectiveness, cost-effectiveness and service users' perceptions of early, well-resourced communication therapy following a stroke: a randomised controlled trial (the ACT NoW Study). *Health technology assessment (Winchester, England), 16*(26), 1-160.

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*(4), 94-98.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38*(9), 1207-1215.

Bozzali, M., Parker, G. J. M., Serra, L., Embleton, K., Gili, T., Perri, R., . . . Cercignani, M. (2011). Anatomical connectivity mapping: a new tool to assess brain disconnection in Alzheimer's disease. *NeuroImage, 54*(3), 2045-2051.

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *The Cochrane Library.*

Breier, J. I., Maher, L. M., Novak, B., & Papanicolaou, A. C. (2006). Functional Imaging Before and After Constraint-Induced Language Therapy for Aphasia Using Magnetoencephalography. *Neurocase, 12*(6), 322-331.

Breier, J. I., Randle, S., Maher, L. M., & Papanicolaou, A. C. (2010). Changes in maps of language activity activation following melodic intonation therapy using magnetoencephalography: Two case studies. *Journal of Clinical and Experimental Neuropsychology, 32*(3), 309-314.

Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., . . . Baumgaertner, A. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet, 389*(10078), 1528-1538.

Brennan, D. M., Georgeadis, A. C., Baron, C. R., & Barker, L. M. (2004). The effect of videoconference-based telerehabilitation on story retelling performance by brain-injured subjects and its implications for remote speech-language therapy. *Telemedicine Journal & e-Health, 10*(2), 147-154.

Budd, M. A., Kortte, K., Cloutman, L., Newhart, M., Gottesman, R. F., Davis, C., . . . Hillis, A. E. (2010). The Nature of Naming Errors in Primary Progressive Aphasia Versus Acute Post-Stroke Aphasia. *Neuropsychology, 24*(5), 581-589.

Burton, M. W., Small, S. L., & Blumstein, S. E. (2000). The role of segmentation in phonological processing: an fMRI investigation. *Journal of Cognitive Neuroscience, 12(*4), 679-690.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience, 14*(5), 365-376.

Calvert, G. A., Brammer, M. J., Morris, R. G., Williams, S. C. R., King, N., & Matthews, P. M. (2000). Using fMRI to study recovery from acquired dysphasia. *Brain and Language, 71*(3), 391-399.

Campana, S., Caltagirone, C., & Marangolo, P. (2015). Combining voxel-based lesion-symptom mapping (VLSM) with A-tDCS language treatment: predicting outcome of recovery in nonfluent chronic aphasia. *Brain stimulation, 8*(4), 769-776.

Cao, Y., Vikingstad, E. M., George, K. P., Johnson, A. F., & Welch, K. M. A. (1999). Cortical language activation in stroke patients recovering from aphasia with functional MRI. *Stroke, 30*(11), 2331-2340.

Cappa, S. F. (1998). Spontaneous recovery from aphasia. *Handbook of neurolinguistics* (pp. 535-545). Academic Press.

Carragher, M., Conroy, P., Sage, K., & Wilkinson, R. (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions. *Aphasiology, 26*(7), 895-916.

Catani, M., & de Schotten, M. T. (2012). *Atlas of human brain connections:* Oxford University Press.

Catani, M., & de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex, 44*(8), 1105-1132.

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., . . . Weintraub, S. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain, 136*(8), 2619-2628.

Chao, L. L., & Knight, R. T. (1998). Contribution of human prefrontal cortex to delay performance. *Journal of Cognitive Neuroscience, 10*(2), 167-177.

Chen, W., Ye, Q., Ji, X., Zhang, S., Yang, X., Zhou, Q., . . . Shan, C. (2015). Mirror neuron system based therapy for aphasia rehabilitation*. Front. Psychol., 6*(1665).

Cherney, L. R., Erickson, R. K., & Small, S. L. (2010). Epidural cortical stimulation as adjunctive treatment for non-fluent aphasia: preliminary findings. *J Neurol Neurosurg Psychiatry, 81*, 1014-1021.

Cherney, L. R., Patterson, J. P., Raymer, A. M., Frymark, T., & Schooling, T. (2008). Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language, and Hearing Research.*

Choi, Y.-H., Park, H. K., & Paik, N.-J. (2015). Role of the posterior temporal lobe during language tasks: a virtual lesion study using repetitive transcranial magnetic stimulation*. NeuroReport, 26*(6), 314-319.

Cipollari, S., Veniero, D., Razzano, C., Caltagirone, C., Koch, G., & Marangolo, P. (2015). Combining TMS-EEG with transcranial direct current stimulation language treatment in aphasia. *Expert review of neurotherapeutics, 15*(7), 833-845.

Cocquyt, E.-M., De Ley, L., Santens, P., van Borsel, J. V., & De Letter, M. (2017). The role of the right hemisphere in the recovery of stroke-related aphasia: A systematic review. *Journal of Neurolinguistics, 44*, 68-90.

Coelho, C. A., McHugh, R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology, 14*(2), 133-142.

