Extending the Scope of Sentence Therapy in Aphasia Across English and Arabic

A thesis submitted to the University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Biology, Medicine and Health

2022

Nourah Alohali

School of Biological Sciences

Division of Neuroscience and Experimental Psychology

Contents

Word count: 77,367

List of tables

List of figures

List of abbreviations

SLT/P Speech-Language Therapist/Pathologist

SLT Speech Language Therapy

PWA People with aphasia

BDAE Boston Diagnostic Aphasia Examination

TMA Transcortical Motor Aphasia

VNeST Verb Network Strengthening Therapy

VAST Verb and Sentence Test

WAB Western Aphasia Battery

LVET Light Verb Elicitation Test

BNT Boston Naming Test

PALPA Psycholinguistic Assessment of Language Processing in Aphasia

RCPM Ravens Colored Progressive Matrices

NARU Neuroscience and Aphasia Research Unit

REC National Research Ethics Committee

AAC Augmentative and Alternative Communication

Abstract

Background: Sentence production deficits in aphasia still are under-investigated, especially in non-English speaking countries, and computer-based approaches increasingly important. The thesis will examine the feasibility of a novel method for remediating sentence production deficits in people with aphasia post-stroke in two languages, English and Arabic. It consists of a multilevel theory-driven hybrid approach that integrates three methods: mapping therapy (Schwartz et al., 1994), Verb Network Strengthening Therapy VNeST (Edmonds et al., 2009), and processing prosthesis for temporal window widening, which is central to SentenceShaper computer software (Linebarger et al., 2001). It was designed to allow flexible application across a wide range of aphasia severity and subtypes. **Methods**: A case series study design and a single-case experimental design with multiple assessment points were implemented. The intervention combined clinician-delivered weekly sessions with independent home practice delivered to 16 participants over 8 to 12 weeks. A range of well-known assessment tools such as Verb and Sentence Test (VAST) (Bastiaanse et al., 2002), the Western Aphasia Battery (WAB) (Kertesz, 2007), and discourse elicitation tasks (cookie theft and dinner party picture stimuli) were used as outcome measures. **Results**: The findings revealed robust improvements in sentence and discourse production skills, in both English and Arabic. Also, generalisation of therapy gains to untargeted skills such as lexical retrieval and comprehension at the word level and sentence comprehension was noted in some participants. **Discussion**: The outcomes supported our hypothesis that our combined approach can produce comparable therapy gains, in sentence and discourse production, to those obtained by implementing each original therapy protocol in isolation, as described in the literature. We consider the feasibility of the combined approach superior to individual application as it reduced the intensity of clinician-directed therapy sessions and fostered independence in home-practice. Also, the user-friendly technology we incorporated as a platform to deliver therapy, across very different languages, added a unique component for practicing self-monitoring and self-correction skills. As a result, the dose of language practice, which is known to be associated with better therapy gains (Cherney et al., 2011), can be increased in an accessible and cost-effective way.

The thesis also presents preliminary work in creating Arabic language assessment tools and therapy materials. The findings highlight both the potential successes that can be achieved in cultural adaptation and translation within rehabilitation research as well as the many unforeseen challenges in applying methods derived in Western culture to diverse cultures world-wide. Finally, the thesis considers the current barriers to increasing the reliability and validity of assessment tasks and suggests ways to address them in future research.

Keywords: Aphasia, sentence production, therapy, Arabic, English

Declaration

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

Copyright statement

The Author

Nourah Alohali was born and raised in Saudi Arabia. She qualified as a Speech-Language Pathologist in 2010 after graduating with a Bachelor of Science degree in Speech-Language Pathology and Audiology from King Saud University, Riyadh, Saudi Arabia. Her work experience includes three years of delivering outpatient and inpatient speech and language therapy services to adults with acquired neurogenic disorders at King Saud University Medical City, a major tertiary care academic medical centre in Riyadh, Saudi Arabia.

To become an effective contributor to the speech and language therapy field in Saudi Arabia, she pursued advanced studies and clinical training in the United States. In 2016, she graduated with a Master of Science from Northwestern University, Evanston, Illinois. The coursework involved working as a full-time graduate student clinician in world-renowned hospitals and rehabilitation centres in Chicago, including the Shirley Ryan AbilityLab, the Ann & Robert H. Lurie Children's Hospital of Chicago and Northwestern Memorial Hospital. This valuable opportunity allowed her to practise in two different countries, speaking two distinct languages, which triggered her constant reflection and comparison. As a result, she arrived at a vision for optimising the quality of speech-language pathology services in her home country. However, the scarce literature on Arabic speakers with acquired language disorders and the lack of standardised assessment tools was a clear obstacle. In 2017, she therefore registered for a full-time PhD programme at the University of Manchester, UK. In her PhD work, she aimed to acquire knowledge and experience to support her future translational research projects and gain insight from her collaboration with the outstanding scientists who contributed to her PhD projects.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my academic supervisor, Dr Paul Conroy. Without his perception, patience, encouragement, unwavering support and guidance, this work could not have been achieved. I am extremely grateful to Dr Stefanie Bruehl for her extensive knowledge and helpful contributions throughout my PhD studies.

This research was made possible by funding awarded by King Saud University, Riyadh, Saudi Arabia. I am grateful to those who believed in my cause and capability and supported my scholarship application.

I wish to express my profound thanks to Dr Marcia Linebarger for creating an Arabic-compatible version of SentenceShaper® and providing me with a copy to use in my project. The project could not have been delivered to the desired standard without this integral contribution. I also wish to thank Prof Dr Roelien Bastiaanse for giving me permission to translate and adapt the Verb and Sentence Test (VAST) into Arabic for pilot testing. Thanks also to Professor Yaron Matras for his insightful suggestions and contribution to the Arabic project.

Special thanks to the Neuroscience and Aphasia Research Unit (NARU) for invaluable research assistance and database access and the SpeakEasy aphasia support group for offering me a recruitment and testing site. I am indebted to the patients and their families who generously gave their time to this study. This project could not have been completed without their contribution and commitment. It has been an honour and a pleasure to work with them.

Many thanks go to Blanca De Dios Perez for her help with the data collection, to Santa Ozolinja, Padraig Roper and Ilaria Pellegrini for their help with the English discourse transcription and coding and to Nora Fahad Al-Sudairi, Dima Taibah, Maryam Al-Resheid and Abdulrahman Bahmaid for their help with the translation and adaptation of the Arabic assessment and therapy materials.

I am immensely grateful to my parents, Nawal Alajaji and Yousef Alohali, for their love, care and support and, most of all, for always believing in me, which gave me the courage to aim high and the confidence to achieve my vision. I am deeply indebted to my aunt Nora Alajaji for generously opening her heart and home to me when I had to pursue my education away from home. I would also like to express my deepest gratitude to my uncle Hamad Alkhowaiter for being a role model and my greatest inspiration.

Finally, thank you to all my friends in Manchester, who made this journey fun, joyful and unforgettable.

CHAPTER 1 Introduction to the thesis (overview of structure and contents)

1.1 Thesis overview and structure

This PhD thesis consists of seven chapters containing four empirical studies presented in the 'traditional' thesis format. This introductory Chapter 1 presents an overview of the thesis' content, structure, and aims. Chapter 2 will set out the background and literature review that motivated our experimental work. It will include a description of the hypotheses of the source of sentence production deficits in aphasia, the theory-driven therapy approaches designed to address them (both paper-based and computer-based), and the therapy-delivery models adopted by speech-language therapists/pathologists SLT/Ps in clinical practice. It also summarizes the main features that distinguish Arabic from English in healthy adults and people with aphasia, which justifies the translation and adaptation of assessment and therapy materials for the Arabic study and explains the challenges. Chapter 3 contains the first empirical study in this thesis, which presents a systematic and sensitive evaluation of sentence production skills across a sample of 29 native English participants with a range of aphasia subtypes and severities. It highlights the process that guided our candidacy selection for the English therapy study. Chapter 4 presents the second empirical study that tested the feasibility of a systematic application of a novel computer-based sentence therapy across a sample of 12 native English participants with a range of aphasia subtypes and severities, a subset of the sample of 29 participants mentioned previously. Chapter 5 accounts for the novelty of the assessment tools and therapy materials introduced in the subsequent Arabic study. It will explain the process we adopted in developing these Arabic tools and materials. Chapter 6 will set out the final empirical study investigating the implementation of the Arabic assessment and therapy tools to identify the characteristics of language production deficits in Arabic speakers with aphasia and examine their response to the sentence therapy program, an equivalent to the program presented in the English study. Those core chapters (Chapter 3, 4, 5 and 6) are written in a journal paper style with some modifications. They each start with a review of the relevant literature that motivated the empirical investigation, then state the research questions, describe the methods, present the results, and conclude with a comprehensive discussion. However, in the method section of each empirical chapter, to avoid repetition due to some overlap in the assessment tools used in the three studies, the reader is referred to previous chapters for the sake of brevity. Also, the reader is occasionally referred to the literature review in the introduction chapter for further readings on relevant topics. Finally, Chapter 7 contains the final discussion that summarises the thesis findings, draws together the theoretical and clinical implications of the research, highlights the study limitations, and presents directions for future research.

1.2 Research questions to be addressed in this thesis

1. What is the current evidence base for sentence therapy and therapy in aphasia? Are the various treatment approaches separate and distinct or is there potential for integration and formation of hybrid therapies? Is there potential for technology?
2. What is the candidacy for sentence therapy across a wide range of people with aphasia, including varying degrees of severity and subtypes?
3. What are the outcomes for a hybrid sentence production therapy across a range of aphasia severity and subtype with respect to lexical, sentence and discourse outcomes?
4. What cultural adaptations and translations need to be implemented in order to apply the hybrid approach to the Saudi Arabic language?
5. What are the outcomes for this hybrid sentence production therapy in a small sample of Arabic speaking people with aphasia? How do these compare to English language results?

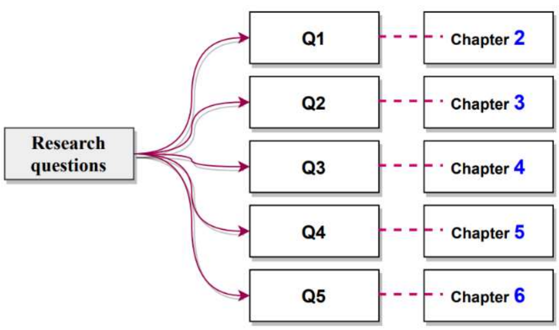


Figure 1 Illustration of the chapters within the thesis as they answered each research question

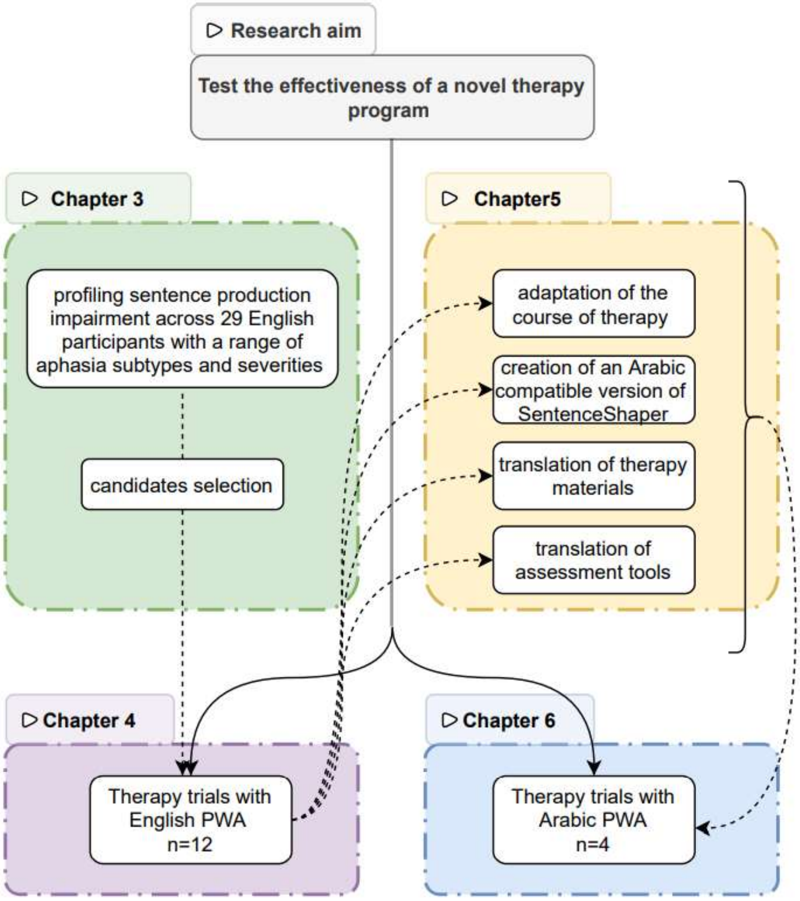


Figure 2 Illustration of how all 4 empirical studies contributed to testing the feasibility of the novel aphasia therapy method in English and Arabic

1.3 Study design

A case series design was implemented in the two empirical studies presented in Chapter 3 and Chapter 4. The participants completed the same treatment program and underwent the same background cognitive and linguistic assessment. This approach enabled comparisons both within and between participants and correlations of their therapy outcomes to baseline language and cognitive profiles. On the other hand, due to the small sample size in the fourth empirical study, presented in Chapter 6, a single-case experimental design with multiple assessment points has been implemented. It enabled the evaluation of therapy-induced changes in language production skills from baseline at multiple time-points throughout the course of therapy.

1.4 Data collection and analysis

The primary investigator administered all the testing, data collection, and data analysis in the four empirical studies, with three exceptions. The cognitive linguistic data labelled "background neuropsychological assessment data" in Chapter 3 and Chapter 4 were not collected by the primary investigator. It was retrieved from a pre-established database of people with aphasia at the Neuroscience and Aphasia Research Unit (NARU) at the University of Manchester. Also, the discourse samples in Chapter 4 were collected by the primary investigator but transcribed by a student collaborator from the University of Manchester and analysed using the Systematic Analysis of Language Transcripts SALT computer software program (Miller and Chapman, 1983). Finally, the normative data collection from Arabic controls in Chapter 5 was conducted by two student collaborators from King Saud University, Riyadh, Saudi Arabia; and transcribed and analysed by the primary investigator.

CHAPTER 2 Background and Literature Review

2.1 Aphasia definition

Aphasia is an acquired language disorder that affects an individual's ability to communicate using language in at least one language modality (i.e., receptive language, expressive language, reading, and/or writing). It follows a neurological insult to the brain caused by any neuropathology such as stroke, brain tumour, arteriovenous malformation, brain infection, traumatic brain injury, or other neuronal changes such as primary progressive aphasia.

2.2 Aphasia demographics

English speaking countries

Although the causes of aphasia are many, the most common cause is stroke. It is estimated that more than 100,000 people suffer from a stroke every year in the United Kingdom (UK), and one-third of the survivors experience aphasia (Stroke Association, 2021). Similarly, aphasia occurs in one-third of stroke survivors in the United States (USA) every year, around 225,000 (National Aphasia Association, 2021).

Arabic speaking countries

Arabic is the 5th most spoken language globally, with 313 million speakers worldwide (Simons and Fennig, 2018). It is the first language in all Middle East countries except Israel, Turkey, and Iran. A systematic review by El-Hajj et al. (2016) estimated the incidence of stroke in the Middle East to range between 22.7 and 250 per 100,000 population per year, and the prevalence ranges between 508 and 777 per 100,000 population. The kingdom of Saudi Arabia is the largest country in the Middle East, covering around 4/5 of the Arabian Peninsula (2,150,000 km^2) with a population of 34,218,169 people (the Saudi general authority for statistics, 2019). A study by Alqahtani and colleagues (2020) estimated the annual incidence of stroke for people residing in Saudi Arabia through a systematic review. The pooled annual incidence of stroke for people residing in Saudi Arabia, from the five studies that met the inclusion criteria, indicated that there are 29 stroke cases for every 100,000 people annually.

2.3 Aphasia types

Numerous patterns of language impairment symptoms could occur depending on the location and extent of injury in the brain. The most common cluster of patterns can be categorised into global aphasia, mixed transcortical aphasia, Broca's aphasia, transcortical motor aphasia, Wernicke's aphasia, transcortical sensory aphasia, conduction aphasia, and anomic aphasia (Goodglass and Kaplan, 1983; Kertesz, 2007). These aphasia subtypes are distinguished through the application of a decision tree (yes or no) to the presence of a) speech fluency, b) auditory comprehension and c) repetition (see Figure 3). To further simplify these classifications, the eight subtypes were divided into two equal groups: fluent and non-fluent (Goodglass and Kaplan, 1972). The fluent group comprises four subtypes: Wernicke's aphasia, transcortical sensory aphasia, conduction aphasia, and anomic aphasia. They share the characteristic of a neuro-typical rate, rhythm, and intonation of speech with little or no motor speech difficulties. The phrase length in their speech output is noted to match or exceed that of neuro-typical controls (Albert et al., 2013; Buckingham Jr and Kertesz, 1974). However, their speech content often lacks meaning (Edwards, 2005). On the other hand, the non-fluent group comprises four subtypes: global aphasia, mixed transcortical aphasia, Broca's aphasia, and transcortical motor aphasia. The label 'non-fluent' refers to the shared characteristic of effortful speech production and misarticulation (i.e., motor speech deficits), limited vocabulary, restricted grammar, short-phrases and spared auditory comprehension (Goodglass and Kaplan, 1972). Nevertheless, it is acknowledged that variations of the well-defined subtypes of aphasia do exist. They are identified as 'undifferentiated aphasia' or 'unclassified aphasia' and combine features of different subtypes of aphasia but do not exactly fit into one recognized type (National Aphasia Association, 2017).