Conroy, P. & Scowcroft, J. (2012). Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study. *Neuropsychological rehabilitation. 22*. 295-318. 10.1080/09602011.2011.641434.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009). The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. *Aphasiology, 23*(6), 707-730.

Conroy, P., Sotiropoulou Drosopoulou, C., Humphreys, G. F., Halai, A. D., & Lambon Ralph, M. A. (2018). Time for a quick word? The striking benefits of training speed and accuracy of word retrieval in post-stroke aphasia. *Brain, 141*(6), 1815-1827.

Crinion & Leff (2015). Using functional imaging to understand therapeutic effects in poststroke aphasia. *Curr Opin Neurol, 28*(4), 330-337.

Crinion, J. T., & Leff, A. P. (2007). Recovery and treatment of aphasia after stroke: functional imaging studies. *Current Opinion in Neurology, 20*, 667-673.

Crinion, J. T., & Leff, A. P. (2015). Using functional imaging to understand therapeutic effects in poststroke aphasia. *Current Opinion in Neurology 28(*4), 330-337.

Crosson, B., McGregor, K., Gopinath, K. S., Conway, T. W., Benjamin, M., Chang, Y.-L., . . . White, K. D. (2007). Functional MRI of Language in Aphasia: A Review of the Literature and the Methodological Challenges. *Neuropsychol Rev, 17,* 157-177.

Crosson, B., Moore, A. B., Gopinath, K., White, K., Wierenga, C. E., Gaiefsky, M. E., . . . Gonzalez Rothi, L. J. (2005). Role of the Right and Left Hemispheres in Recovery of Function during Treatment of Intention in Aphasia. *Journal of Cognitive Neuroscience, 17*(3), 392-406.

Crosson, B., Moore, A. B., McGregor, K. M., Chang, Y.-L., Benjamin, M., Gopinath, K., . . . White, K. D. (2009). Regional Changes in Word-Production Laterality After a Naming Treatment Designed to Produce a Rightward Shift in Frontal Activity. *Brain Lang, 111*(2), 73-85.

Degen, R. M., Fink, M. E., Callahan, L., Fibel, K. H., Ramsay, J., & Kelly, B. T. (2016). Brain contusion with aphasia following an ice hockey injury. *The Physician and Sportsmedicine, 44*(3), 324-326.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287-314.

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *Journal of Neuroscience, 15*(9), 5870-5878.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage, 31*(3), 968-980.

Dial, H. R., Hinshelwood, H. A., Grasso, S. M., Hubbard, H. I., Gorno-Tempini, M. L., & Henry, M. L. (2019). Investigating the utility of teletherapy in individuals with primary progressive aphasia. *Clinical Interventions in Aging, 14,* 453.

Dignam, J. K., Copland, D. A., Rawlings, A., O'Brien, K., Burfein, P., & Rodriguez, A. D. (2016). The relationship between novel word learning and anomia treatment success in adults with chronic aphasia. *Neuropsychologia, 81*, 186-197.

Dignam, J. K., Rodriguez, A. D., & Copland, D. A. (2016). Evidence for intensive aphasia therapy: consideration of theories from neuroscience and cognitive psychology. *PM&R, 8*(3), 254-267.

Doogan, C., Dignam, J., Copland, D., & Leff, A. (2018). Aphasia recovery: when, how and who to treat?. *Current neurology and neuroscience reports, 18*(12), 1-7.

Duffau, H., Gatignol, P., Moritz-Gasser, S., & Mandonnet, E. (2009). Is the left uncinate fasciculus essential for language? . *J Neurol, 256*, 382-389.

Duncan, E. S., & Small, S. L. (2018). Changes in dynamic resting state network connectivity following aphasia therapy. *Brain imaging and behavior, 12*(4), 1141-1149.

Efstratiadou, E. A. (2018). Investigation of different therapy approaches for aphasia in the Greek language (Doctoral dissertation, City, University of London).

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation, 13*(3), 337-363.

Fillingham, J., Sage, K., & Lambon Ralph, M. (2005). Further explorations and an overview of errorless and errorful therapy for aphasic word-finding difficulties: The number of naming attempts during therapy affects outcome. *Aphasiology, 19*(7), 597-614.

Fiorillo, C. D. (2013). Two dimensions of value: dopamine neurons represent reward but not aversiveness. *Science, 341*(6145), 546-549.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research, 12*(3), 189-198.

Fridriksson, J. (2010). Preservation and modulation of specific left hemisphere regions is vital for treated recovery from anomia in stroke. *Journal of Neuroscience,* 30(35), 11558-11564.

Fridriksson, J., Baker, J. M., Whiteside, J., Eoute, D., Moser, D., Vesselinov, R., & Rorden, C. (2009). Treating visual speech perception to improve speech production in nonfluent aphasia. *Stroke, 40*(3), 853-858.

Fridriksson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A. L., Bonilha, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca’s aphasia to produce fluent speech*. Brain, 135*(12), 3815-3829.

Fridriksson, J., Moser, D., Bonilha, L., Morrow-Odom, K. L., Shaw, H., Fridriksson, A., . . . Rorden, C. (2007). Neural correlates of phonological and semantic-based anomia treatment in aphasia. *Neuropsychologia, 45*(8), 1812-1822.

Fridriksson, J., Moss, J., Davis, B., Baylis, G. C., Bonilha, L., & Rorden, C. (2008). Motor speech perception modulates the cortical language areas. *NeuroImage, 41*(2), 605-613.

Fridriksson, J., Richardson, J. D., Baker, J. M., & Rorden, C. (2011). Transcranial direct current stimulation improves naming reaction time in fluent aphasia: a double-blind, sham-controlled study*. Stroke, 42*(3), 819-821.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., & Penny, W. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*: Academic Press.

Geranmayeh, F., Brownsett, S. L. E., & Wise, R. J. S. (2014). Task-induced brain activity in aphasic stroke patients: what is driving recovery? *Brain, 137,* 2632-2648.

Geva, S., Correia, M., & Warburton, E. A. (2011). Diffusion tensor imaging in the study of language and aphasia. *Aphasiology, 25*(5), 543-558.

Gili, T., Fiori, V., De Pasquale, G., Sabatini, U., Caltagirone, C., & Marangolo, P. (2017). Right sensory-motor functional networks subserve action observation therapy in aphasia*. Brain Imaging and Behaviour, 11*, 1397-1411.

Gill, S. K., & Leff, A. P. (2014). Dopaminergic therapy in aphasia. *Aphasiology, 28*(2), 155-170.

Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and language, 40*(1), 67-88.

Gold, B. T., & Kertesz, A. (2000). Preserved visual lexicosemantics in global aphasia: A right-hemisphere contribution? *Brain and Language, 75*(3), 359-375.

Gold, B. T., Balota, D. A., Kirchhoff, B. A., & Buckner, R. L. (2005). Common and dissociable activation patterns associated with controlled semantic and phonological processing: evidence from fMRI adaptation. *Cerebral Cortex, 15*(9), 1438-1450.

Goldenberg, G., & Spatt, J. (1994). Influence of size and site of cerebral lesions on spontaneous recovery of aphasia and on success of language therapy. B*rain and Language, 47*(4), 684-698.

Goodglass, H., & Kaplan, E. (1983). The assessment of aphasia and related disorders (2nd ed.). Philadelphia: Lea & Febiger.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). The assessment of aphasia and related disorders (3rd ed.). Austin: Pro-Ed.

Goodglass, H., Kaplan, E., & Weintraub, S. (2001). BDAE: The Boston Diagnostic Aphasia Examination. Philadelphia, PA: Lippincott Williams & Wilkins.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., . . . Boeve, B. F. (2011). Classification of primary progressive aphasia and its variants. *Neurology, 76*(11), 1006-1014.

Gow, D. W., & Ahlfors, S. P. (2017). Tracking reorganization of large-scale effective connectivity in aphasia following right hemisphere stroke. *Brain and Language, 170*, 12-17.

Grechuta, K., Rubio Ballester, B., Espin Munne, R., Usabiaga Bernal, T., Molina Hervás, B., Mohr, B., ... & Verschure, P. (2019). Augmented dyadic therapy boosts recovery of language function in patients with nonfluent aphasia: a randomized controlled trial. *Stroke, 50*(5), 1270-1274.

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2017). Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex, 86*, 275-289.

Hara, T., Abo, M., Kakita, K., Mori, Y., Yoshida, M., & Sasaki, N. (2017). The Effect of Selective Transcranial Magnetic Stimulation with Functional Near-Infrared Spectroscopy and Intensive Speech Therapy on Individuals with Post-Stroke Aphasia. *Eur Neurol, 77*, 186-194.

Hara, T., Abo, M., Kobayashi, K., Watanabe, M., Kakuda, W., & Senoo, A. (2015). Effects of Low-Frequency Repetitive Transcranial Magnetic Stimulation Combined with Intensive Speech Therapy on Cerebral Blood Flow in Post-Stroke Aphasia. *Trans. Stroke. Res., 6*, 365-374.

Hartwigsen, G., & Saur, D. (2019). Neuroimaging of stroke recovery from aphasia–Insights into plasticity of the human language network. *NeuroImage, 190,* 14-31.

Heath, C., Conroy, P., Pellegrini, I., & Bruehl, S. (in preparation), Investigating the neural basis of post-stroke aphasia therapy; a systematic review of neuroscience-based therapy studies.

Heath, C., Conroy, P., Stewart, A., Bruehl, S. (in preparation), A comparison of different approaches to behavioural treatment of chronic post-stroke aphasia: Results and Discussion

Heiss, W.-D., & Thiel, A. (2006). A proposed regional hierarchy in recovery of post-stroke aphasia. *Brain and Language, 98*(1), 118-123.