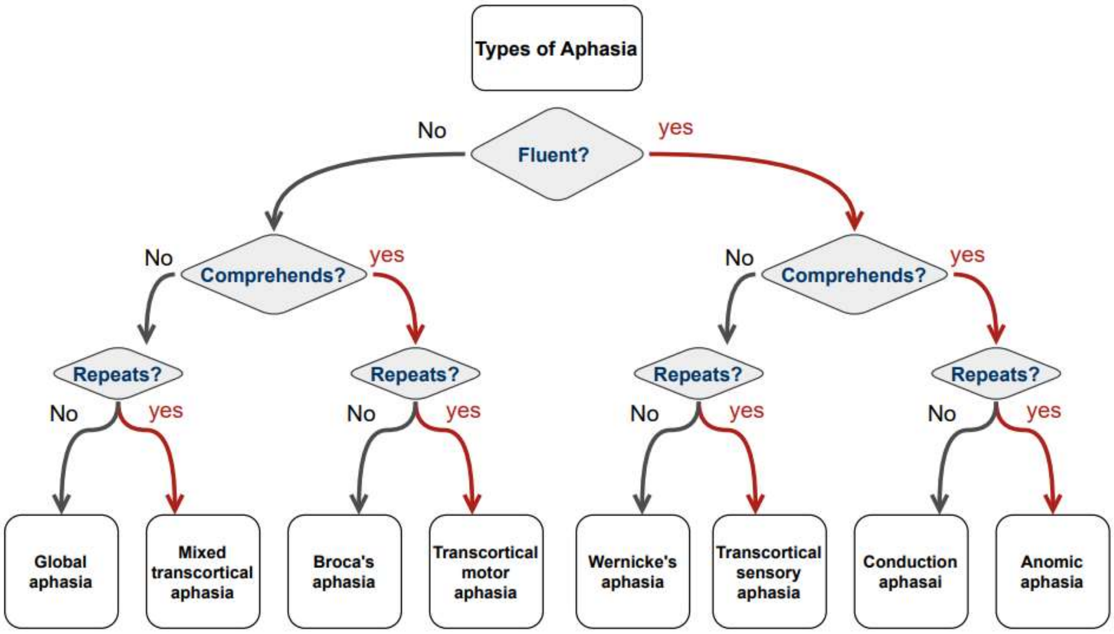


Figure 3 Aphasia classifications (Goodglass and Kaplan, 1983)

2.4 Consequences of aphasia

Aphasia can limit the individual's ability to express thoughts and feelings and understand others' verbal and written communication. It is invariably highly disabling for people's social and professional lives and affects their independence in daily life activities. The adverse consequences extend to include the individual's psychological well-being and quality of life (Cruice et al., 2003). Therefore, it is not a surprise that aphasia's negative impact on patients' quality of life can be much higher than those of other devastating diseases such as cancer and Alzheimer's disease (Lam and Wodchis, 2010).

2.5 Models of sentence production

The symptoms affecting verbal expression are among the most obvious and socially limiting of features of aphasia. The literature examining expressive symptoms has focused substantially on lexical processing, as a core function of language processing more broadly. However, from a functional and treatment perspective, sentence production is a skill which is equally as striking in aphasic behaviour. Although, the sentence production literature, relating to models of sentence processing and theoretical approaches to treatment, is not as extensive as that for lexical processing, it is now well established and provides a useful starting point for understanding sentence production deficits in aphasia (Marshall, 2015; Thompson et al., 2015).

2.5.1 Serial models

Several language models have been developed to explain the processes involved in sentence production. One of the earliest models was proposed by Garrett, based on speech error analysis in healthy subjects (Garrett, 1980; Garrett, 1982; Garrett, 1988). It proposed that levels of processing in sentence production are serial and discrete. Within this framework, each level has to be completed before the next one begins; Accordingly, no interaction between levels is involved. Garrett's model consists of three stages: conceptualization, formulation, and articulation (Garrett, 1975; Garrett, 1976). Sentence production begins with conceptualization, also referred to as the message level, which includes the generation of a conceptual proposition of the message the speaker intends to express. The formulation stage then follows in which syntactic planning take place. Here, the model distinguishes two main processes: the functional level and the positional level. At the functional level, major lexical concepts and their functional relations are specified (i.e., thematic role assignment).

However, words are not explicitly ordered and phonological forms are not specified until the next level of processing, the positional level. Moreover, at the positional level the syntactic frame of the planned sentence is generated, thematic roles are mapped into the syntactic roles, and function words are inserted. Garrett argued that since semantic content of words are selected at the functional level whereas function words are selected at the positional level, there is a structural dissociation between lexical retrieval and syntactic planning processes at the sentence formulation stage. Then, at the last level of the formulation stage, the sound level, the phonological representation of function words and other grammatical elements are specified. Finally, at the articulation stage, the phonological representations get translated into a series of phonological features that drive the articulatory apparatus.

Another recognized serial model of sentence production is Levelt's model. Similar to Garrett's model, it distinguishes three main areas of processing that underlie language production: conceptualization, formulation, and articulation (Levelt, 1989). At the conceptualization level, also called the message level, a preverbal message is generated following the intention to communicate a certain message. However, Levelt's model differs from Garrett's model in that the functional and positional levels are encompassed in one process, labelled the grammatical encoding, instead of being divided into serial levels. Additionally, Levelt's model further expanded the properties of the positional level by including processes such as constituents assembly and inflection processes (Bock and Levelt, 1994b).

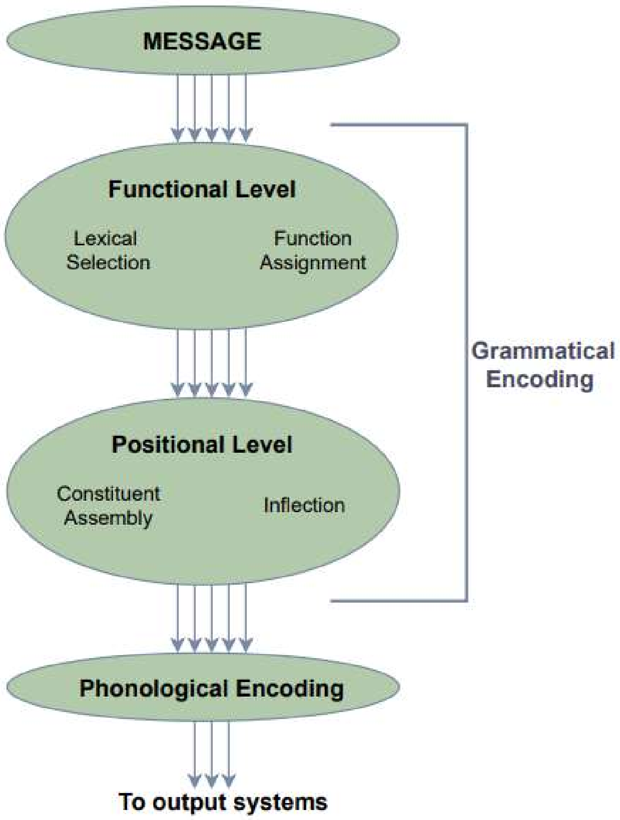


Figure 4 Language production model (Bock and Levelt, 1994b)

2.5.2 Interactive Activation Models

Chang et al. (2006) developed an interactive model of sentence production based on insights derived from research on language acquisition in children (syntax) as well as sentence priming studies (Bock, 1986). In Chang et al. (2006) model, the process of sentence production is composed of two relatively independent components, which are the meaning system and a sequencing system, and a dual path interactive activation occurs between them. The model proposes that " the meaning system learns the association between the message and the thematic roles of different lexical concepts" while " the sequencing system learns to sequence these roles and insert morphosyntactic markers" (Thompson et al., 2015, p.332). Additionally, it highlights a correlation between the process of learning language rules (i.e., implicit learning of syntactic rules) and the use of these learned language rules in sentence production.

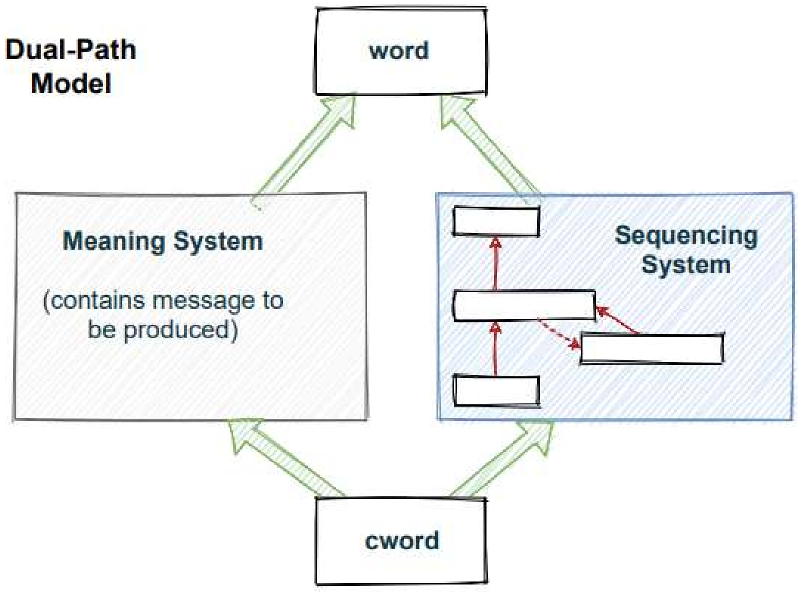


Figure 5 Dual-path model of sentence production (Chang et al., 2006)

Overall, sentence production models agree on the following essential component processes: generating the message, grammatical encoding (involves *content processes* and *structure processes*) and morphophonological encoding (Thompson et al., 2015).The conceptual preparation of generating the message includes both macroplanning, influenced by the communicative context and intent, and microplanning, which yields a set of semantic and pragmatic specification of the intended message that serve as the starting point for sentence generation. Then, grammatical encoding includes *content processes* and *structure processes*. The *content processes* involve "lexicalization" which translates the semantic representation of a word (i.e., accessing mental lexicon) into its phonological form. Levelt (1989) proposed, that all words in the mental lexicon are represented by *lemmas* which encodes semantic and syntactic information, whereas the phonological form of words are represented by *lexemes* (i.e., contains abstract minimal information about the word's structure and stress pattern). According to the lexicalist syntactic theory, within *structure processes* sentence planning is lexically driven since lexical entries contain grammatical information in addition to its semantic content. Lastly, within the morphophonological encoding stage, morphological encoding encompasses a range of operations that select and retrieve grammatical elements that are necessary to achieve a complete well-formed sentence and conveys the intended message (e.g., grammatical elements include functional morphemes (such as negation), subject-verb agreement, determiners for noun phrases, etc.). On the other hand, phonological encoding involves the retrieval of the *lexeme* followed by post-lexical planning that specifies the phonetic description which is then translated into a motor plan that drives the articulatory apparatus (Thompson et al., 2015).

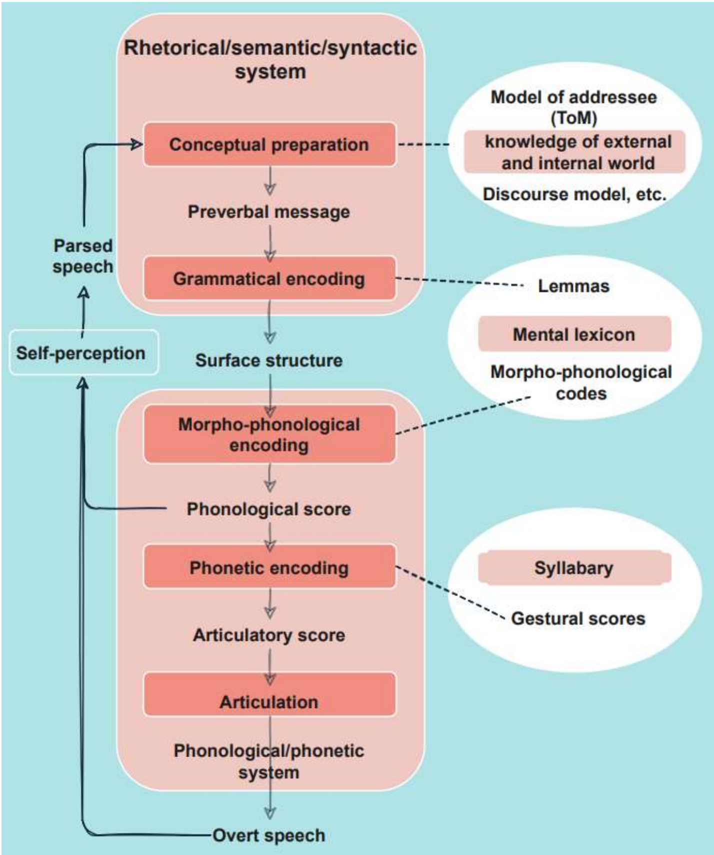


Figure 6: Levelt (1999, p.87) “a blueprint of the speaker”

In psycholinguistic theories, spoken language processing is often divided into two linguistic processes: comprehension and production. In both processes, the verb plays a central role.

Sentences are made of nouns and verbs in addition to other lexical items. Although the ‘lemma’ of nouns and verbs share similar types of encoded information that specify meaning, gender and distribution (e.g., the requirement of a determiner), additional integral syntactic information is associated with the verb lemma only. The verb has information stored that determines the subcategorization frame (e.g., what phrasal structure can follow the verb), thematic and argument structure of a sentence. Moreover, the verb expresses aspect of time (e.g., past tense, future tense, etc.) and requires inflection (e.g., person and number) that corresponds to the subject in a given sentence. Those features make the lemma of a verb more complex than the lemma of a noun (Bastiaanse et al., 2002).

Sentence formulation requires the retrieval of verbs and the associated syntactic information. In most instances, the sentence cannot be grammatically correct without a verb. Also, failure to access the syntactic information encoded within the verb could result in a wide range of sentence errors (e.g., thematic role assignment errors) which leads to communication breakdown.

Furthermore, the ability to process single verbs is of great relevance to sentence comprehension. In neuro-typical adult listeners, sentence comprehension involves parsing the verb and the syntactic items in a given sentence. For example, *The girl pushes the boy*, *push* is the verb, the *girl* is the subject and the *boy* is the direct object. In the passive sentence *the boy is pushed by the girl,* it is noted that the verb is in past participle form and the sentence includes a by-phrase to which the *girl* belongs (the *boy* is now the subject). Next, the verb representation in the mental lexicon is searched for associated grammatical and thematic roles. In the sentence *The girl pushes the boy*, the verb *push* assigns the role of agent to the subject and the role of theme to the object. Finally, to fully understand the sentence, the mapping process is required, in order to link grammatical roles to thematic roles. In the same example sentence mentioned above, the subject *girl* is the agent (the doer of the action), and the *boy* fulfils the theme role (receiving the action) (Bastiaanse et al., 2002).

2.6 Aphasia Speech and Language Therapy

A growing body of evidence in the literature supports speech and language therapy's efficacy in ameliorating symptoms of aphasia, specifically by accelerating language recovery and improving communication effectiveness. A Cochrane review by Brady and colleagues that included 57 randomized controlled trials and a total of 3002 participants concluded that speech and language therapy yielded significant clinical and statistical gains in "functional communication" in expressive language as well as in reading and writing (Brady et al., 2016).

Sentence Therapy

Engaging in conversation is the most common activity of everyday communication in people with aphasia and healthy older people equally (Davidson et al., 2003). Initiating and participating in a conversation requires the individual to formulate sentences such as questions, declarative, exclamative, or imperative sentences that convey their intended message. Given the complexity of the processes involved in sentence production, there are a range of ways sentence production skills can break-down in aphasia. This is reflected in the several approaches to sentence production therapy which can now be found in the clinical aphasiology literature.

Many aphasia sentence processing therapies have been established based on theories of normal production models, such as Garrett's (1988) and Bock & Levelt's (1994a) models. These models provided the foundation for research to identify the locus of a deficit in sentence production in PWA and design an intervention plan accordingly. Several studies have examined PWA's poor performance on sentence comprehension and production tasks, especially with reversible sentences, and supplemented their investigation with tests that revealed reserved grammaticality judgment ability and intact sensitivity to syntactic structures (Linebarger, 1990; Linebarger, 1987; Linebarger et al., 1983; Saffran and Schwartz, 1988). They concluded that sentence impairment does not stem from a loss in central sentence processing competence but rather from a failure to integrate thematic roles (e.g., agent, patient) to grammatical roles (e.g., subject, object) of a parsed sentence constituents (Saffran and Schwartz, 1988; Schwartz et al., 1987). This impairment has been termed the 'mapping deficit hypothesis'. In normal sentence production models, such as Garrett's (1988) and Bock & Levelt's (1994a), this disruption occurs at a level described as "grammatical encoding". It involves two stages: the "functional level" in which semantic and thematic elements of words are specified (i.e., abstract form) and the predicate-argument structure is generated, and the "positional level" in which the surface form of the sentence is constructed and the phonological features of the words retrieved. Many studies in the aphasia literature suggest that a breakdown in grammatical encoding is the source of PWA sentence deficits. It could affect one or both stages of the process (i.e., functional or positional levels), or the transfer of information between them (Berndt, 2001; Mitchum et al., 2000).

Individuals with mapping impairment exhibit difficulty understanding complex sentence structures such as reversible sentences. Moreover, their speech is characterized by marked verb omissions, and when verbs are produced, they are usually with reduced verb argument structure (Marshall, 2015). It can occur in both fluent and non-fluent aphasia, although they are typically associated with non-fluent and especially agrammatic presentations (Marshall et al., 1997; Mitchum et al., 1995).

Saffran, Schwartz and colleagues (Saffran and Schwartz, 1988; Schwartz et al., 1987) proposed further explanation of the potential sources of the mapping deficit. In the "lexical" variant, encoded thematic information in the verb becomes inaccessible due to aphasia. Examples that support this notion are derived from the poor performance on sentence comprehension tasks when the stimulus involves simple canonical sentences. On the other hand, the "procedural" variant is suggested to stem from faulty assignment (i.e., mapping) procedures that link sentence structure to meaning (i.e., grammatical roles and thematic roles). This variant's characteristics involve a clear discrepancy in performance on comprehension and production tasks between canonical and non-canonical sentences, with better performance on canonical sentences. Also, it is associated with relatively intact verb knowledge. Accordingly, the implication is to target strengthening the connection between sentence structure (i.e., subject, verb, object) and meaning (i.e., agent, theme, etc.,), specifically "who did what to whom", rather than focus on training the production of various sentence forms.

The introduction of the mapping deficit hypothesis by Schwartz and colleagues (1987) posed an implication for the development of several approaches that aimed to test the outcomes of strengthening the mapping operations explained in the paper. Accordingly, they were categorized as "mapping therapy". The concept of mapping therapy was shaped into various approaches, but the common aim remains the same; "they all emphasize where event participants are mapped in relation to the verb" (Marshall, 2015, p.387).

The following section will summarise our analysis of the current approaches to sentence therapy in aphasia. It guided our design of a novel therapy program, presented in Chapter 4, by selecting from the literature a combination of therapy approaches that, based on theoretical underpinnings, were anticipated to produce maximum benefit in a time-efficient way and foster independence in home-practice.