Hilari, K., Galante, L., Huck, A., Pritchard, M., Allen, L. and Dipper, L. (2018). Cultural adaptation and psychometric testing of The Scenario Test UK for people with aphasia. *International Journal of Language & Communication Disorders*, doi: 10.1111/1460-6984.12379

Hocking, J., & Price, C. J. (2009). Dissociating verbal and nonverbal audiovisual object processing. *Brain and language, 108*(2), 89-96.

Hodgson, C., & Lambon Ralph, M. A. (2008). Mimicking aphasic semantic errors in normal speech production: Evidence from a novel experimental paradigm. *Brain and language, 104*(1), 89-101.

Hoeft, F., Barnea-Goraly, N., Haas, B. W., Golarai, G., Ng, D., Mills, D., . . . Reiss, A. L. (2007). More is not always better: increased fractional anisotropy of superior longitudinal fasciculus associated with poor visuospatial abilities in Williams syndrome. *Journal of Neuroscience, 27*(44), 11960-11965.

Hoffmann, M., & Chen, R. (2013). The Spectrum of Aphasia Subtypes and Etiology in Subacute Stroke. *Journal of Stroke and Cerebrovascular Diseases, 22*(8), 1385-1392.

Howard, D. (2000). Cognitive neuropsychology and aphasia therapy: The case of word retrieval. In I. Papathanasiou (Ed.), Acquired neurogenic communication disorders: A clinical perspective. London and Philadelphia: Whurr.

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology, 20*(9), 921-950.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Ioannidis, J. P. (2005). Why most published research findings are false*. PLoS medicine, 2*(8), e124.

Jackson, R. L., Hoffman, P., Pobric, G., & Lambon Ralph, M. A. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. *Journal of Neuroscience, 36*(5), 1490-1501.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain, 129,* 2132-2147.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17*(2), 825-841.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*(2), 782-790.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical image analysis, 5*(2), 143-156.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). Boston naming test. Philadelphia: Lea & Febiger.

Kasselimis, D. S., Simos, P. G., Peppas, C., Evdokimidis, I., & Potagas, C. (2017). The unbridged gap between clinical diagnosis and contemporary research on aphasia: A short discussion on the validity and clinical utility of taxonomic categories. *Brain & Language, 164*, 63-67.

Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology, 10*(2), 159-180.

Kemerdere, R., de Champfleur, N. M., Deverdun, J., Cochereau, J., Moritz-Gasser, S., Herbet, G., & Duffau, H. (2016). Role of the left frontal aslant tract in stuttering: a brain stimulation and tractographic study. *Journal of neurology, 263*(1), 157-167.

Kertesz, A. (1982). The Western Aphasia Battery. New York: Grune and Stratton, Inc.

Kiran, S. (2012). What is the nature of poststroke language recovery and reorganization? *ISRN neurology*, 2012.

Kiran, S., & Thompson, C. K. (2019). Neuroplasticity of language networks in aphasia: Advances, updates, and future challenges. *Frontiers in Neurology, 10,* 295.

Kiran, S., Meier, E. L., Kapse, K. J., & Glynn, P. A. (2015). Changes in task-based effective connectivity in language networks following rehabilitation in post-stroke patients with aphasia*. Front. Hum. Neurosci., 9*(316).

Kroenke, K.-M., Kraft, I., Regenbrecht, F., & Obrig, H. (2013). Lexical learning in mild aphasia: Gesture benefit depends on patholinguistic profile and lesion pattern. *Cortex, 49*, 2637-2649.

Kummerer, D., Hartwigsen, G., Kellmeyer, P., Glauche, V., Mader, I., Kloppel, S., . . . Saur, D. (2013). Damage to ventral and dorsal language pathways in acute aphasia. *Brain, 136*, 619-629.

Lam, J. M. C., & Wodchis, W. P. (2010). The Relationship of 60 Disease Diagnoses and 15 Conditions to Preference-Based Health-Related Quality of Life in Ontario Hospital-Based Long-Term Care Residents. *Med Care, 48*, 380-387.

Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post CVA: both language and cognitive status are key predictors. *Neuropsychological Rehabilitation, 20*(2), 289-305.

Leff, A. P., Crinion, J. T., Scott, S., Turkheimer, F., Howard, D., & Wise, R. (2002). A physiological change in the homotopic cortex following left posterior temporal lobe infarction. *Annals of neurology, 51*(5), 553-558.

Leonard, C., Laird, L., Burianova, H., Graham, S., Grady, C., Simic, T., & Rochon, E. (2015). Behavioural and neural changes after a "choice" therapy for naming deficits in aphasia: preliminary findings. *Aphasiology, 29*(4), 506-525.

Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology, 22*(9), 923-947.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences, 22*, 1-75.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 570.

Lutkenhoff, E. S., Rosenberg, M., Chiang, J., Zhang, K., Pickard, J. D., Owen, A. M., & Monti, M. M. (2014). Optimized brain extraction for pathological brains (optiBET). *PloS one, 9*(12), e115551.

Lystad, R. P., & Pollard, H. (2009). Functional neuroimaging: a brief overview and feasibility for use in chiropractic research. *The Journal of the Canadian Chiropractic Association, 53*(1), 59.