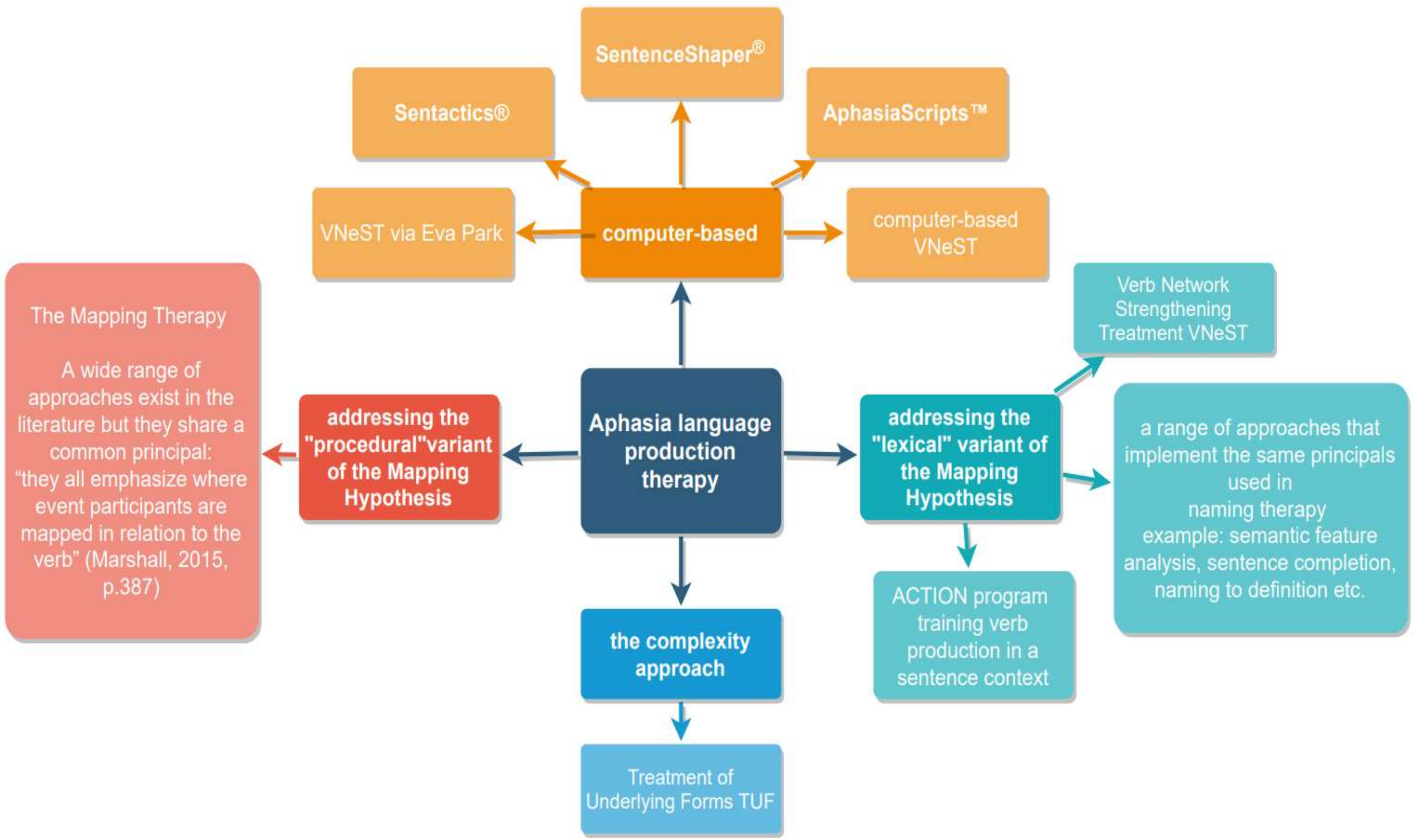


Figure 7 Overview of the therapy approaches that were examined for the design of our novel sentence therapy approach presented in Chapter 4

2.6.1 Therapies designed to remediate thematic impairment

a. Therapies targeting the "lexical" variant of the mapping deficit

In the aphasia literature, observations have supported the notion that sentence impairments are, in some cases, related to a lexical deficit (i.e., verb access impairment) (Marshall, 2015; Saffran et al., 1980). For example, various studies have documented strong association between sentence production impairments and verb impairments in PWA (Berndt et al., 1997; Luzzatti et al., 2002; Mätzig et al., 2009). Also, other literature has noted the facilitative effects of providing the verb as a lexical cue in the production of sentences and alleviating deficits (Marshall, 2015; Marshall et al., 1998). This notion is perhaps not surprising, given the established agreement that essential syntactic information is stored in the verb's lexical representation. According to Biran and Fisher (2015), it includes "information about the verb's predicate-argument structure (PAS, number of arguments), the thematic role of each argument (agent, theme, etc.), and the verb's subcategorization frames (types of syntactic phrases that can complement the verb)", which restricts the type of arguments and phrases that can attach to a given verb (Biran and Fisher, 2015, p.30). As expected, inaccessibility to this information would result in faulty sentence processing.

From this point, several therapy approaches have emerged to restore verb processing, and therefore, sentence construction skills. Two recent reviews, one by Faroqi-Shah and Baker (2017), and one by Marshall (2017), reviewed a representative sample of the existing approaches to remediate verb deficit in the aphasia literature. Some of these approaches implemented the same principals used in naming therapy (i.e., noun retrieval) such as semantic feature analysis, confrontational naming with semantic and/or phonemic cues, word to picture matching, naming to definition, sentence completion, and repetition (Carragher et al., 2013; Edwards and Tucker, 2006; McCann and Doleman, 2011; Raymer and Ellsworth, 2002; Wambaugh et al., 2002). Other studies eliminated the naming task and trained verb production in a sentence context such as ACTION therapy program (Links et al., 2010). On the other hand, a study by Webster and colleagues (2005) kept the naming and semantic tasks and extended the therapy program to include sentence production tasks of one-, two-, and three- argument structures. Other studies presented argument structure-based intervention where the participant is prompted to name a verb, generate various arguments in response to

Wh-questions presented by a clinician, and to produce a sentence that includes the target verb and the generated arguments.

Verb Network Strengthening Treatment VNeST is an example of this approach. Edmonds, Nadeau, & Kiran (2009) defined the VNeST as a "semantic treatment that aims to improve lexical retrieval of content words in sentence context by promoting systematic retrieval of verbs (e.g., *measure*) and their thematic roles—i.e., agent (doer of the action, e.g., *carpenter, chef*) and patient (receiver of the action, e.g*., lumber, sugar*)" (Edmonds et al., 2009, p.402). As indicated by its name, the Verb Network Strengthening Treatment VNeST is designed to strengthen the semantic network (i.e., represented as neural networks) associated with a given verb and its related nouns through repeated activation and use (Edmonds et al., 2009). This approach is known as Hebbian learning, proposed by Hebb (1949), is commonly summarized as "what fires together wires together" (Shatz, 1992, p.64). The premise is supported by evidence from studies on neuro-typical controls, which showed that the activation of a target verb leads to neural co-activation (i.e., bidirectional priming) with its associated thematic roles (e.g., agent, patient, instrument) as well as co-activation with other semantically related verbs (Edmonds and Mizrahi, 2011; Ferretti et al., 2001; McRae et al., 2005). As a result, the priming of a target verb reached a widespread network of semantically related verbs and nouns, as illustrated in Figure 8. Therefore, VNeST aims to create this broad activation of semantic networks in people with aphasia to facilitate the retrieval of both trained and untrained verbs and their associated thematic roles.

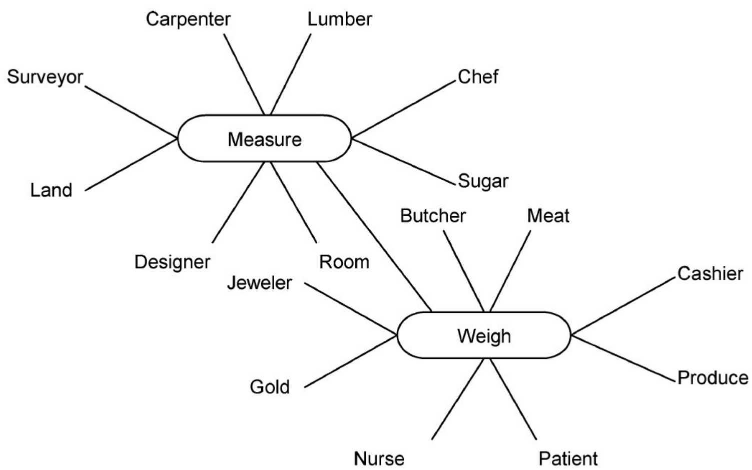


Figure 8 Schematic of the relationship between the verb–thematic network (Edmonds et al., 2009, p.405)

As illustrated in Figure 9 below, VNeST's procedure involves six steps. Step one focuses on generating multiple scenarios around a trained verb. The task would start by setting the scene, as shown in step one. The clinician would present the target verb, a card with the target word written on it (e.g., drive) and place it in between two cards that have the words "who" and "what" written on them. The clinician would then point to each card and read the words aloud. Next, the participant is asked to generate a sentence by answering, "who can/might (drive) something/someone?". Three different sentences/scenarios using three different agents and patients is required for this task. Also, throughout the task, cues are provided as needed. These range from minimal semantic/contextual cues (e.g. occupational, locations, etc.) to maximal cues (e.g., providing the correct agent or patient along with the target verb and asking the patient to choose the card that has the matching agent/patient from a field of four). Step two, the participant is asked to read the three generated sentences (i.e., agent-verb-patient triad) aloud. Target responses would look like: 'Dad_drive_boat', 'Chauffeur_drive_limousine', and 'paramedic_drive_ambulance'. Step three targets sentence expansion. The participant is asked to select one of the three generated sentences. Then to expand it by answering wh-questions such as: where, when, and why (i.e., written on cards similar to "who" and "what"). When a sentence is achieved, the participant is asked to read it aloud. Step four involves asking the participants to make semantic judgments about sentences (i.e., with the same target verb) read by the clinician, with appropriate and inappropriate thematic role pairing. The participant would indicate their answer by saying "yes" when the sentence made sense and "no" when it did not. Step five targets independent retrieval by answering " what verb have you been working on?" presented by the clinician. Finally, step six requires the participant to recall and repeat the sentences generated in step one independently without cues from the clinician or visual aids (i.e., cards). When all of the six steps are achieved, the next target verb is introduced, and the same steps are followed until 10 verbs are trained. After that, the same 10 verbs will be cycled through again. The aim is to train all 10 verbs within one week (Edmonds, 2014).

VNeST studies have shown promising results. A review by Edmonds (2016) summarised the findings of those studies, which encompassed 19 native English participants with a range of aphasia types and severity. It revealed that the intervention was successful in improving noun naming in 86% of the sample (measured by either OANB or BNT and verb naming subtest from the NAVS) and action naming in 58% of the sample (Edmonds, 2016, p.127). The generalization to semantically related but untreated stimuli has also been noted (e.g., the ability, in a sentence construction task, to produce an agent (nurse) when the given untrained verb (weigh) is semantically related to the trained verb (measure) and patient (baby) in response to a novel picture stimuli (Edmonds et al., 2009)). Also, 75% of the sample improved on untrained sentence production skills (measured by a constrained sentence production task), and 59% of the sample showed significant gains in sentence production in discourse (measured by complete utterances /CUs).

Moreover, a number of studies reported therapy-induced improvement in verb retrieval following the implementation of the mapping approach in treating sentence processing (Byng, 1988; Marshall et al., 1993; Schwartz et al., 1994), which further supports the “lexical” variant hypothesis of the mapping deficit.

Treatment Steps for Verb Network Strengthening Treatment (VNeST).

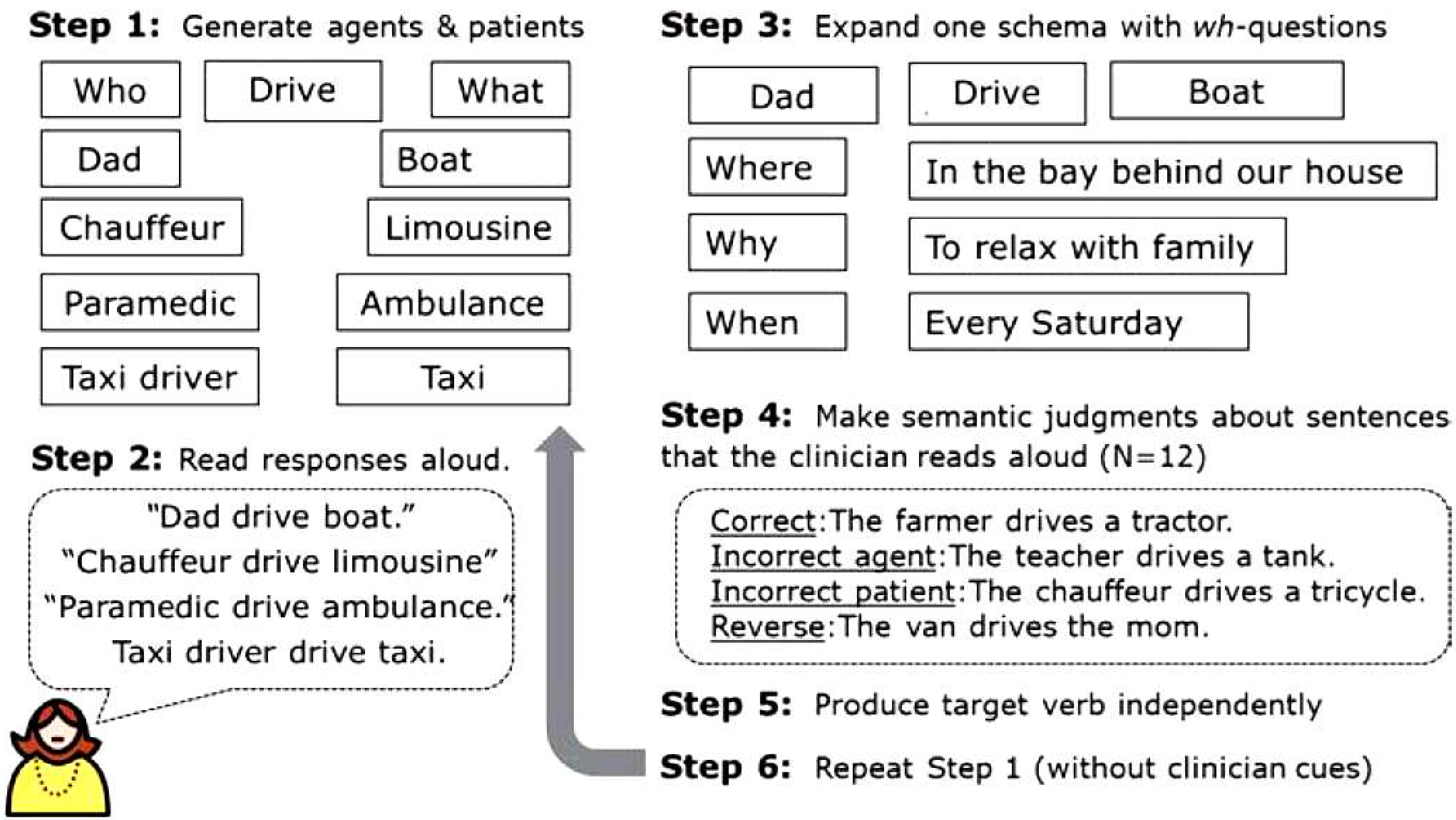


Figure 9 VNeST treatment steps (Edmonds, 2014, p.81)

b. Therapies targeting the "procedural" variant of the mapping deficit

Many forms of the "procedural" variant of the mapping therapy have been introduced in the literature. Two features are common in all these forms: they aim to repair/strengthen the connection between positional- and functional-level representations, and they do not address lexical difficulties explicitly. They can broadly be divided into comprehension based approaches with no production tasks (Berndt and Mitchum, 1997; Jones, 1986; Mitchum et al., 1995; Schwartz et al., 1994) and approaches that placed greater emphasis on sentence production (Rochon et al., 2005). Also, the second type can be further differentiated into studies that implement the traditional hierarchy of difficulty in training sentence structures from simple to complex and those that implement the complexity approach in focusing on training complex sentence structures.

One form of mapping therapy involves a picture description task and focuses on word-order implications on a sentence's meaning. Therefore, it incorporates methods to support the production of the target word order. In this approach, the participant is required to select one of two picture scenes that illustrate the same action (e.g., writes). Then describe it by constructing a sentence using the presented sentence fragments and with the aid of a given scaffold of the sentence frame. The sentence fragments are written on four separate colour-coded cards to highlight nouns and verbs. The cards are presented in a field of four. The target subject, verb, and object the cards include a distractor subject (e.g., the robber, writes, the letter, the monk). Next, the participant is asked to change the assembled sentence to describe the other picture. Then feedback is given to highlight the connection between word order and meaning (Byng et al., 1994; Nickels et al., 1991). The next step involves verbal sentence production to describe the picture scenes. In addition to the visual scaffold of the sentence frame, cues are provided by the clinician to aid in word order and retrieval of target verb as needed. At this stage, feedback is also provided to highlight the implication of word order. Following the same steps, the construction and production of different sentence types are trained. Lastly, the participant is asked to generate sentences to describe personal events. Several studies implemented a modified version of the therapy framework described above, and few included non-canonical sentences in their training (Carragher et al., 2015; Harris et al., 2012; Marshall, 2017; Rochon et al., 2005).

Although the mapping therapy literature encompasses a wide variety of procedures, sample sizes, and outcome measures, they generally reported favourable results in improving sentence processing (Byng, 1988; Byng et al., 1994; Dorze et al., 1991; Fink et al., 1998; Jones, 1986; Marshall, 1995; Marshall et al., 1997; Mitchum et al., 1997a; Mitchum et al., 2000; Nickels et al., 1991; Rochon et al., 2005; Schwartz et al., 1994). Additionally, some of these studies also reported improvements in everyday language, including increased structural complexity of sentences in spoken narratives (Byng, 1988; Nickels et al., 1991; Rochon et al., 2005; Schwartz et al., 1994). Nevertheless, those improvements were specific to the trained modality (comprehension or production) (Berndt and Mitchum, 1997; Mitchum et al., 1997a; Rochon et al., 2005). Also, results are inconclusive regarding the generalisation of treatment effects to untreated sentence structures (Rochon et al., 2005; Schwartz et al., 1994).

c. Treatment of underlying forms TUF (Thompson and Shapiro, 2005)

Treatment of underlying forms (TUF) represents an extension of mapping therapy. It was developed to remediate sentence production and comprehension deficits in people with agrammatic aphasia. The complexity account of treatment efficacy CATE hypothesis distinguishes TUF (Ballard and Thompson, 1999; Jacobs and Thompson, 2000; Thompson et al., 1997a; Thompson, 2008; Thompson and Shapiro, 1995; Thompson et al., 2003; Thompson et al., 1993; Thompson et al., 1996) from other sentence therapy approaches. It proposes that training more complex structures, which contradicts the traditional method of the gradual increase of complexity in training, yields better outcomes. It improves trained items and induces generalization of treatment effects to untrained but linguistically related sentences. In comparison, training simple sentences first did not show generalisation to complex structures. Moreover, the TUF studies reported a generalisation of therapy gains to connected speech on both lexical and structural measures (Ballard and Thompson, 1999).

In a TUF task, participants are presented with models of the target sentence structure read aloud to them by the examiner (for example, Wh-question sentence structure: Who is the soldier pushing in the street?). They are then presented with canonically sequenced sentence elements written in separate cards, with a relevant Wh-word and a question mark card. They are then asked to modify the order of cards to construct the target sentence structure (i.e., Wh-question). Therapy begins with identifying the verb and its thematic roles. Then, "the object NP is replaced by the relevant Wh-card (who) and the question mark added to the end of the sentence" (Marshall, 2015), and the participant is asked to produce it. It is then followed by a demonstration of the subject/auxiliary verb inversion. Lastly, the Wh-word card is moved to the beginning of the sentence, and the participant is asked to produce the modified sentence (i.e., in its Wh-question structure). The task requires repetition of the steps, and with these subsequent repetitions, the therapist's assistance is faded until the participant's independence is reached.