Maher, L. M., & Raymer, A. M. (2004). Management of anomia. *Topics in stroke rehabilitation, 11*(1), 10-21.

Maeshima, S., Kuwata, T., Uematsu, Y., Itakura, T., & Komai, N. (1994). Transcortical mixed aphasia in a case of thalamic glioblastoma. *Aphasiology, 8*(3), 197-204.

Mah, Y. H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain*, *137*(9), 2522-2531.

Maher, L. M., & Raymer, A. M. (2004). Management of anomia. *Topics in stroke rehabilitation, 11*(1), 10-21.

Marangolo, P., Fiori, V., Sabatini, U., De Pasquale, G., Razzano, C., Caltagirone, C., & Gili, T. (2016). Bilateral transcranial direct current stimulation language treatment enhances functional connectivity in the left hemisphere: preliminary data from aphasia. *Journal of Cognitive Neuroscience.*

Marcotte, K., Adrover-Roig, D., Damien, B., Preaumont, M. d., Genereux, S., Hubert, M., & Ansaldo, A. I. (2012). Therapy-induced neuroplasticity in chronic aphasia. *Neuropsychologia, 50*, 1776-1786.

Marcotte, K., Perlbarg, V., Marrelec, G., Benali, H., & Ansaldo, A. I. (2013). Default-mode network functional connectivity in aphasia: Therapy-induced neuroplasticity. *Brain & Language, 124,* 45-55.

Martin, P. I., Naeser, M. A., Theoret, H., Tormos, J. M., Nicholas, M., Kurland, J., . . . Pascual-Leone, A. (2004). Transcranial magnetic stimulation as a complementary treatment for aphasia. *Paper presented at the Seminars in speech and language.*

McKinnon, E. T., Fridriksson, J., Glenn, G. R., Jensen, J. H., Helpern, J. A., Basilakos, A., . . . Bonilha, L. (2017). Structural Plasticity of the Ventral Stream and Aphasia Recovery. *ANN NEUROL, 82*, 147-151.

Meinzer, M., Flaisch, T., Breitenstein, C., Wienbruch, C., Elbert, T., & Rockstroh, B. (2008). Functional re-recruitment of dysfunctional brain areas predicts language recovery in chronic aphasia. *NeuroImage, 39,* 2038-2046.

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

Meinzer, M., Paul, I., Wienbruch, C., Djundja, D., & Rockstroh, B. (2009). Electromagnetic brain activity in higher frequency bands during automatic word processing indicates recovery of function in aphasia. *Eur J Phys Rehabil Med, 45*, 369-378.

Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgartner, A., Schiffbauer, H., . . . Breitenstein, C. (2009). Imaging short- and long-term training success in chronic aphasia. *BMC Neuroscience, 10*(118).

Mitchell, R. L. C., & Crow, T. J. (2005). Right hemisphere language functions and schizophrenia: the forgotten hemisphere? *Brain, 128*, 963-978.

Mohr, B., Difrancesco, S., Harrington, K., Evans, S., & Pulvermuller, F. (2014). Changes of right-hemipsheric activation after constraint-induced, intensive language action therapy in chronic aphasia: fMRI evidence from auditory semantic processing. *Front. Hum. Neurosci., 8*(919).

Mohr, B., MacGregor, L. J., Difrancesco, S., Harrington, K., Pulvermuller, F., & Shtyrov, Y. (2016). Hemispheric contributions to language reorganisation: An MEG study of neuroplasticity in chronic post stroke aphasia. *Neuropsychologia, 93*, 413-424.

Monti, A., Ferrucci, R., Fumagalli, M., Mameli, F., Cogiamanian, F., Ardolino, G., & Priori, A. (2013). Transcranial direct current stimulation (tDCS) and language. *J Neurol Neurosurg Psychiatry, 84*, 832-842.

Moore, E., Schaefer, R. S., Bastin, M. E., Roberts, N., & Overy, K. (2017). Diffusion tensor MRI tractography reveals increased fractional anisotropy (FA) in arcuate fasciculus following music-cued motor therapy. *Brain and Cognition, 116*, 40-46.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour, 1*(1), 1-9.

Musso, M., Weiller, C., Kiebel, S., Müller, S. P., Bülau, P., & Rijntjes, M. (1999). Training-induced brain plasticity in aphasia. *Brain, 122*(9), 1781-1790.

Naeser, M. A., Martin, P. I., Baker, E. H., Hodge, S. M., Sczerzenie, S. E., Nicholas, M., . . . Samaraweera, R. (2004). Overt propositional speech in chronic nonfluent aphasia studied with the dynamic susceptibility contrast fMRI method. *NeuroImage, 22*(1), 29-41.

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Helm-Estabrooks, N., . . . Fregni, F. (2005). Improved naming after TMS treatments in a chronic, global aphasia patient–case report. *Neurocase, 11*(3), 182-193.