A series of studies investigated the outcomes of the above-described TUF (Ballard and Thompson, 1999; Jacobs and Thompson, 2000; Thompson et al., 1997a; Thompson et al., 1997b; Thompson and Shapiro, 1995; Thompson et al., 2003; Thompson et al., 1993; Thompson et al., 1996). Favourable outcomes in both production and comprehension of sentences were documented, supporting this approach's effectiveness (Thompson and Shapiro, 2005). This approach requires the participant to have an adequate ability to read words and sentences and understand Wh-questions, which is usually compromised in people with aphasia.

TUF was first introduced as a paper-based therapy approach, and currently, a computerised version of the therapy is being developed, which is called "Sentactics"(Thompson et al., 2010). The implementation and outcomes of this interactive computer program will be revisited shortly.

2.6.2 Computer-based therapy

a. SentenceShaper® (Linebarger and Romania, 2000)

SentenceShaper® is a computer program constructed as a "processing prosthesis" to support PWA in building sentences and narratives (Linebarger and Romania, 2000). The body of aphasia literature links sentence production impairments to either one or both of the following causes. First, the deficit arises from the loss of central representations required to perform syntactic analysis; it is termed the ''syntactic deficit'' hypothesis (Schwartz et al., 1985). Second, a limitation in cognitive resources is the cause of sentence impairments in PWA. One type of these cognitive resources is related to a shortage in the performance of the working memory during production; it is termed the "temporal window" hypothesis (Kolk, 1995; Kolk and Van Grunsven, 1985). It suggests that language production and comprehension impairments in people with agrammatism stem from a slowed activation and/or a rapid decay of linguistic information (i.e., sentence constituents). This shortage prevents the speaker from retaining the retrieved sentence elements simultaneously in working memory long enough to construct them in a sentence.

The majority of computerized aphasia therapy programs are model-driven and were developed to remediate an underlying deficit in linguistic representation/processing. However, SentenceShaper® is distinctive in its approach to extend the 'temporal window'. It targets the cognitive processes involved in language production rather than explicitly aiming to restore central linguistic representations. Accordingly, it has been classified as a processing prosthesis (i.e., communication aid). This role of the prosthesis is to widen the temporal window by allowing the retrieved sentence elements to be retained as long as required to integrate them into a sentence or discourse. Therefore, it alleviates the time pressure from the process of sentence production. Also, with the record, playback, and edit features of the program, the participant can listen to their productions and self-correct when needed. Thus, it promotes the construction of more complex, elaborated, and well-formed sentences and discourse. These characteristics were built into the program based on the observation (at least in non-fluent participants) that comprehension skills are often stronger than production skills. Therefore, stronger sentence comprehension skills will be the resource used to identify faulty productions and guide self-correction. Accordingly, this approach aids residual grammatical competence in PWA. The SentenceShaper literature examined the treatment effects of using SentenceShaper to practice narrative construction without explicit impairment-based training. However, a single case study by McCall and colleagues (2009) examined the efficacy of including a clinician-lead structure-specific training to produce multi-clause sentences. The results were promising, with improvements noted in connected speech on linguistic measures.

To summarise the SentenceShaper® procedure, it is a ''user-initiated interactive program that allows the speaker to record his/her own voice and then replay the recorded utterance for immediate self-judgement or to trigger new material through sentence completion (cloze effect). The user then combines the recorded segments into sentences and narratives in an incremental fashion, replaying these larger creations for self-judgement or in an attempt to activate new material." (McCall et al., 2009, p.440)

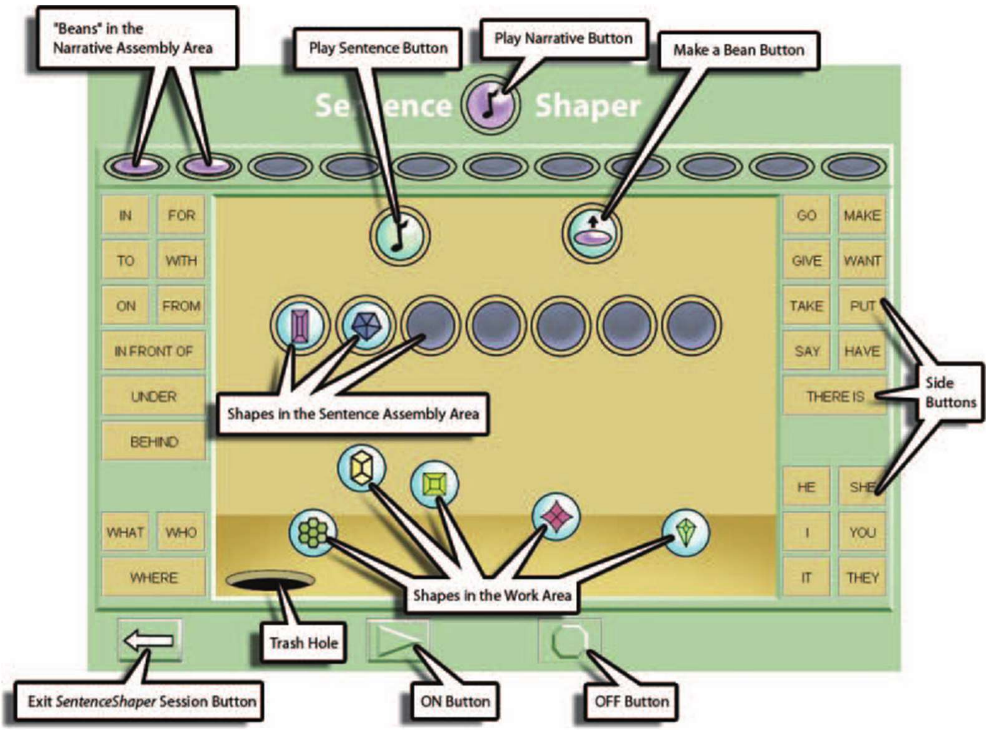


Figure 10 SentenceShapter (True et al., 2010)

A series of studies have investigated the outcomes of SentenceShaper®. The first study that showed the treatment effects of SentenceShaper by Linebarger et al. (2001) included two participants with non-fluent aphasia. The study design alternated the use of SentencShaper with another therapy approach. The findings showed that independent home practice without clinician-directed intervention was sufficient to induce remarkable improvements in the participants spontaneous (unaided) narratives in story retelling task, after 15 hours of practice. The noted improvement included an increase in the Mean Length of Utterances MLU. The intervention method in this study involved introductory in-person sessions to teach the participant how to operate the program effectively, and weekly lab visits for monitoring and encouragement. The assigned home practice included creating narratives to retell movies or television shows, express opinions, and describe life events.

The treatment effects of a narrative-based SentenceShaper treatment protocol were then examined by Linebarger et al. (2007)and McCall et al. (2009) and revealed treatment gains in speech rate, narrative informativeness, and structural gains including sentence length, proportion of words in sentences, and grammaticality of sentences. A subsequent case study by Albright and Purves (2008), showed evidence of increased morphosyntactic complexity in the narrative production of a participant with non-fluent aphasia following the intervention.

A handheld portable version of this computer software, "SentenceShaper To Go™" was introduced to support PWA communication in a range of daily life encounters (e.g., going to the doctor) (Linebarger et al., 2008). It was designed to serve as an AAC device where the participant can build narratives for an anticipated communicative encounter and play them in context when needed. Most recently, in 2019, an iOS app version of SentenceShaper® had been released. It was designed for application as a treatment tool as well as a communication aid. Also, this version is compatible with languages that apply right-to-left text direction, including Arabic.

b. Sentactics® (Thompson et al., 2010)

Sentactics® is an interactive computer program that delivers the Treatment of Underlying Forms TUF therapy protocol by a virtual therapist. It is developed by Thompson and colleagues (2010). The TUF protocol is based on the complexity account of treatment efficacy (CATE) (Thompson et al., 2003) that proposes that training complex structures, instead of the traditional increasing hierarchy of complexity, extends generalisation of therapy gains beyond the target sentence structure to include untrained linguistically related but less complex sentence structures.

A computer version of TUF, Sentactics®, was introduced as a solution to overcome the limitations of the clinician-delivered paper-based TUF therapy. For example, the original protocol requires at least 20 in-person therapy sessions to train a specific sentence structure. The affordability of this high number of sessions may restrict many PWA. Moreover, delivering this method requires the treating SLT/P to have extensive background knowledge in linguistics and a considerable amount of training.

Therefore, computer-based therapy was proposed to improve the availability and cost-effectiveness of this method. Thompson and colleagues (2010) compared the outcomes of both approaches, TUF paper-based and Sentactics®, and concluded that the outcomes were similar regardless of delivery mode. Improvements were noted in the production and comprehension of complex sentences and generalisation to untrained linguistically related, less complex structures. Therapy gains were also captured by standardized tests such as the Northwestern Assessment of Verbs and Sentences NAVS (Cho-Reyes and Thompson, 2012) and a narrative construction task (e.g., Cinderella narrative). The author also pointed to the enhanced quality of discourse post-intervention, which may indicate generalisation to discourse production. However, the trend was also noted in controls, and further research is required to support this finding.

Although both TUF and the mapping therapy implement the same concept of training verbs in a sentence context to strengthen the connection between thematic roles and grammatical roles, the reported outcomes of TUF seems to exceed those of the mapping therapy. The TUF/ Sentactics® therapy gains were not limited to trained sentence structures (e.g., sentences with object relative clauses) but extended to include untrained linguistically related less complex sentence structures (e.g., object clefts and object wh-questions). The reported improvements were noted in both modalities comprehension and production. In comparison, the mapping therapy studies reported limited generalisation of therapy gains beyond the sentence structures targeted in therapy (Schwartz et al., 1994). In both forms of intervention, TUF and mapping therapy, cross-level generalisation of therapy gains to discourse has been noted (Webster et al., 2015).

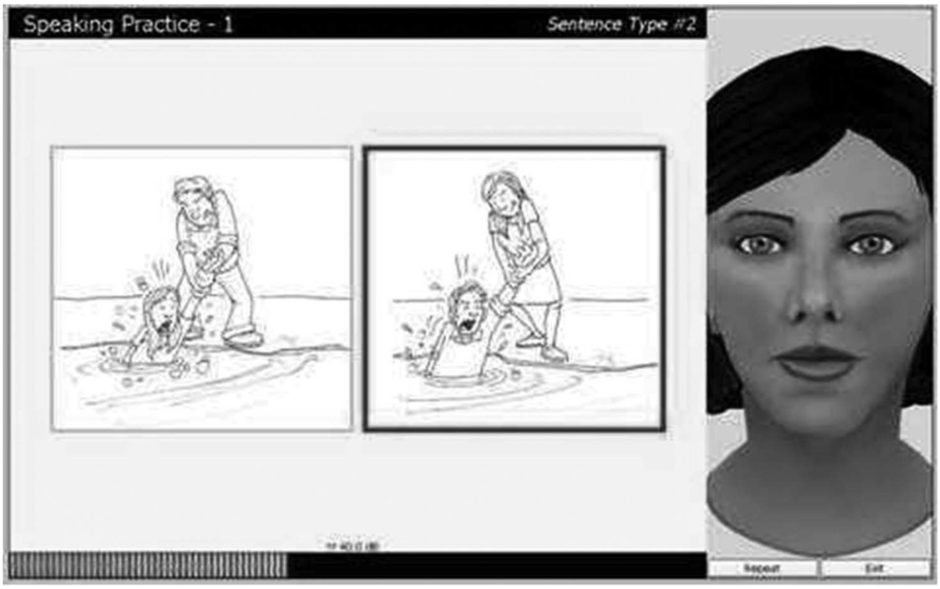


Figure 11 Sentactics® sentence production test screen. (Thompson et al., 2010, p.1250)

c. AphasiaScripts™ (Cherney et al., 2008a)

AphasiaScripts™ is also computer software that utilizes a virtual therapist to deliver therapy tasks. The foundation of this approach is the repeated practice of pre-recorded personalized scripts. The sentences are generated in a conversation context or monologue with the virtual therapist's aid (to model sentence production and as a conversational partner). One of the approach's strengths is the inclusion of the participant's personal goals in therapy planning, as it enhances motivation and compliance. It also promotes functional communication gains and social independence by selecting functional goals to practice, such as ordering food in a restaurant or recounting personal information. Another interesting feature is the adjustable level of support for sentence production. The cues can be faded as needed until independence is reached. This approach also fosters self-directed practise and autonomy.

Cherney et al. (2008a) reported therapy-induced improvements; however, they were limited to the practised materials, and no generalisation to untrained scripts was observed. Nevertheless, two of the three participants in this study showed an increase in their overall language performance, as measure by the Western Aphasia Battery WAB (Kertesz, 2007). Lee, Kaye, and Cherney (2009) reported that therapy gains were positively correlated with the amount of therapy the participants received.

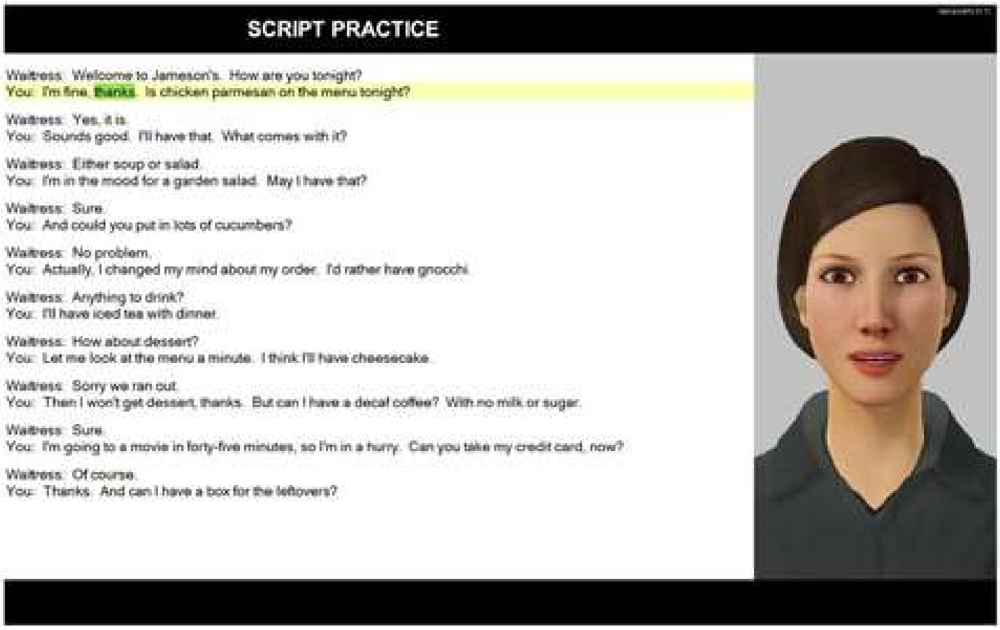


Figure 12 Example script from AphasiaScripts™ (Cherney et al., 2014, p.S348)

d. VNeST program delivered virtually via EVA Park

EVA Park is a multi-user online virtual communication environment (Marshall et al., 2016). The aim of creating this platform was to enable PWA to communicate with each other and with support workers. It offers a simulated world of various settings (e.g., a health centre, a hair salon, a restaurant), in which participants communicate in real-time mainly through speech. The first EVA Park trials, which included 20 participants, showed noticeable improvements in functional communication through language stimulation. However, the delivery of explicit aphasia therapy was not tested at that time.

Marshall et al. (2018) reported a case study that examined the outcomes of delivering a modified version of VNeST therapy program, originally designed for face-to-face therapy, remotely via EVA Park. No measurable therapy-induced gains were noted in verb retrieval for both confrontational naming and sentence production contexts (i.e., in response to sentence elicitation picture task). Also, no improvements were noted in connected speech or functional communication gains as measured by the CADL-2 (Holland et al., 1999).

Figure 13 Modified Semantic Feature Analysis (SFA) delivered in EVA Park: participant (left) and treating therapist (right) (Marshall et al. 2018, Page 1060)

e. Computer-delivered VNeST

Two computer-based versions of VNeST, which implemented an adapted protocol of VNeST, have been introduced. Furnas and Edmonds (2014) administered VNeST- C through telerehabilitation over the internet. The study extended the original therapy protocol to include training in both spoken and typed modalities. Two participants were enrolled in the pilot study and received 24 sessions over 8 weeks (three times per week, two hours per session). In all sessions, treatment was guided by the clinician throughout. Both participants responded well to this type of remote therapy and showed improvements in lexical retrieval at the sentence level of both trained and untrained verbs. However, the generalisation of therapy gains to discourse was limited.

2.7 Designing a therapy plan

The literature on speech and language therapy for aphasia has highlighted two main pathways for designing intervention plans for PWA. The first pathway includes selecting either an impairment-focused or participation-focused therapy approaches (De Bleser and Papathanasiou, 2003). The second pathway involves combining and bundling therapies (i.e., labelled a care bundle), which constitutes either a simple additive model or a theoretically-based integrative approach (Hinckley, 2017).

The *care bundle* category was defined by Hinckley (2017) as " a group of evidence-based therapies that addresses a particular symptom or disorder" (Hinckley, 2017, p.341). The author explained that although each therapy approach in a given *care bundle* is supported by evidence, a combined delivery of these approaches in a bundle for maximum efficacy is usually not backed up by evidence in the literature. Nevertheless, it is a common practice by many SLT/P s, based on clinical reason and expertise, without recognizing it as a *care bundle*. According to the author's review, only few *care bundles*, designed for aphasia speech and language therapy, were directly studied and reported in the literature.

Two main methods were identified in selecting and combining different therapy approaches to create a care bundle, the additive model and the integrative model. The main trend noticed in the additive model was that it combined an impairment-focused approach with a participation-focused approach. The goal is to facilitate the generalization of a trained skill to a target activity by adding a participation-focused treatment sequentially within an impairment-focused session or across these sessions. On the other hand, the integrative model involves the interlace of different therapy approaches together, "specifically in the service of one or more client-selected activities" (Hinckley, 2017, p. 342). The identified activity, in which the client's performance requires improvement, serves as a target and an outcome measure for therapy (e.g. ordering food from a menu in a restaurant). Accordingly, the clinician identifies the set of skills required to accomplish the desired activity by decomposing the tasks involved in a top-down manner. Therefore, the assessment of speech and language impairment and therapy goal selection is limited to skills that are directly responsible for achieving those tasks/activities. For example, the skill of writing, even if deficient, is not included in the assessment and therapy plan when the desired activity is to order food from a restaurant.

Hinckley (2017) reported a promising observation of the efficacy of a range of studies (Herbert et al., 2003; Robson et al., 1998; Springer et al., 1991) that implemented the additive model in combining an impairment-focused approach with a participation-focused approach. The findings indicated that this approach was successful in achieving the generalisation of targeted language skills. On the other hand, the integrative model has been supported by studies such as Frederiksen and White (1989) that demonstrated the effectiveness of integrating whole-task with part-task training. The authors concluded that this approach achieved superior outcomes in terms of shorter skill acquisition time and larger skill transfer to untrained contexts.