Naeser, M. A., Martin, P. I., Nicholas, M., Baker, E. H., Seekins, H., Kobayashi, M., . . . Kurland, J. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: an open-protocol study*. Brain and Language, 93*(1), 95-105.

Naeser, M. A., Martin, P. I., Theoret, H., Kobayashi, M., Fregni, F., Nicholas, M., . . . Pascual-Leone, A. (2011). TMS suppression of right pars triangularis, but not pars opercularis, improves naming in aphasia. *Brain and language, 119*(3), 206-213.

Nardo, D., Holland, R., Leff, A. P., Price, C. J., & Crinion, J. T. (2017). Less is more: neural mechanisms underlying anomia treatment in chronic aphasic patients. *Brain, 140*(11), 3039-3054.

Nenert, R., Allendorfer, J. B., Martin, A. M., Banks, C., Ball, A., Vannest, J., . . . Szaflarski, J. P. (2017). Neuroimaging Correlates of Post-Stroke Aphasia Rehabilitation in a Pilot Randomized Trial of Constraint-Induced Aphasia Therapy. *Med Sci Monit, 23,* 3489-3507.

Nord, C. L., Valton, V., Wood, J., & Roiser, J. P. (2017). Power-up: a reanalysis of power failure in neuroscience using mixture modeling. *Journal of Neuroscience, 37*(34), 8051-8061.

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy. *Annals of the New York Academy of Sciences, 1169*(1), 431-436.

Nosek, B. A., Aarts, A. A., Anderson, J. E., Kappes, H. B., & Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).

Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., ... & Chater, T. (2019). Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (Big CACTUS): a multicentre, single-blinded, randomised controlled trial. *The Lancet Neurology, 18*(9), 821-833.

Palmer, R., Witts, H., & Chater, T. (2018). What speech and language therapy do community dwelling stroke survivors with aphasia receive in the UK? *PloS one, 13(*7).

Parker, G. J. M., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A. M., Ciccarelli, O., & Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage, 24*, 656-666.

Pascual-Leone, A., Walsh, V., & Rothwell, J. (2000). Transcranial magnetic stimulation in cognitive neuroscience- virtual lesion, chronometry, and functional connectivity. *Current Opinion in Neurobiology, 10*, 232-237.

Patterson, K., Graham, N., Ralph, M. A. L., & Hodges, J. (2006). Progressive non-fluent aphasia is not a progressive form of non-fluent (post-stroke) aphasia. *Aphasiology, 20*(9), 1018-1034.

Peck, K. K., Moore, A. B., Crosson, B. A., Gaiefsky, M., Gopinath, K. S., White, K., & Briggs, R. W. (2004). Functional Magnetic Resonance Imaging Before and After Aphasia Therapy: Shifts inHemodynamic Time to Peak During an Overt Language Task. *Stroke, 35*, 554-559.

Perani, D., Cappa, S., Tettamanti, M., Rosa, M., Scifo, P., Miozzo, A., . . . Fazio, F. (2003). A fMRI study of word retrieval in aphasia. *Brain and Language, 85*(3), 357-368.

Pierce, J. E., Menahemi-Falkov, M., O'Halloran, R., Togher, L., & Rose, M. L. (2017). Constraint and multimodal approaches to therapy for chronic aphasia: A systematic review and meta-analysis. *Neuropsychological Rehabilitation.*

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains*. Psychological review, 103*(1), 56.

Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of evaluation in clinical practice, 18*(3), 689-694.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage, 62*(2), 816-847.

Pulvermüller, F., & Berthier, M. L. (2008). Aphasia therapy on a neuroscience basis. *Aphasiology, 22*(6), 563-599.

Pulvermuller, F., Hauk, O., Zohsel, K., Neininger, B., & Mohr, B. (2005). Therapy-related reorganization of language in both hemispheres of patients with chronic aphasia. *NeuroImage, 28*, 481-489.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*(20), 7865-7870.

Pulvermuller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint-Induced Therapy of Chronic Aphasia After Stroke. *Stroke, 32*, 1621-1626.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raboyeau, G., De Boissezon, X., Marie, N., Balduyck, S., Puel, M., Bezy, C., . . . Cardebat, D. (2008). Right hemisphere activation in recovery from aphasia: lesion effect or function recruitment? *Neurology, 70*(4), 290-298.

Richter, M., Miltner, W. H. R., & Straube, T. (2008). Association between therapy outcome and right-hemispheric activation in chronic aphasia. *Brain, 131,* 1391-1401.

Rosen, H., Petersen, S., Linenweber, M., Snyder, A., White, D., Chapman, L., . . . Corbetta, M. (2000). Neural correlates of recovery from aphasia after damage to left inferior frontal cortex. *Neurology, 55*(12), 1883-1894.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*(2), 217-236.

Sarasso, S., Santhanam, P., Määtta, S., Poryiazova, R., Ferrarelli, F., Tononi, G., & Small, S. L. (2010). Non-fluent aphasia and neural reorganization after speech therapy: insights from human sleep electrophysiology and functional magnetic resonance imaging. *Archives italiennes de biologie, 148*(3), 271.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., . . . Abel, S. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences, 105*(46), 18035-18040.

Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain, 129*(6), 1371-1384.

Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for Plasticity in White-Matter Tracts of Patients with Chronic Broca's Aphasia Undergoing Intense Intonation-based Speech Therapy. *Paper presented at the The Neurosciences and Music III- Disorders and Plasticity, McGill Univ, Montreal, CANADA.*

Schlaug, G., Norton, A., Marchina, S., Zipse, L., & Wan, C. Y. (2010). From singing to speaking: facilitating recovery from nonfluent aphasia. *Future Neurol, 5*(5), 657-665.

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of memory and language, 54*(2), 228-264.

Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain, 135*(12), 3799-3814.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400-2406.

Sebastian, R., & Breining, B. L. (2018). Contributions of Neuroimaging to Understanding Language Deficits in Acute Stroke. *Paper presented at the Seminars in speech and language.*

Seghier, M. L., Ramlackhansingh, A., Crinion, J., Leff, A. P., & Price, C. J. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *Neuroimage, 41*(4), 1253-1266.

Sharp, M. E., Foerde, K., Daw, N. D., & Shohamy, D. (2016). Dopamine selectively remediates 'model-based'reward learning: a computational approach. *Brain, 139(*2), 355-364.

Sidaros, A., Engberg, A. W., Sidaros, K., Liptrot, M. G., Herning, M., Petersen, P., . . . Rostrup, E. (2008). Diffusion tensor imaging during recovery from severe traumatic brain injury and relation to clinical outcome: a longitudinal study. *Brain, 131*(2), 559-572.

Simanovic, O., Mrkonjic, Z., Zukic, S., Vidovic, M., & Imamovic, K. (2011). Post-stroke language disorders. *Acta Clin Croat, 50*, 79-94.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17*(3), 143-155.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., . . . Flitney, D. E. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23,* S208-S219.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory, 6*(2), 174.

Spielmann, K., Durand, E., Marcotte, K., & Ansaldo, A. I. (2016). Maladaptive Plasticity in Aphasia: Brain Activation Maps Underlying Verb Retrieval Errors. *Neural plasticity, 2016*.

Spitzer, L., Binkofski, F., Willmes, K. & Bruehl, S. (2021). The novel Cognitive Flexibility in Aphasia Therapy (CFAT): A combined treatment of aphasia and executive functions to improve communicative success. *International Journal of Speech-Language Pathology, 23(2),* 168-179.

Spitzer, L., Binkofski, F., Willmes, K. & Bruehl, S. (2021). The novel Cognitive Flexibility in Aphasia Therapy (CFAT): A combined treatment of aphasia and executive functions to improve communicative success. *International Journal of Speech-Language Pathology, 23(2),* 168-179.

Sridharan, D., Levitin, D. J., & Menon, V. (2008). A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences, 105*(34), 12569-12574.

Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Frontiers in Human Neuroscience, 7*.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain, 134*(10), 3083-3093.

Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G., & Pulvermuller, F. (2017). Communicative-Pragmatic Assessment Is Sensitive and Time-Effective in Measuring the Outcome of Aphasia Therapy. *Front. Hum. Neurosci., 11*(223).

Stark, B. C., & Warburton, E. A. (2018). Improved language in chronic aphasia after self-delivered iPad speech therapy. *Neuropsychological rehabilitation, 28*(5), 818-831.

Stoeckel, C., Gough, P. M., Watkins, K. E., & Devlin, J. T. (2009). Supramarginal gyrus involvement in visual word recognition*. Cortex, 45*(9), 1091-1096.

Swinburn, K., Porter, G., & Howard, D. (2005). Comprehensive aphasia test. Hove: Psychology Press.

Szaflarski, J. P., Vannest, J., Wu, S. W., DiFrancesco, M. W., Banks, C., & Gilbert, D. L. (2011). Excitatory repetitive transcranial magnetic stimulation induces improvements in chronic post-stroke aphasia. *Medical science monitor: international medical journal of experimental and clinical research, 17*(3), CR132.

Takamura, T., & Hanakawa, T. (2017). Clinical utility of resting-state functional connectivity magnetic resonance imaging for mood and cognitive disorders. *Journal of Neural Transmission, 124*(7), 821-839.

Thompson, C. K. (2000). Neuroplasticity: Evidence from aphasia. *Journal of communication disorders, 33*(4), 357.

Thompson, C. K., & den Ouden, D.-B. (2008). Neuroimaging and recovery of language in aphasia. *Current neurology and neuroscience reports, 8*(6), 475.

Thulborn, K. R., Carpenter, P. A., & Just, M. A. (1999). Plasticity of language-related brain function during recovery from stroke. *Stroke, 30*(4), 749-754.

Tippett, D. C. (2015). Update in Aphasia Research. *Curr Neurol Neurosci Rep, 15*(49).

Tippett, D. C., Niparko, J. K., & Hillis, A. E. (2014). Aphasia: Current concepts in theory and practice. *Journal of neurology & translational neuroscience, 2*(1), 1042.