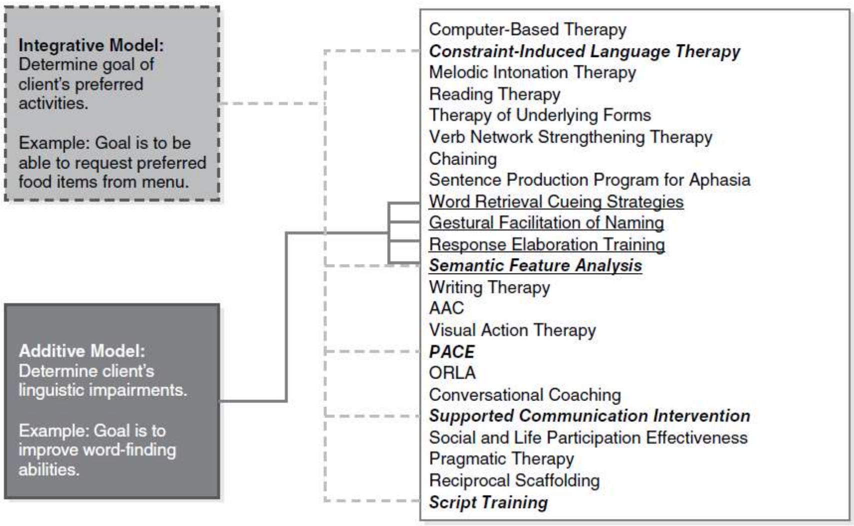


FIGURE 10.3 Therapy selection differs when activity goals or impairments are used as the primary characteristic for selection. a) In this example of an integrative model, five therapies are selected because they incorporate the same interactive goal as the targeted activity (dashed lines); b) In this example of an additive model, four therapies are selected because they intend to address the client's impairments (solid lines). One therapy (SFA) fits both examples.

Figure 14: Examples of the integrative model and additive model (Hinckley, 2017, p.348)

2.8 Interim Summary

The intervention approaches described in this chapter reflect the most clinically well specified and most fully evaluated approaches reported in the aphasia sentence therapy literature. With this foundation, Evidence-Based Practice in aphasia therapy is available to Speech-Language pathologists to support their clinical decisions. Unfortunately, resources are not equally available for Arabic speakers. Therapy studies that examined an impairment-based intervention's outcomes are limited to one case study (Al-Shdifat et al., 2018) to the best of our knowledge. Since English and Arabic are fundamentally different languages, it is essential to examine the efficacy of implementing therapy approaches borrowed from the English aphasia literature. Also, to explore novel interventions that may better address issues specific to Arabic. The following section will illustrate the distinctive characteristics of Arabic.

2.9 Arabic

Arabic is a Semitic language, a branch of the Afro-Asiatic languages. It originated from the central and northern parts of the Arabian Peninsula and spread to become the standard language of many countries in the Middle East and North Africa. Today it is considered the largest living member of the Semitic languages with at least 313 million native speakers (Simons and Fennig, 2018). It is also the fifth most spoken language after Chinese, English, Hindi, and Spanish (Khwaileh et al., 2017).

Figure 15 Arabic speaking countries

2.9.1 Dialects in Arabic

In Arabic, there are two variations of the same speaker's language, a situation called “diglossia” (Ferguson, 1959). The first variant, colloquial Arabic, is acquired at home and thus considered the native language of Arabic speakers. The other variant, the Modern Standard Arabic (MSA), is learned at school, and its uses are limited to formal settings such as written texts and formal oral communication (e.g., news, academic discourse, and official speeches). Therefore, an Arabic speaker's fluency in the colloquial dialect does not imply an equal command of MSA. The evolution of colloquial Arabic relative to the origin of classic Arabic is historically indefinable (Aoun et al., 2009; Procházka, 2006).

A broad range of differences between Arabic dialects exists. Arabic sociolinguists classified regional dialects into four main groups: Gulfian, Egyptian, Levantine and North African (Zughoul, 2007). Individual groups may contain more than one sub-dialect. The distinction of colloquial Arabic from MSA includes variations in phonology, morphology, syntax, and semantics (Laks, 2013).

2.9.2 The Arabic root system

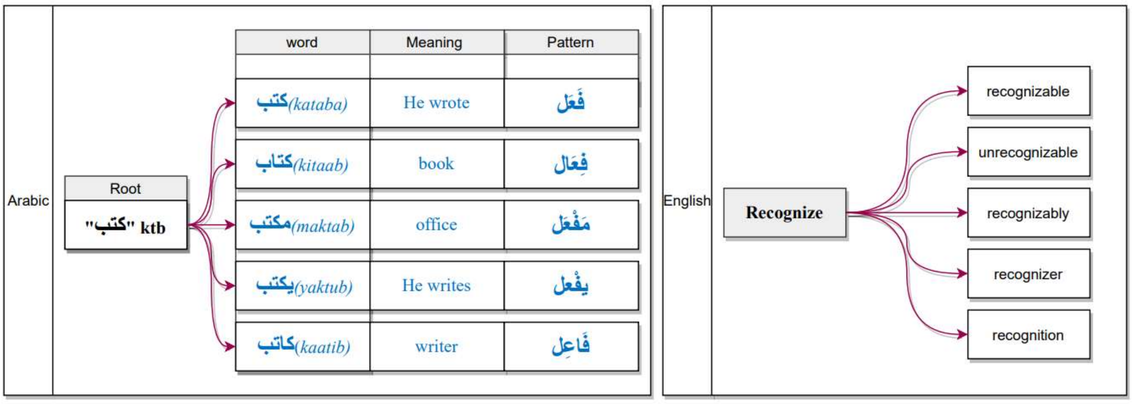
Arabic, and Semitic languages in general, have a unique underlying form-meaning relationship. A given word can have a network of words related in meaning (same semantic core) (Alhawary, 2009), and the lexical differentiation is carried out through a system of derivation.

It is important to understand the root and pattern system in Arabic in order to understand its grammar. The word-formation process in Arabic depends on a systematic derivation of words from their lexical roots, which typically consist of three consonants called radicals. This process of derivation can change a word's form class (e.g. creating an adjective شمالي from noun شمال) and subclass (creating a transitive verb أسقط from an intransitive base سَقَط). According to Ryding (2014), "Arabic is a "synthetic" or "fusional" type of language (like Latin) wherein several morphemes may "fuse" together in one word, indicating various kinds of grammatical and lexical information."(Ryding, 2014, p.45).

A four consonant roots are also common and considered an extension of triconsonant roots. On the other hand, the biconsonants, composed of two radicals, are infrequent type of word roots, and found in the most elementary vocabulary such as water 'maʔ' and father 'ʔab' (Procházka, 2006). It is not always easy to identify morphemes and morpheme boundaries when analysing Arabic word structure; therefore, background knowledge of Arabic roots variation and morphological processes is usually required.

In Arabic, lexical roots (radicals) appear in the same sequence in all derived word patterns. However, a prominent feature that distinguishes it from English, is that the derivation process can involve changes to the core of the word (root letters) in addition to the beginning and end of the word. It includes the addition of letters or combination of letters in between the root letters (كاتب). In contrast, in English, the changes are limited to the beginning and the end of the word.

Table 2.1 Example of words that are derived from the root "كتب ktb" in Arabic, and an equivalent example in English



Words are typically morphologically complex and consisting of more than one morpheme. For example, the word 'مطعم' (/*maṭʕam*/) 'restaurant' is composed of the lexical root morpheme 'طعم' (ṭʕ*m*) 'taste' and the grammatical pattern morpheme that specifies a place 'مـ'. A counterpart example in English would be 'conservatory' and 'conserve'. Nevertheless, words in Arabic do occur occasionally in the form of solid stems such as the noun *yad* 'hand'.

The following are examples of morphological analysis of Arabic words in Modern Standard Arabic MSA:

a) Noun decomposition: the word مطعمٌ *maṭʕamun* 'restaurant' is composed of four inflectional morphemes (number, gender, case, and definiteness), and two derivational morphemes (root and pattern). In general, masculine and singular morphological properties of words in Arabic are usually subtle and unmarked (Ryding, 2014).

o Four overt morphemes can be identified as the following:

- Lexical *root* طعم ṭ-ʕ-*m* 'taste'
- Lexical *pattern* indicating noun-of-place 'مـ' . This pattern 'مفعل' includes grammatical components that specify morphological features such as *number* (singular) and *gender* (masculine).
- Nominative *case-marker* suffix '-u'
- *Indefinite* marker suffix '-n'

A distinctive feature of colloquial Arabic is the omission of nominative *case-marker* and *Indefinite* marker suffix –u and –n in *ma*tʕ*am**<u>un</u>***.

a) Verb decomposition: the word يدرسُ *yadrusu* 'he is studying' is composed of six inflectional morphemes (root, stem pattern, tense, person, gender, mood, number, and voice) and two derivational morphemes (root and pattern).
   - Eight overt morphemes can be identified as the following:
   - Lexical *root* **درس** *d-r-s* 'study'
   - Verb *stem pattern* '**يفعل**' '**يـ**', which also indicate present *tense* and *voice* marks
   - Inflectional prefix ya- indicating *person* and *gender*
   - Inflectional suffix -u indicating *mood* and *number* markings for indicative singular

   Basic semantic information of words in Arabic is contained in their roots.

2.9.3 Sentences in Arabic

In Arabic, words are classified into: verbs "actions", names "nouns", and particles "letters". The "names" classification encompasses "adjectives" as well due to their comparable morphology. In Arabic, Adjectives can take the same position as a noun in a noun phrase NP, acting as a head of NP and bearing the markers of definiteness and indefiniteness. Accordingly, they are syntactically less distinguishable from nouns compared to their equivalents in English. (Ingham, 1994).

Sentences in Arabic can be distinguished into mainly two types: "verbal" sentences and "nominal" sentences. Classical grammarians categorizes a sentence that starts with a verb as a "verbal sentence" (i.e., verb-subject-object VSO sentence pattern), and a sentence that starts with a noun as a "Nominal sentence" (i.e., includes both SVO and OVS sentence patterns). Some modern Arab grammarians argue that a classification that takes into account the presence or absence of a verb as of more structural relevance. Thus, they define a "Verbal sentence" as a sentence containing a verb, and a "Nominal sentence" as a sentence with no verb (Ingham, 1994). Nevertheless, the categorization of sentences without a verb remains a topic of debate.

a. Word order of sentences in Arabic

The most common word order in Arabic is VSO; however, different word orders are also possible, influenced by discourse context (Procházka, 2006; Ryding, 2014). The typical word order VSO of classic Arabic is not consistent in regional dialects. For example, in Egyptian colloquial Arabic, the preferred word order is SVO (Gamal-Eldin, 1967). However, Ingham (1994) investigation of Southern Iraq, Najdi, Qatar dialects (i.e., also known as Gulfian dialect) revealed a relatively equal frequency of VSO and SVO word orders.

b. Subject-verb agreement in Arabic

In Arabic, a full agreement is required when a subject precedes the verb in a sentence (e.g., SV); the verb needs to reflect both gender and number of its subject. On the other hand, only a partial agreement is required when a verb precedes a subject in the dual or plural form (e.g., VS). In this case, the verb needs only to reflect the gender of its subject. These agreement constraints are exclusive to human subjects since it is the only form of plural subjects reflected in Arabic agreement morphology. In a non-human dual or plural subject, the subject-verb agreement will always be a feminine singular regardless of word order (Ryding, 2014).

c. The complexity of Arabic sentences

The complexity of parsing Arabic sentences stems from: " 1) the length of the sentence and the complex Arabic syntax, 2) the omission of diacritics (vowels) in written Arabic "altashkiil", 3) the free word order nature of Arabic sentence, and 4) the presence of an elliptic personal pronoun "alDamiir almustatir."'(Othman et al., 2003, p.1).

2.9.4 Arabic Orthography

In Arabic, the Modern standard Arabic version is used in written materials (e.g., newspapers, books), and colloquial Arabic has no written form. A distinguishing feature of Arabic orthography is that it is written from right to left. The written forms of verbs and most nouns in Arabic are composed of consonantal roots. Specific combinations of affixes and vowels added to the consonantal roots create various words (Berman and Bolozky, 1978).

The orthographic symbol set consists of 28 letters, eight diacritical marks, and the Hamza (ء). Diacritics can be divided to:

1. three short vowels, m ◌َ , ◌ُ , and ◌ِ .
2. three nunation diacritics ◌ً , ◌ٌ , and ◌ٍ (short vowels pronounced followed by an /n/), which is considered a mark of nominal indefiniteness in Standard Arabic.
3. a consonant doubling diacritic (◌ّ), which indicates a repetition of the marked consonant.
4. a diacritic for marking when there is no diacritic (◌ْ).

The Hamza can serve as either a letter (e.g., ء) or a diacritic when attached to a letter (e.g., ئ، ؤ) (Habash et al., 2007).

The presence of diacritics in a written text specifies the phonological form of orthographical sequences. It also makes the orthography-phonology relation fully transparent (Ibrahim et al., 2002). Diacritics appear almost exclusively in religious texts and school textbooks. Therefore, reading common Arabic texts requires the reader to have background knowledge of diacritics to accurately pronounce the written word. A similar counterpart of this process in English, is when a reader relies on the context to decide if the written word should be pronounced read /ri:d/ (present tense) or read /rɛd/ (past tense) (Habash et al., 2007).

There are two additional features that add to the complexity of Arabic orthography. First, letters are represented in different shapes depending on their location in a word. It is possible for one phoneme to be represented by different graphemes. For example, the letter "هـ" can appear in the following graphemes: هـ initial, ـهـ medial, and ـه or ه final word positions. Twenty-two of the 28 letters in Arabic have four shapes, and six have two shapes. Moreover, dots in Arabic orthography are used extensively and sometimes it is the only feature that distinguishes two identical structures. For example, (ث) and (ت) are identical in structure but the number of dots changes the corresponding phoneme from /θ/ to /t/.

2.9.5 Arabic Phonology

Arabic consists of 28 consonant phonemes (i.e., /ʔ/أ - /b/ب - /t/ت - /θ/ث - /dʒ/ج - /ħ/ح - /x/خ- /d/د - /ð/ذ - /r/ر - /z/ز - /s/س - /ʃ/ش - /ş/ص - /đ/ض - /ŧ/ط - /ðˤ/ظ - /ʕ/ع - /ɣ/غ - /f/ف - /q/ق - /k/ك - /l/ل - /m/م - /n/ن - /h/ه - /w/و - /j/ي) and three vowels /a, u, i/. A distinguished variance between Arabic and English phonemes is that the sounds /p/ and /v/ are not included in Arabic phonemes.

A prominent feature of both consonant and vowel systems in Arabic is the phonemic short-long contrast. It allows conveying different meanings of the same word depending on the locus of insertion. For example, [dʒamal] means ‘camel’, and when a phonemic contrast is applied to the consonant /m/ in [dʒamːal], the meaning is changed to ‘beautify’. Also, when a phonemic contrast is applied to the vowels /u/ in [muraːfiquna] ‘our companion’ the meaning is changed to [muraːfiquːnaː]‘our companions’(Procházka, 2006). Other grammarians use different terminologies to describe these phonemic processes, such as vowel lengthening (e.g., rasala ‘he sent’ raasala ‘he corresponded’) and germination or doubling (e.g., darasa ‘ he studied’ darrasa ‘ he taught’)(Ryding, 2014).

Words in Arabic never begin with a vowel. Also, vowels within a word must be separated with at least one consonant (e.g., CVCV), but not more than two (e.g., CVCCV) (Procházka, 2006).

2.10 Aphasia characteristics in Arabic

Few studies have investigated the patterns of sentence production deficits in Arabic speakers with aphasia of the Jordanian and Palestinian dialects.

Albustanji et al. (2013) investigated the characteristics of agrammatism in Jordanian Arabic speaking individuals. In their study, participants exhibited more difficulty in producing Wh-questions than Yes/No questions. They explained that changing word order and producing an additional morpheme required to formulate a Wh-question is likely the reason for this discrepancy. Additionally, production of grammatical morphemes in a sentence completion task showed that most errors were related to tense inflection, followed by agreement inflections and negation particles. The authors concluded that the language deficit patterns found in the Jordanian Arabic speaking individuals with aphasia were in agreement with patterns identified in Semitic languages such as Hebrew and Palestinian Arabic (Friedmann, 2006).

A study by Friedmann (2006) examined the characteristics of language production deficits in a group of participants with Broca's aphasia. The sample, composed of 2 Arabic speaking individuals of the Palestinian dialect and 16 Hebrew speaking participants, showed impaired production of Wh-questions and intact production of yes/no questions. A cross-linguistic comparison showed variability with the performance of Dutch, English, and German speaking individuals with the same site of lesion, as their production of yes/no questions was impaired. The author referred to the syntactic tree-pruning hypothesis to explain this pattern in performance. Accordingly, the discrepancy is attributed to a failure in accessing high nodes of the syntactic tree (Friedmann and Grodzinsky, 1997) as Wh-question production requires access to higher syntactic nodes in Arabic and Hebrew while yes/no questions do not.

Khwaileh et al. (2015) studied morpho-syntactic processing of regular and irregular plurals in Jordanian Arabic speaking individuals with agrammatism. The identified error types included: gender inflection errors in regular plural constructions and over-regularization in irregular plural construction. The authors concluded that the forms of irregular plurals in Arabic are stored; in contrast, the forms or regular plurals are derived, which aligns with findings in English.

2.11 Chapter Summary & Aims of this PhD study

Sentence impairment in aphasia is common. The characteristics could range from reduction in the use of complex sentences and embedded clauses to more severe forms such as the total omission of morphological markers (i.e., telegraphic speech). The aphasia sentence processing literature has mainly focused on studying the extreme forms of the deficit, such as agrammatism in Broca's aphasia. An updated fine-grained description of the variation in sentence production skills across a sample of participants representing a range of aphasia subtypes, including fluent aphasia, and severity levels is currently lacking. Therefore, the first study in this thesis, presented in the next chapter, aims to provide a detailed, graded snap-shot of sentence construction skills across a heterogeneous sample of 29 participants with aphasia (PWA).

The overarching aim of the thesis is to develop a novel theory-driven computer-based language intervention that is effective in remediating sentence deficits in aphasia and compatible with two languages, English and Arabic. Therefore, in addition to the knowledge we will gain from examining sentence deficits across aphasia types and severities in the first empirical study, the outcomes will inform our candidacy selection for the second empirical study that includes therapy implementation to a subset of the screened group of participants.

The sentence production therapy literature is characterized by small studies evaluating novel therapies with a given subset of patients (e.g., participants with agrammatic aphasia). A systematic application of one hybrid sentence production therapy across a large, varied and heterogeneous range of participants could be a more informative way to analyse the factors that influenced therapy gains and to evaluate optimal candidacy for future therapy selection.

In the third empirical study we endeavoured to translate and adapt this novel therapy program, which was designed for English PWA, into Arabic and examine its outcomes with Arabic PWA. The findings will contribute to the Arabic aphasia literature that is currently lacking, and support evidence-based practice in speech-language therapy delivered to this population.

In this chapter, 2, we have reviewed the theories of normal language production models that served subsequent hypotheses, such as the mapping deficit hypothesis, in identifying the locus of sentence production breakdown in aphasia in and the therapy approaches that emerged from it to address the deficits (Schwartz et al., 1987). We also reviewed the computer-based approaches to aphasia therapy that were either available to the consumer or still in development, to explore their usability to our study. In doing so, we have identified three main approaches to sentence therapy that were originally designed for unique purposes. We hypothesized that a combination of the three will complement each other and will achieve our goal of creating an affordable and accessible intervention that yields maximum therapy gains with time-efficiency and fosters autonomy in home practice. This is also an approach that is suitable for translation and adaption into Arabic, which is fundamentally different than English as explained in this chapter.

Our novel theory-driven computer-based aphasia sentence therapy approach is composed of: the mapping therapy approach (Schwartz et al., 1994), Verb Network Strengthening Therapy VNeST (Edmonds et al., 2009), and SentenceShaper® (Linebarger and Romania, 2000). The mapping therapy approach repair and strengthen the mapping operations between grammatical roles (subject, object, etc.) and thematic roles (agent, theme, etc.) which is required to determine “who did what to whom” in a sentence. On the other hand, the VNeST approach serves to “strengthen the semantic representation of the verb and its relationship to various thematic roles” and “potentially strengthen specification of PAS and basic sentence syntax” (Edmonds, 2016, p.125). In doing so, VNeST is expected to remediate sentence production deficits based on the integration of semantics, syntax, and thematic role assignment its tasks offer. The semantic network activation it promotes yields a wide spread improvements in lexical retrieval that extends beyond targeted words, which means the time-efficiency of therapy is optimised. However, VNeST does not target the grammaticality of sentences in training, in which the mapping therapy covers. Lastly, SentenceShaper® provides processing support, an artificial way of widening the temporal window to overcome cognitive processing limitations that is hypothesized to be the source of sentence production breakdown in aphasia (Kolk and Van Grunsven, 1985). Also, unlike the mapping therapy and the VNeST, SentenceShaper® is a computer software that can serve as a language neutral medium (compatible with English and Arabic) and is flexible to customisation to deliver a range of aphasia therapy approaches. It is also user friendly and can effectively foster independence in home-practice, which can be a cost-effective means for increasing the intensity of language training in.

CHAPTER 3 Screening sentence production skills across a cohort of people with aphasia

3.1 Introduction

Sentence production is a core skill within human language and interaction. Sentences represent the linguistic realization of generating propositions within which we describe, comment on, and verbally engage with the world. A grammatically correct syntactic structure is crucial to delivering an intended message successfully. Failure to do so will eventually lead to communication breakdowns. The performance of participant PB, a case study presented in Marshall et al. (1997), illustrates how a missing element in the syntactic structure can hinder verbal communication effectiveness. In a picture description task, PB explained, "*One woman and a cat is buying the man and paying the money the till"*(Marshall et al., 1997, p.859). Without looking at the picture stimulus, it is hard to understand what PB was trying to say (a woman sells a cat to a man). A closer assessment of the production reveals adequate retrieval of target words, both nouns and verbs, but a deficit in assigning thematic roles that compromised the sentence's meaning.

In general, there seems to be evidence to indicate that to convey an intended message, the produced sentence has to be contentful, informative, complete, and grammatically correct. Here we define 'contentful' as sentences constructed with content vocabulary (e.g. 'the man is reading the book') instead of pro-forms (e.g., 'He is doing that') which is less informative. This typical production of sentences requires several cognitive processes to be engaged. It includes retrieving target lexical items, linking thematic (semantic) roles to grammatical (syntactic) roles, sequencing words in the right order, inserting function words and appropriate grammatical morphemes.

In sentences, lexical items carry semantic and syntactic information (i.e., 'lemma' level (Levelt, 1989)) that is vital for sentence processing. The verb in particular plays a central role in both sentence comprehension and production, as the verb encodes information that determines the thematic structure, argument structure and the subcategorization frame of a sentence as well as expresses aspect of time (see Chapter 2 for more information). In typical sentence production, when the verb is retrieved from the mental lexicon, the associated thematic and grammatical information becomes available, which details the number and type of arguments (thematic roles) associated with that particular verb. For example, in the sentence '*the mother is washing the dishes*' the verb '*washing*' is a transitive verb with an agent and a theme role that are fulfilled by the subject and a direct object in an active sentence (i.e., canonical sentence). The next process in sentence production involves mapping the thematic roles into the grammatical roles determined by the syntactic structure. In the above example, the agent is mapped onto the subject '*mother*' and the theme is mapped onto the object '*dishes*'. The final phase involves sequencing the words in the right order and inserting the grammatical morphemes (e.g., verb-subject agreement and verb tense). Accordingly, syntax involves how to produce a certain message and not the content of it. For example, "syntax may dictate that a noun must be selected at a certain point, but not which specific noun" (Parisi, 2013, p.202).

People with Aphasia often experience disruption at one or more of these typical processes, which leads to sentence production deficits (Thompson et al., 2015). The impairment could be at the lexical level (i.e., single word) originating from a semantic deficit affecting the access to meaning or due to inability to access the phonological form of the target word. The observed characteristics of sentence deficits in PWA with semantic difficulties include limited use of sentence structures and over reliance on single phrases, especially noun phrases (Berndt et al., 1997) PWA with phonological deficits have been noted to be able to produce a sentence frame even when having difficulties retrieving the target words (Berndt et al., 1997; Fink et al., 1993). Moreover, sentence deficits can also be linked to the inability to access argument structure information encoded within the verb, which is required to determine the type and number of thematic roles that can be associated with a particular verb. Also, it has been noted that PWA show greater verb retrieval difficulties as the complexity of the verb argument structure increases (e.g., verbs with three argument structures impose greater difficulty than verbs with one- and two-argument structures), including tasks that involve single word retrieval (Kim and Thompson, 2000). In connected speech, people with agrammatic aphasia showed preference for verbs with simple verb argument structures (Thompson and Shapiro, 1995).

Verbs, based on their semantic weight, can be classified as "light" (i.e., defined as high frequency semantically empty words that generally convey nonspecific meanings such as go, put and have), or "heavy" (i.e., communicates more semantic information such as chop, drive and measure) (Berndt et al., 1997). It has been frequently suggested in the literature that people with aphasia rely on producing light verbs to overcome verb retrieval difficulties (Berndt et al., 1997; Gordon, 2008; Jespersen, 1992).

The literature on sentence production in aphasia has described deficits into two main categories: patterns associated with the fluent aphasia subgroup and those associated with the non-fluent aphasia subgroup. In fluent aphasia, studies on sentence and discourse production have concluded that one of its most distinctive features is "empty speech" (Nicholas et al., 1985), indicating its uninformative nature compared to neurotypical controls. It was noted that participants tend to resolve to circumlocution and to skip propositions when experiencing word-finding difficulty. Therefore, it was suggested that the compensatory strategies they have adopted to overcome their lexical retrieval difficulties were the reason for their incoherence rather than the lack of awareness of the information they aim to convey. Overall, language production of fluent aphasia as a subgroup is characterized by incomplete information, information gaps, irrelevant information, and disruption to the progression of the message being expressed (Christiansen, 1995). In terms of linguistic features, Nicholas et al. (1985) reported 14 categories of the "empty speech" features. It includes empty phrases such as common idioms (e.g., "something like that"), indefinite terms (e.g., "thing", "stuff" etc.,), deictic terms (e.g., "this", "that" etc.,), pronouns without antecedents, etc. Gleason et al. (1980) reported that the speech of people with Wernicke's aphasia was distinguished by "the use of many words, concatenated sentences, deixis, and the use of verbs rather than nouns"(Gleason et al., 1980, p.381). The authors concluded that although people with fluent aphasia (Wernicke's aphasia was the focus of their study) were not notably impaired syntactically, their sentence construction skills were characterized by reduced complexity of syntactic structure and well-formedness compared to neuro-typical participants; similar findings have been reported in the literature (Bird and Franklin, 1996; Butterworth and Howard, 1987; Edwards, 1995).

On the other hand, the language production of participants with non-fluent aphasia contains a higher proportion of content words compared to fluent aphasia and many deletions or errors in function words, morphology, and thematic role assignment (Armstrong, 2000; Gleason et al., 1980; Goodglass and Kaplan, 1983; Schwartz et al., 1987). Such errors invariably limit both the number of words produced and the informativeness of the sentences produced, and in severe cases, the effectiveness of verbal communication is diminished. Also, Gleason et al. (1980) reported that in narrative samples of participants with Broca's aphasia, the ratio of nouns was higher than verbs.

In summary, a large number of PWA have difficulties constructing grammatical sentences and retrieving the target words to express their intended message. In people with non-fluent aphasia, sentence production deficits are commonly attributed to grammatical complexity, whereas in people with fluent aphasia lexical deficits seems to be the main problem, although grammatical deficits are not excluded (Bastiaanse et al., 2002).

The two most prominent theories relating to differences in language production skills between fluent and non-fluent aphasia were the avoidance theory of (Heeschen, 1985) and the adaptation theory (Kolk and Heeschen, 1990; Kolk et al., 1985). Heeschen (1985) explained that sentence deficits exist in both subgroups, but the observed variance in production skills is due to the distinction of their adopted strategies to overcome the deficit. Goodglass (1993) hypothesized that speech output fluency influences the strategy adopted by each group. Individuals with agrammatism (observed mainly in non-fluent aphasia) tend to avoid grammatical challenges and resolve to use elliptical speech. On the other hand, individuals with paragrammatism (observed in fluent aphasia) do not avoid the difficulty but instead force their way through it and make paraphasic errors as a result. The adaptation theory proposed by Kolk and Van Grunsven (1985) proposed that agrammatic speech is a volitional strategy adopted by PWA to avoid making grammatical errors by producing only very simple sentence structures. The authors explained that agrammatic speech is influenced by the slow access or rapid decay of syntactic representations (Haarmann and Kolk, 1991).

The above described distinction reported in the literature between the two subgroups was mainly driven by somewhat extreme cases of aphasia in terms of fluency such as Broca's aphasia (non-fluent) or Wernicke's aphasia (fluent) with less attention given to other classifications within each subgroup (e.g., conduction, transcortical motor, etc.,). Also, given the variability between individual participants the comparison is not without limitations. Additionally, the feasibility of distinguishing two main subgroups is further challenged by the overlap in some characteristics of the language deficit (e.g., omission and substitution errors in morphology was noted in both fluent and non-fluent participants) (Bird and Franklin, 1996). Accordingly, this system fails to categorise cases with mixed characteristics of both fluent and non-fluent aphasia, and therefore they are labelled undifferentiated/ unclassified aphasia (Bates et al., 2005; Clough and Gordon, 2020; Crary et al., 1992; Kasselimis et al., 2017).

There is a large literature within aphasiology considering the effects of focal brain damage on sentence production skills; sentence production models (Bock and Levelt, 1994a; Garrett, 1975) are commonly referred to in identifying the disrupted level of processing. Various research reports have sought to categorise sentence production deficits according to the aphasia sub-type in which they have been identified, particularly fluent versus non-fluent presentations (Edwards and Bastiaanse, 1998; Edwards and Tucker, 2006; Rochon et al., 2005; Thompson et al., 2003). However, in this current study, we will aim to capture fine-grained variation in sentence production skills across aphasia as a whole, including the range of severity and symptomatic sub-types. A novel scoring system for measuring sentence production skills will be introduced to account for specific sentence features not covered by the existing scoring protocol of the selected outcome measures. In doing so, we aim to provide a more continuous and graded picture of these skills relative to studies that have contrasted somewhat extreme points on criteria such as fluency.

We will also endeavour to investigate the correlation between specific language skills, within-subjects and across-subjects, that are linked to sentence production. According to psycholinguistic theories described above (see Chapter 2 for more information), it has been established that verbs play a central role in sentence comprehension and sentence production. The verb 'lemma' contains semantic and syntactic information that is essential for sentence processing. Failure to retrieve verbs or access the encoded information can be attributed to either a semantic deficit or a phonological deficit. Therefore, we will look into the individual correlation between phonological skills and sentence production skills as well as between semantic skills and sentence production skills. Additionally, we will examine the correlation between the ability to retrieve verbs and the ability to produce sentences. Since the success of verb production in PWA can be affected by the type of verb (i.e., frequency, transitivity, name relatedness) or the context of retrieval (i.e., availability of sentential cues, or sentence contexts which require verb inflection such as finite verbs), we will include both tasks in our analysis. These will consist of a confrontational action naming task at the word level, and fill-in sentences with a verb task. Similarly, we will investigate the correlation between performance on each individual task with performance on sentence production task.

A number of influential studies have proposed that sentence impairment is not caused by central loss of sentence processing competency but rather from a failure to map thematic roles to grammatical roles of a parsed sentence (i.e., mapping deficit hypothesis) (Linebarger et al., 1983; Saffran and Schwartz, 1988; Schwartz et al., 1987). Their conclusion was supported by evidence of a dissociation in PWA's performance on grammaticality judgement tasks and sentence comprehension tasks (Linebarger, 1990; Linebarger, 1987; Linebarger et al., 1983; Saffran and Schwartz, 1988). Adequate performance on the grammaticality judgment task shows an intact sentence parsing ability. Accordingly, comparing the participant's performance on grammaticality judgement task to their performance on sentence comprehension task may clarify whether their difficulty stem from a deficit in processing grammatical information or in mapping. Nevertheless, it is important to note that both skills can be affected by a lexical deficit. In this study, we will investigate the correlation between the participants' performance on verb comprehension, sentence comprehension and grammaticality judgment tasks with their performance on sentence production task.

Light verbs are defined as high frequency semantically empty words that generally convey nonspecific meanings (Berndt et al., 1997). They have lower imageability compared to concrete words, which is a semantic feature related to the mental image or sensory experience words evoke (Paivio et al., 1968). The degree of word imageability is typically associated with verb processing speed and accuracy (Alyahya et al., 2018a). Gordon (2007) hypothesised that "that there is a continuum of dependence of lexical production on syntax and semantics, with words like determiners at one end (almost purely syntactic), concrete nouns and heavy verbs at the other (almost purely semantic), and light verbs at an intermediate point". The connectionist model of sentence production by Gordon and Dell (2003), called the "division of labour", demonstrated that selective impairment in light verbs in aphasia and over reliance on full verbs can be induced by lesioning syntactic system in their model, which resembles the characteristics of agrammatic aphasia. On the other hand, when the syntactic system is preserved and the semantic system is lesioned within the same model, the reverse pattern was noted. The participants showed heavy reliance on full verbs and reduced production of light verbs. A study by Thorne and Faroqi-Shah (2016) supported the notion of the trade-off between light verb/syntactic abilities and semantic abilities in narrative language production as proposed by "division of labour" theory. Nevertheless Thorne and Faroqi-Shah (2016) argued that other factors could explain the observed difference in performance between light and full verb production in PWA; it can be attributed to semantic diversity, aphasia severity, or lexical frequency. Overall, only a small number of studies investigated the influence of verb semantics complexity on verb retrieval and sentence production. Some studies found that PWA relied on light verbs to overcome difficulties in retrieving full verbs (Berndt et al., 1997; Jespersen, 1992). However, the outcomes remain inconclusive and further research is required on a larger sample of participants (Barde et al., 2006; Cruice et al., 2003; Gordon, 2008).

A study by Barde et al. (2006) reported conclusive results that participants with syntactic impairment were specifically impaired in light verb retrieval skills. Accordingly, in the current study we will explore the correlation between performance on sentence production tasks and on light verb retrieval tasks.

In addition to advancing our knowledge by profiling verb and sentence production skills in a relatively large, varied and heterogeneous range of participants with aphasia through a systematic and sensitive evaluation, we aim to translate these findings into selection of candidates for the empirical therapy study, Chapter 4.

This chapter aims to address the following research questions:

1. What patterns of sentence production skills will be evident across a varied cohort of people with aphasia, with particular reference to fluent versus non-fluent aphasia presentations? How do these skills relate to related measures of lexical production and comprehension?
2. How reliable is a novel scoring system for measuring sentence production skills?
3. How does constrained production of light verbs in sentences vary with aphasia severity, and subtype?
4. How do sentence production skills vary according to measures of phonological and semantic processing skills and cognitive skills?
5. What candidacy factors can we observe in relation to participant engagement in sentence production therapy?

3.2 Methods

3.2.1 Participants

A convenient, non-selective sample was chosen for this study. The participants were recruited from a pre-established database of people with aphasia at the Neuroscience and Aphasia Research Unit (NARU) at the University of Manchester. For practicality, the participants were selected based on geographical proximity; those living closer to the University of Manchester were first invited to participate in the study. Participants with all aphasia subtypes, severities, and varying cognitive skills levels post left hemisphere stroke were included, except for participants with global aphasia as determined by the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass et al., 2001). The presence of apraxia, commonly associated with aphasia (McNeil et al., 2009), was not an exclusion criterion in this study. Nevertheless, exclusion criteria included the presence of active neurological disorder (e.g., brain tumour, uncontrolled seizures, traumatic brain injury, etc.,) and poor corrected vision or hearning.

Twenty-nine participants were recruited, including 16 fluent and 13 non-fluent participants, 12 females and 17 males. The sample represented a range of aphasia classifications according to their results on BDAE assessment (Goodglass and Kaplan, 1972). It included 15 individuals with anomia, 8 with Broca's, 3 with mixed non-fluent, 2 with conduction aphasia, and 1 with Transcortical Motor Aphasia (TMA). They all presented with a history of single left hemisphere ischemic or haemorrhagic stroke at the chronic stage of recovery, at least one-year post-onset. Moreover, 7 participants underwent additional testing using the Western Aphasia Battery (WAB) (Kertesz, 2007) to confirm the reliability of their aphasia classification, which was conducted in a subsequent study presented in Chapter 4. Time post-onset varied between participants, ranging from 27 to 211months (Mean= 92.6, SD= 47.5). The age of participants ranged from 44 to 87 years old (Mean= 63.7, SD= 10.9). Years of education ranged from 9 to 19 years (Mean= 12.2, SD= 2.4). All participants were native English speakers. At the time of testing, participants reported normal or corrected-to-normal hearing and visual acuity. Informed consent was taken from all participants with the approval of the National Research Ethics Committee, REC reference: 01/8/094.