Tochadse, M., Halai, A. D., Lambon Ralph, M. A., & Abel, S. (2018). Unification of behavioural, computational and neural accounts of word production errors in post-stroke aphasia. *NeuroImage: Clinical, 18*, 952-962.

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. A. (2011). Lichtheim 2: Synthesizing Aphasia and the Neural Basis of Language in a Neurocomputational Model of the Dual Dorsal-Ventral Language Pathways. *Neuron, 72*, 385-396.

van Casteren, M., & Davis, M. H. (2007). A program to assist in matching the conditions of factorial experiments. *Behaviour Research Methods, 39*(4), 973-978.

van de Sandt-Koenderman, M. W. M. E., Mendez Orellana, C. P., van der Meulen, I., Smits, M., & Ribbers, G. M. (2016). Language lateralisation after Melodic Intonation Therapy: an fMRI study in subacute and chronic aphasia. *Aphasiology.*

van der Meulen, I., van de Sandt-Koenderman, W. M. E., Duivenvoorden, H. J., & Ribbers, G. M. (2010). Measuring verbal and non-verbal communication in aphasia: reliability, validity, and sensitivity to change of the Scenario Test. *International Journal of Language & Communication Disorders, 45*(4), 424-435.

van Hees, S., McMahon, K., Angwin, A., de Zubicaray, G., Read, S., & Copland, D. A. (2014). A Functional MRI Study of the Relationship Between Naming Treatment Outcomes and Resting State Functional Connectivity in Post-Stroke Aphasia. *Human Brain Mapping, 35,* 3919-3931.

van Hees, S., McMahon, K., Angwin, A., de Zubicaray, G., Read, S., & Copland, D. A. (2014). Changes in white matter connectivity following therapy for anomia post stroke. *Neurorehabilitation and neural repair, 28*(4), 325-334.

Visser, M., & Lambon Ralph, M. A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. *Journal of Cognitive Neuroscience, 23*(10), 3121-3131.

Vitali, P., Abutalebi, J., Tettamanti, M., Danna, M., Ansaldo, A.-I., Perani, D., . . . Cappa, S. F. (2007). Training-induced brain remapping in chronic aphasia: a pilot study. *Neurorehabilitation and Neural Repair, 21*(2), 152-160.

Wan, C. Y., Zheng, X., Marchina, S., Norton, A., & Schlaug, G. (2014). Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca's aphasia. *Brain & Language, 136*, 1-7.

Warburton, E. A., Price, C. J., Swinburn, K., & Wise, R. J. S. (1999). Mechanisms of recovery from aphasia: evidence from positron emission tomography studies. *Journal of Neurology, Neurosurgery & Psychiatry, 66*(2), 155-161.

Watkins, K. E., & Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of Cognitive Neuroscience, 16*(6), 978-987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994.

Weiller, C., Isensee, C., Rijntjes, M., Huber, W., Müller, S. P., Bier, D., . . . Diener, H. C. (1995). Recovery from Wernicke's aphasia: a positron emission tomographic study. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 37*(6), 723-732.

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

Welbourne, S. R., Woollams, A. M., Crisp, J., & Lambon Ralph, M. A. (2011). The role of plasticity-related functional reorganization in the explanation of central dyslexias. *Cognitive neuropsychology, 28*(2), 65-108.

Whitworth, A., Webster, J., & Howard, D. (2014). A cognitive neuropsychological approach to assessment and intervention in aphasia: A clinician's guide. Psychology Press.

Williams, N. R., Bentzley, B. S., Hopkins, T., Pannu, J., Sahlem, G. L., Takacs, I., . . . Short, E. B. (2018). Optimization of epidural cortical stimulation for treatment-resistant depression. *Brain stimulation, 11*, 239-240.

Xu, J., Wang, J., Fan, L., Li, H., Zhang, W., Hu, Q., & Jiang, T. (2015). Tractography-based parcellation of the human middle temporal gyrus. *Scientific reports, 5,* 18883.

Yagata, S. A., Yen, M., McCarron, A., Bautista, A., Lamair-Orosco, G., & Wilson, S. M. (2017). Rapid recovery from aphasia after infarction of Wernicke's area. *Aphasiology, 31*(8), 951-980.

Yang, M., Li, Y., Li, J., Yao, D., Liao, W., & Chen, H. (2017). Beyond the Arcuate Fasciculus: Damage to Ventral and Dorsal Language Pathways in Aphasia. *Brain Topogr, 30*, 249-256.

Zahn, R., Schwarz, M., & Huber, W. (2006). Functional activation studies of word processing in the recovery from aphasia. *Journal of Physiology-Paris, 99*(4-6), 370-385.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature neuroscience, 15*(4), 528.

Zumbansen, A., Peretz, I., & Hébert, S. (2014). The combination of rhythm and pitch can account for the beneficial effect of Melodic Intonation Therapy on connected speech improvements in Broca’s aphasia. *Frontiers in Human Neuroscience, 8.*