Table 3.1 Basic demographical information (n=29)

Participant No.	Initials	Age (years)	Gender	Education (years)	Time-post onset (months)	BDAE aphasia classification
1	MH	67	M	11	41	Conduction
2	RH	66	M	17	27	Anomia
3	WE	65	M	10	103	Anomia
4	GHa	56	M	16	38	Anomia
5	DF	46	F	11	109	Anomia
6	EBo	46	M	11	84	Anomia
7	WC	87	M	9	45	Anomia
8	JS	71	F	19	81	Anomia
9	AL	54	F	12	156	Anomia
10	DM	53	M	17	124	Broca’s
11	PW	75	F	11	190	Mixed non-fluent
12	RR	60	M	13	90	Broca’s
13	AD	77	F	11	106	Broca's
14	GP	60	M	11	76	Anomia
15	AG	59	M	15	211	Broca's
16	CH	44	F	11	72	Anomia
17	MD	74	M	11	71	Mixed non-fluent
18	AB	52	M	13	112	Anomia
19	PR	73	F	12	84	Transcortical Mixed
20	JBO	79	M	13	186	Mixed non-fluent
21	PM	74	M	11	142	Broca's
22	PB	56	M	11	53	Anomia
23	DP	64	F	11	40	Anomia
24	JK	69	F	11	80	Anomia
25	JP	72	F	11	78	Anomia
26	MO	57	M	11	30	Broca’s
27	ST	69	F	11	77	Broca’s
28	DA	52	F	11	86	Broca’s
29	SH	70	M	11	94	Mixed non-fluent

3.2.2 Key Tasks

The measures included the Verb and Sentence Test (VAST) (section A and B: using stimulus booklets 1, 2, and 3) (Bastiaanse et al., 2002), the Light Verb Elicitation Test (LVET) (Carragher et al., 2013), and the Western Aphasia Battery (WAB) (Kertesz, 2007). Also, supplemental background neuropsychological data for 26 participants were retrieved from the NARU database. It included the Boston Naming Test (BNT) (Goodglass et al., 2001), the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) – Subtest 8: immediate repetition of non-words (Kay et al., 1996), 96-trial synonym judgment test (Jefferies et al., 2009), and Ravens Colored Progressive Matrices (RCPM)(Raven et al., 1962). The participants declared no prior knowledge of the testing materials which eliminated learning effects. Testing time was restricted to two hours to avoid experimental fatigue, and all the assessment tasks were completed in one session, except for two participants (PR and AG) who required two sessions. The primary researcher conducted the assessment and scoring, which took place at each participant's home, a familiar and quiet environment with minimal distractions.

a. The Verb and Sentence Test (VAST) (Bastiaanse et al., 2002)

The VAST is a language assessment battery designed to identify verb and sentence deficits in adults with aphasia in production and comprehension modalities. From 10 subtests within the test battery, only seven subtests were selected (200 trials) for this study. They were determined sufficient to answer the research questions while maintaining the sessions' duration within acceptable limits. The verbal production tasks included verb production assessment in multiple contexts such as confrontation naming, fill-in verbs in sentences, and sentence construction in picture description tasks. The auditory comprehension tasks included verb comprehension at the word level, sentence comprehension, and grammaticality judgment tasks. It examined performance on 4 sentence types active, passive, object cleft, and subject cleft sentences.

The maximum potential score on the VAST subtests was determined by the performance of neuro-typical controls, which was acquired from 80 native English speakers (Bastiaanse et al., 2002). The following is the range of scores the sample of controls have achieved on comprehension tasks: verb comprehension 38-40 points, sentence comprehension 35-40 points, grammaticality judgment 37-40 points; and the production tasks: action naming 37-

40 points, fill-in finite verbs and infinitives in sentences 8-10 points each, and sentence construction: 16-20 points. The VAST scoring manual provides a guideline for scoring each subtest individually. In our study, for further examination, we added three additional scores: total comprehension score (maximum 120 points representing the sum of scores on all 3 comprehension subtests), total production score (maximum 80 points representing the sum of scores on all 4 production tasks) and the composite score (representing performance on all 7 comprehension and production subtests using the following formula: (80+120/200) *100).

In each subtest, the examiner presented the instructions verbally and then demonstrated the task with a practice item. All the tasks included picture stimuli, presented in a field of four for comprehension tasks and in a field of one in production tasks, except the grammaticality judgment task. The participants were asked to respond to the comprehension tasks by pointing to the target picture and verbally producing their answers to respond to the production tasks. For the grammaticality judgment task, the participants were asked to indicate their judgment of the sentence they heard, whether it was grammaticality correct or not, through one of the following means: verbally by saying "good" or "bad", gesture thumbs up or thumbs down, or pointing to the word "good" with a checkmark symbol or the word "bad" with a cross symbol written on a paper. The examiner recorded participants' first spontaneous productions only. Any occasional cues provided after they have given their first answer were for the purpose of motivation and were not recorded or scored.

The scoring approach adhered to the VAST manual's instructions (Bastiaanse et al., 2002) in all except two subtests: the sentence construction and the fill-in sentences with verbs subtests. Since this study aimed to examine the characteristics of sentence production skills more closely and systematically, a specific scoring rubric was developed to analyze sentence content, grammaticality, and completeness (lexical retrieval, morphology, thematic and grammatical role assignment, word order, etc.). Appendix 1 contains this purpose-designed novel scoring protocol, which we developed to capture the range of clinical performance across people with aphasia. The reliability of this novel scoring rubric was determined through interrater reliability testing that included 25% of the sample. Also, the VAST was implemented fully for the subtests described, regardless of error production.

b. Light Verb Elicitation Test (LVET) (Carragher et al., 2013)

The light verb elicitation test examines the retrieval of light verbs in constrained fill-in verbs in a sentence task. It examines the retrieval of 10 light verbs (e.g., go, put, have etc.,) in a total of 30 trials, in which each verb is tested three times in three different sentences (see Appendix 2). The maximum potential score is 30. The performance of native English speakers (healthy controls) ranged between 27 and 30. For time efficiency, the test was terminated after consecutive errors on eight items.

3.2.3 Background neuropsychological assessments

a. Boston Naming Test (BNT) (Goodglass et al., 2001)

The BNT is a confrontational naming task that assesses noun retrieval skills at the word level in individuals with aphasia. It contains 60 items of picture illustrations of objects in drawings. The task instructions, verbally presented by the examiner, include presenting the participant with a picture stimulus in a field of one and a prompt to name the object that appears in each picture. Only independent productions within 10 seconds of presenting the picture were recorded and scored. The test was terminated after eight consecutive wrong answers. A study by Tombaugh and Hubiey (1997) reported that the average score achieved by neuro-typical native English adults on the BNT was 53 points.

Test items: bed, tree, pencil, house, whistle, scissors, comb, flower, saw, toothbrush, helicopter, broom, octopus, mushroom, hanger, wheelchair, camel, mask, pretzel, bench, racquet, snail, volcano, seahorse, dart, canoe, globe, wreath, beaver, harmonica, rhinoceros, acorn, igloo, stilts, dominoes, cactus, escalator, harp, hammock, knocker, pelican, stethoscope, pyramid, muzzle, unicorn, funnel, accordion, noose, asparagus, compass latch, tripod, scroll, tongs, sphinx, yoke, trellis, palette, protractor, and abacus.

b. The Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) – Subtest 8: immediate repetition of non-words (Kay et al., 1996)

The non-word repetition task was designed to test the integrity of the sub-lexical acoustic phonological conversion (auditory analysis and repetition skills). The authors suggest that performance may depend on phonological short-term memory. The test involves a total of

30 items of equal groups of 1-syllable, 2-syllable, and 3-syllable non-words. Examples include: ipical, splant, and vater. The following is the task instruction presented verbally to the participant "I'm going to say something to you. It's not a word, but it sounds as if it might be. See if you can say it after me". The maximum possible score for this test is 30 points.

c. 96-trial synonym judgment test (Jefferies et al., 2009)

The synonym judgment test was designed to examine participants' comprehension of concrete and abstract words with varying levels of imageability and frequency. In each trial, the participant is presented with one main written stimulus word and three written options written below the main word. The examiner reads the words out, then points to the stimulus and asks the patient, "which of the three words below are closest in meaning to this/target?". The participant will practice with 7 example exercises to ensure understanding before moving on to the test items. For example: student (target), pupil (synonym), summer (distractor), radio (distractor). The maximum score for this test is 96 points.

d. Ravens Colored Progressive Matrices (RCPM)(Raven et al., 1962)

The RCPM is a non-verbal assessment of cognitive function. The test consists of 36 trials with a maximum score of 36. In each trial, the participant is presented with a two-dimensional visual geometric design with a missing piece. Then, the participant is instructed to identify the missing item that completes the pattern from six or eight choices presented at the bottom of the page. "The test is similar to a jigsaw puzzle, except that the patterns are mostly meaningless, although the colours and some geometric figures are verbalizable" (Kertesz and McCabe, 1975, p.388).

3.3 Results

3.3.1 Research Question 1: What patterns of sentence production skills will be evident across a varied cohort of people with aphasia, with particular reference to fluent versus non-fluent aphasia presentations? How do these skills relate to related measures of lexical production and comprehension?

Sentence construction skills across participants

Participants' scores on the VAST language assessment battery are shown in Table 3.2, with participants ordered according to their performance on the sentence construction task. The wide range of scores illustrates the heterogeneity of the group. It spanned from failure to construct any clearly understandable simple sentence (e.g. participants PW, Jbo, and MD) to successful production of complete, grammatically correct, and informative sentences in all 20 trials (e.g. participants Ebo, WC, and Gha). For those participants who were able to produce some form of sentences, their performance varied widely. For example, some participants were not able to achieve a complete sentence for the target "The boy is pushing the girl": DA "The man waiting … woman", RR "Push away … (gesture) as well", and AD "Pushing her". Others were able to formulate a sentence but with some errors: JP "The man is saving her, save the girl", and DM "The boy pushing mum". Moreover, a few participants were able to formulate a sentence successfully, such as GP "The man pushes the woman", and MH "The man is pushing the girl". Table 3.3 shows the variety of sentence construction attempts across the cohort in response to one picture stimulus.

A close examination of the sentence construction scores (see Table 3.2: third set of column scores from the right) revealed that approximately one-third of the group scored 90% and above in production accuracy. Also, another third scored between 70% - 89% accuracy. The scores of the remaining third (i.e., nine participants) ranged between 0 and 70% accuracy. The cohort can be split into approximately two equal groups by setting 80% accuracy as a cut-off score. Accordingly, 14 participants scored above 80% accuracy level, and 15 participants scored below 80%. The group that scored above 80% accuracy is composed of participants with fluent aphasia except for one participant (DM). On the other hand, the group that scored below 80% accuracy is mainly composed of participants with non-fluent aphasia except three participants (MH, JS, and AB).

Table 3.2 VAST scores per participant (raw scores and percentage scores “%”)

Participant initials		Aphasia subtype*	Verb comprehension		Sentence comprehension		Grammaticality judgment		Total comprehension		Verb production		Fill-in verbs in sentences		Sentence construction		Total production		Composite score	
Maximum score			40	%	40	%	40	%	120	%	40	%	20	%	20	%	80	%	100	%
1	Ebo	F	40	100	37	93	39	98	116	97	31	78	16	80	20	100	67	84	91	91
2	WC	F	39	98	30	75	39	98	108	90	30	75	20	100	20	100	70	88	89	89
3	Gha	F	40	100	37	93	37	93	114	95	33	83	15	75	20	100	68	85	91	91
4	PB	F	38	95	34	85	39	98	111	93	33	83	19	95	19	95	71	89	91	91
5	AL	F	37	93	40	100	40	100	117	98	34	85	17	85	19	95	70	88	93	93
6	DP	F	40	100	32	80	35	88	107	89	39	98	19	95	19	95	77	96	92	92
7	CH	F	38	95	33	83	38	95	109	91	36	90	17	85	19	95	72	90	90	90
8	GP	F	39	98	22	55	32	80	93	78	32	80	14	70	18	90	64	80	79	79
9	WE	F	35	88	37	93	36	90	108	90	35	88	19	95	18	90	72	90	90	90
10	JK	F	32	80	12	30	24	60	68	57	21	53	17	85	18	90	56	70	62	62
11	DM	NF	39	98	17	43	25	63	81	68	39	98	18	90	17	85	74	93	78	78
12	DF	F	37	93	27	68	31	78	95	79	27	68	15	75	17	85	59	74	77	77
13	JP	F	37	93	21	53	30	75	88	73	22	55	13	65	17	85	52	65	70	70
14	RH	F	37	93	20	50	33	83	90	75	16	40	11	55	17	85	44	55	67	67
15	AG	NF	40	100	37	93	39	98	116	97	40	100	18	90	16	80	74	93	95	95
16	MH	F	32	80	27	68	35	88	94	78	8	20	6	30	15	75	29	36	62	62

17	JS	F	0	0	0	0	0	0	0	0	18	45	11	55	15	75	44	55	22	22
18	PR	NF	27	68	37	93	34	85	98	82	32	80	9	45	15	75	56	70	77	77
19	MO	NF	38	95	31	78	37	93	106	88	25	63	12	60	15	75	52	65	79	79
20	ST	NF	38	95	25	63	30	75	93	78	31	78	11	55	15	75	57	71	75	75
21	AD	NF	33	83	25	63	36	90	94	78	21	53	15	75	13	65	49	61	71	71
22	DA	NF	27	68	22	55	30	75	79	66	13	33	9	45	12	60	34	43	57	57
23	PM	NF	35	88	31	78	37	93	103	86	22	55	12	60	12	60	46	58	74	74
24	AB	F	40	100	27	68	28	70	95	79	20	50	5	25	10	50	35	44	65	65
25	RR	NF	40	100	20	50	20	50	80	67	22	55	11	55	9	45	42	53	61	61
26	SH	NF	25	63	13	33	30	75	68	57	18	45	10	50	6	30	34	43	51	51
27	PW	NF	38	95	26	65	29	73	93	78	3	8	2	10	0	0	5	6	49	49
28	Jbo	NF	35	88	23	58	33	83	91	76	24	60	9	45	0	0	33	41	62	62
29	MD	NF	6	15	8	20	0	0	14	12	0	0	0	0	0	0	0	0	7	7
Mean			34		26		31		90		25		13		14		52		71	
SD			9.5		9.6		9.9		27		10		5		6		20		21	

Table 3.3 Examples of sentences produced by participants in response to item 6 in the sentence construction subtest of the VAST (arranged according to their total score from high to low)

Participant No.	Initials	Target: "The boy is pushing the girl"	Raw score (total score= 9)
1	Ebo	The man pushes the woman	9
2	WC	The man is pushing the woman	9
3	Gha	The boy is pushing the girl	9
4	PB	The boy pushes the girl	9
5	AL	The boy is pushing a woman	9
6	DP	The man pushes the woman	9
7	CH	The man is pushing the girl	9
8	GP	The man pushes or pulls the woman	9
9	WE	The man is pushing the girl	9
10	JK	The man is taking the girl, pushing her around, away	9
11	DM	The boy pushing mum	7
12	DF	The boy is pushing the girl	9
13	JP	The man is saving her, save the girl	7
14	RH	The boy is pushing the girl	9
15	AG	The man is shoving the girl	9
16	MH	The man is pushing the girl	9
17	JS	He is pushing his wife along	8
18	PR	The chap (unintelligible production) the woman	5
19	MO	He pushing her, she doesn't want to go	5
20	ST	The man pushing the shoulder	6
21	AD	Pushing her	4
22	DA	The man waiting … woman	6
23	PM	The man is pushing the girl	9
24	AB	Doesn't want to go to buy his…	0
25	RR	Push away … (gesture) as well	2
26	SH	Pushing, banging	3
27	PW	Unintelligible production	0
28	Jbo	Unintelligible production	0
29	MD	Non response	0

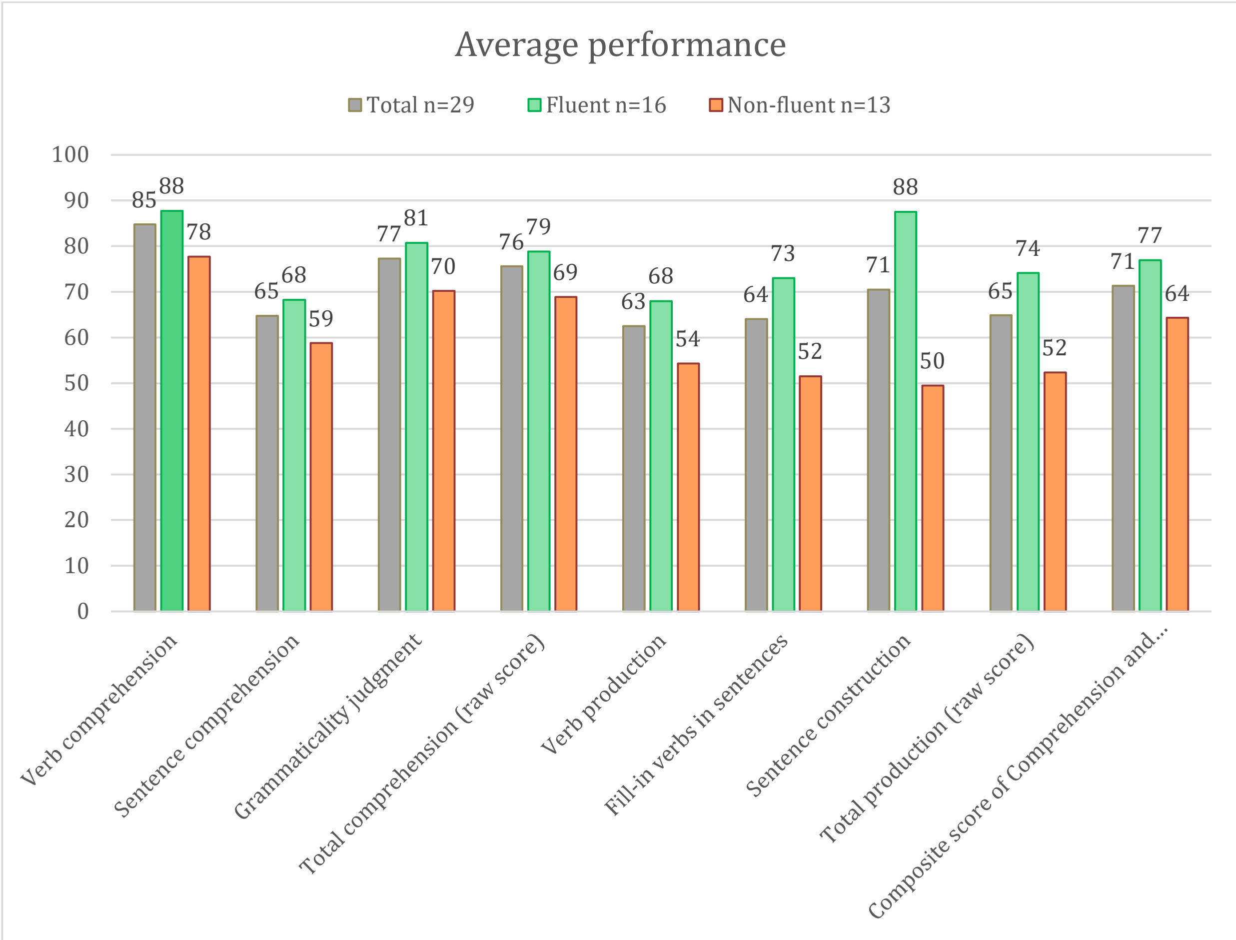


Figure 16 Mean performance of the fluent vs non-fluent groups compared to the mean performance of the total group on the VAST subtests (percentage scores%)

As well as affording us a global view on sentence production skills across a large cohort of people with aphasia, these data also permit a more detailed analysis of the relationship between specific measures and hence discreet linguistic processing skills.

Correlation between sentence production skills and comprehension skills

Total group performance

Spearman's rank correlation coefficient was administered to examine the relationship between sentence production scores and comprehension scores. The results of the correlation analysis are presented in Table 3.4 (below). The adjusted Bonferroni correction for multiple comparisons, in this case four comparisons ($p<0.05/4$), was $p<0.0125$. For the whole sample ($n=29$), a positive correlation between sentence production and verb comprehension approached but did not reach significance ($r_s=0.416$, $n=29$, $p=0.025$, two-tailed), typically sentence production and verb comprehension scores broadly aligned. However, sentence production and sentence comprehension were significantly correlated ($r_s=0.49$, $n=29$, $p=0.007$, two-tailed), as were sentence production and grammaticality judgment ($r_s=0.576$, $n=29$, $p=0.001$, two-tailed). The correlation between sentence production and the total score on the 3 comprehension tasks of the VAST also reached significance ($r_s=0.602$, $n=29$, $p=0.001$, two-tailed).

Performance based on aphasia subtype

We also analysed the group's performance based on their aphasia subtype. As shown in Figure 16 (above), the fluent group's performance on the production tasks (verb production, fill-in verbs in a sentence, and sentence construction) was substantially superior to the non-fluent group compared to relatively modest differences in performance on comprehension tasks. Also, there were no subtests on which the fluent group performed below the level of the non-fluent group.

To examine the relationship between sentence production scores and comprehension scores in each aphasia subtype, we implemented Spearman's rank correlation coefficient statistical test. Similarly, the adjusted Bonferroni correction for multiple comparisons was conducted, in this case eight comparisons ($p<0.05/8$), was $p<0.00625$. As seen in Table 3.4, for the fluent group ($n=16$), the correlation between sentence production and verb comprehension fell short of statistical significance ($r_s=0.568$, $n=29$, $p=0.022$, two-tailed). However, other comparisons reached significance, specifically sentence production and sentence comprehension (r_s =0.676, n=16, p=0.004, two-tailed), sentence production and grammaticality judgment (r_s =0.777, n=16, p=0.000, two-tailed), as well as sentence production and total score on the 3 comprehension tasks of the VAST (r_s =0.749, n=16, p=0.001, two-tailed). In broad terms, comprehension and production for fluent participants performance were closely aligned.

In contrast, the non-fluent group (n=13) showed no significant correlation between either sentence production and verb comprehension (r_s =0.404, n=13, p=0.17, two-tailed), or sentence production and sentence comprehension (r_s =0.444, n=13, p=0.129, two-tailed). Similarly, no significant correlations were evident for sentence production and grammaticality judgment (r_s =0.428, n=13, p=0.144, two-tailed), or sentence production and total performance on the 3 comprehension tasks within the VAST (r_s =0.528, n=13, p=0.064, two-tailed). It highlighted the marked performance discrepancy for these non-fluent participants with comprehension scores stronger than production.

Table 3.4 Spearman's rank correlation analysis of the VAST sentence production scores with comprehension scores of the VAST subtests (verb comprehension, sentence comprehension, and grammaticality judgment)

Sentence production	Verb comprehension		Sentence comprehension		Grammaticality judgment		Total comprehension	
	r_s	p	r_s	p	r_s	p	r_s	p
Total n=29	0.416	0.025	**0.49***	0.007	**0.576***	0.001	**0.602***	0.001
Fluent n=16	0.568	0.022	**0.676****	0.004	**0.777****	.000	**0.749****	0.001
Non-fluent n=13	0.404	0.17	0.444	0.129	0.428	0.144	0.528	0.064

*The adjusted Bonferroni correction for multiple comparisons; in this case, four comparisons (p<0.05/4) were applied. Correlation is significant at p<0.0125 (two-tailed)

** The adjusted Bonferroni correction for multiple comparisons; in this case, eight comparisons (p<0.05/8) were applied. Correlation is significant at p<0.00625 (two-tailed).

The correlation between sentence production skills and verb retrieval in different contexts

Total group performance

Additional statistical testing investigated the association between sentence production skills and the ability to produce verbs in several contexts. Thus, Spearman's rank correlation coefficient was administered to test the correlation between verb retrieval in a sentence context in response to a picture description task and verb retrieval in two other contexts: confrontational naming and fill-in verbs in sentences tasks. The results of the correlation analysis are presented in Table 3.5 (below) for a subset of 28 participants (due to missing data for one participant with fluent aphasia). The adjusted Bonferroni correction for multiple comparisons, in this case two comparisons ($p<0.05/2$), was $p<0.025$. The findings revealed that for the whole sample ($n=28$), there was a significant positive correlation between sentence production and action naming ($r_s=0.679$, $n=28$, $p=0.000$, two-tailed) as well as between sentence production and fill-in verbs in sentences ($r_s=0.817$, $n=28$, $p=0.000$, two-tailed).

Performance based on aphasia subtype

The group was further divided into fluent and non-fluent to compare the performance between aphasia subgroups. The adjusted Bonferroni correction for multiple comparisons, in this case four comparisons ($p<0.05/4$), was $p<0.0125$. The results of the fluent group ($n=15$) showed significant positive correlation between sentence production and fill-in verbs in sentences ($r_s=0.792$, $n=15$, $p=0.000$, two-tailed) and between sentence production and action naming ($r_s=0.633$, $n=15$, $p=0.011$, two-tailed). For the non-fluent group ($n=13$) there were significant positive correlations for sentence construction and both fill-in verbs in sentences ($r_s=0.728$, $n=13$, $p=0.005$, two-tailed) and action naming ($r_s=0.811$, $n=13$, $p=0.001$, two-tailed). Overall, there were no major discrepancies between production skills across the cohort or between aphasia subtypes.

Table 3.5 Spearman's rank correlation analysis of the VAST sentence production scores with two different production subtests' scores of the VAST(action naming and fill-in verbs in sentences).

Sentence production	Action naming		Fill-in verbs in sentences	
	r_s	p	r_s	p
Total n=28	**0.679***	0.000	**0.817***	0.000
Fluent n=15	**0.633****	0.011	**0.792****	0.000
Non-fluent n=13	**0.811****	0.001	**0.728****	0.005

* The adjusted Bonferroni correction for multiple comparisons; in this case, two comparisons ($p<0.05/2$) were applied. Correlation is significant at $p<0.025$ (2-tailed).

** The adjusted Bonferroni correction for multiple comparisons; in this case, four comparisons ($p<0.05/4$) were applied. Correlation is significant at $p<0.0125$ (2-tailed).

3.3.2 Research Question 2: How reliable is the novel scoring system for measuring sentence production skills.

To determine the reliability of the novel scoring rubric used to analyse the participants' sentence production skills (see Appendix 1), we conducted an inter-rater reliability assessment. A Speech-Language Therapist/Pathologist SLT/P with no knowledge or experience in administering the VAST test performed the independent rater's role. Using the scoring rubric and a sample of the picture stimuli used in the VAST, the independent rater scored the performance of 31% of the sample (9 out of 29 participants). The picture stimuli were provided to aid the rater in judging the relevance of the produced sentence to the target. As long as the produced sentence successfully described the picture scene, it was considered correct even if it did not match the target sentence suggested in the test manual. For example, one item within the VAST sentence construction subtest constantly confused the participants (it contained a picture of a clown smiling and the participants found it challenging to identify the target event; however, they produced an acceptable alternative such as "the clown is posing in front of the audience"). Table 3.6 (below) presents the two raters' scoring outcomes and the percentage of agreement between the two. With a mean agreement score of 97% across the sample, this analysis suggested a strong agreement between raters and supported the scoring rubric.

Table 3.6 Agreement of scoring produced by rater 1 and rater 2 for the sentence production subtest of the VAST. The reliability testing included 9 samples from 9 different participants (total sample n=29).

Participant	Rater 1	Rater 2	Agreement
	Total raw score (160 points) for 20 trials		% percentage
1	98*	114*	90
2	138	138	100
3	135	140	97
4	123	125	99
5	121	128	96
6	134	134	100
7	70	72	99
8	117	126	94
9	147	145	99
		Mean	97
		Range	90-100

* the discrepancy in scoring could be attributed to rater 2's limited experience. Participant 1's (first on the list above) production was the first sample rater 2 scored independently following two practice exercises with the primary investigator.

3.3.3 Research Question 3: How does constrained production of light verbs in sentences vary with aphasia severity and subtype?

As shown in Figure 17 the whole group's performance on the LVET task was lower than their performance on the other production tasks, such as confrontation word production for both nouns and verbs (BNT and VAST action naming), fill-in verbs in sentences, and sentence construction. A clear distinction in performance between the fluent and non-fluent groups was evident in the LVET task. The fluent group's average score was markedly superior to the average score of the non-fluent group with a ratio of 2.4:1.

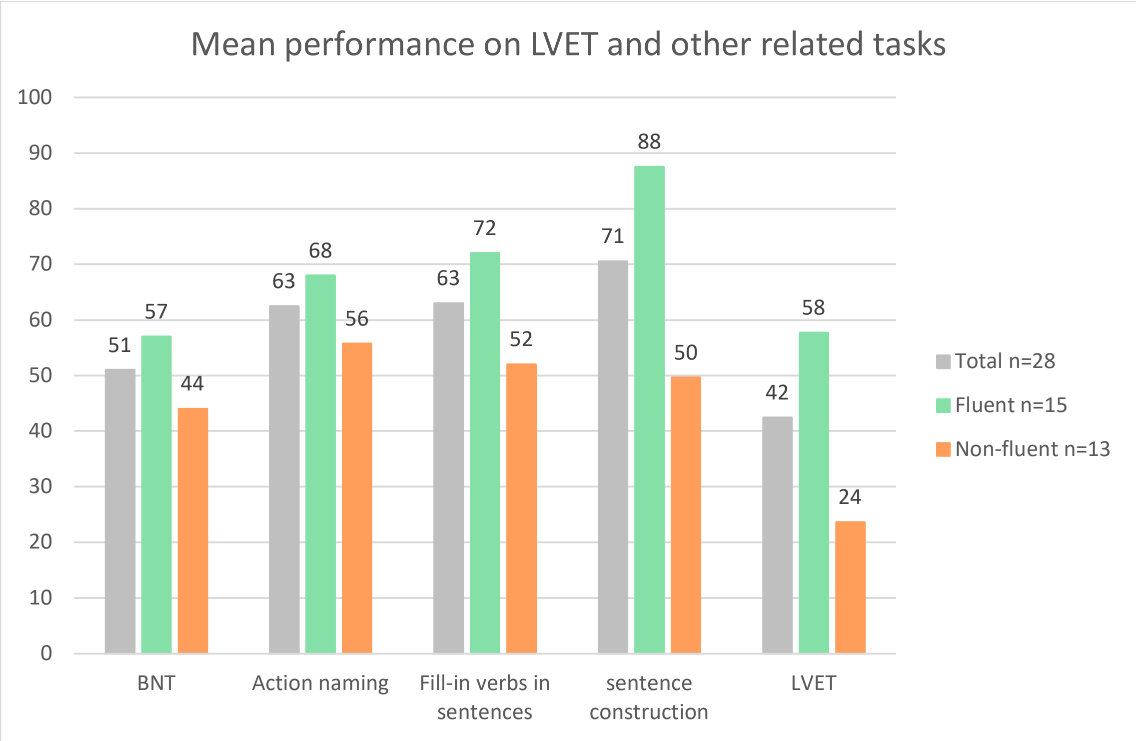


Figure 17 Mean performance on LVET compared to mean performance on four related tasks: BNT, VAST action naming, VAST fill-in verbs in sentences, and VAST sentence production task.

Table 3.7 Background psycholinguistic data (percentage scores)

Participant	BNT	LVET	Ravens	Auditory non-word repetition (PALPA 8)	96 synonym judgment task
Maximum score	100	100	100	100	100
1. Ebo	55	97	97	100	91
2. WC	55	60	53	30	95
3. Gha	78	53	83	80	96
4. PB	85	90	81	96.66	80
5. AL	88	100	92	90	94
6. CH	60	60	92	60	88
7. GP	57	57	97	43.3	90
8. WE	55	80	92	46.67	88
9. DM	72	0	92	60	96
10. DF	50	57	89	53.3	78
11. JP	57	57	83	70	83
12. RH	2	40	83	3.33	90
13. AG	78	73	75	73.3	90
14. MH	5	0	81	0	70
15. JS	43	87	100	50	97
16. PR	38	27	81	56.67	83
17. MO	47	67	86	30	65
18. ST	50	23	47	53.33	-
19. AD	50	70	64	23.3	82
20. PM	52	3	47	10	70
21. AB	42	0	89	26.67	75
22. RR	23	0	89	10	82
23. SH	23	17	56	6.66	70
24. PW	5	0	50	0	74
25. Jbo	10	23	78	23.3	66
26. MD	38	0	39	26.67	57
Average	46.9	43.9	77	43.2	79
SD	23.7	33.7	17.6	30.2	11.0

Figure 18 (below) clearly shows the linguistic space occupied across the cohort with respect to sentence production and light verb production. Though the fluent versus non-fluent sub-groups overlap substantially, there is a clear fluent 'space' in the top right quadrant of this Figure (good light verb retrieval and good sentence production) and a clear non-fluent 'space' in the lower quadrants (poor light verb retrieval and variable sentence production).

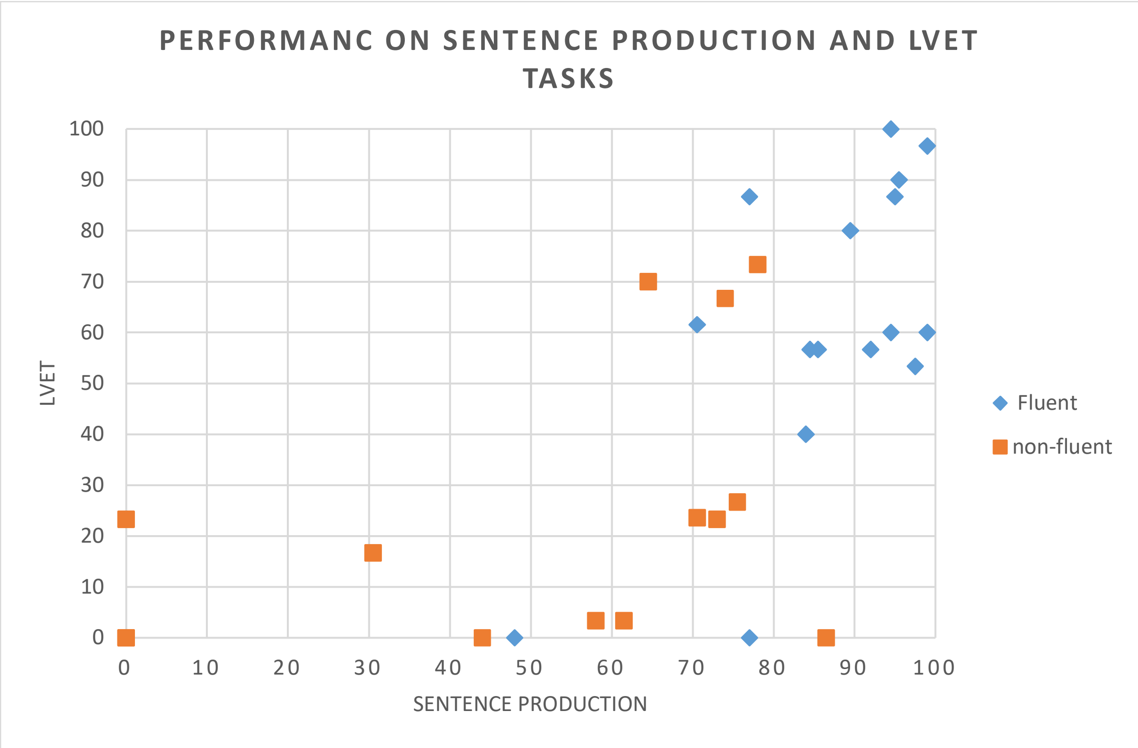


Figure 18 Illustrated the performance of the fluent and non-fluent subgroups on the VAST sentence production task

3.3.3.1 Correlation between scores on light verb production in a sentence completion task and full verb production in three tasks: confrontational naming, fill-in verbs in a sentence, and sentence production

The correlation between light verb production in a sentence completion task and full verb production in three different contexts was examined using Spearman's rank correlation coefficient. The verb production contexts included action naming in response to a confrontation naming task, verb retrieval in a sentence completion task, and verb production in a sentence context in response to a picture description task. The results of the correlation analysis are presented in Table 3.8 (below), which included a subset of 28 participants due to missing data for one participant with fluent aphasia. The adjusted Bonferroni correction for multiple comparisons, in this case three comparisons ($p<0.05/3$), was $p<0.016$. The findings show that for the whole sample (n=28), there was a significant positive correlation between performance on LVET and action naming (r_s =0.547, n=28, p=0.003, two-tailed), performance on LVET and scores on fill-in verbs in a sentence (r_s =0.715, n=28, p=0.000, two-tailed), and performance on LVET and verb production in a sentence context (r_s =0.678, n=28, p=0.000, two-tailed).

The fluent group (n=15) showed a significant positive correlation between performance on LVET and scores on fill-in verbs in a sentence (r_s =0.669, n=15, p=0.006, two-tailed), with the adjusted Bonferroni correction for multiple comparisons, in this case six comparisons ($p<0.05/6$), was $p<0.008$. Other comparisons did not reach statistical significance: LVET and action naming (r_s =0.574, n=15, p=0.025, two-tailed), and LVET and verb production in a sentence context (r_s =0.557, n=15, p=0.031, two-tailed). For the non-fluent group (N=13), there were no significant correlations: LVET and action naming (r_s =0.492 , n=13, p=0.087, two-tailed), LVET and scores on fill-in verbs in a sentence (r_s =0.415 , n=13, p=0.159, two-tailed), and LVET and verb production in a sentence context (r_s =0.466 , n=13, p=0.108, two-tailed). The findings are consistent with the non-fluent subgroup's markedly poor performance on light verb retrieval and the implication that words of such low imageability fell beyond their expressive vocabulary range. Thus, they found the LVET task exceptionally difficult relative to the other production-focused tasks.

Table 3.8 Spearman's rank correlation analysis of the LVET scores with the action naming, fill-in verbs in sentences, and sentence production scores of theVAST subtests.

LVET	Action naming		Fill-in verbs in sentences		Sentence Production	
	r_s	p	r_s	p	r_s	p
Total *n*=28	**0.547***	0.003	**0.715***	0.000	**0.678***	0.000
Fluent *n*=15	0.574	0.025	**0.669****	0.006	0.557	0.031
Non-fluent *n*=13	0.492	0.087	0.415	0.159	0.466	0.108

* The adjusted Bonferroni correction for multiple comparisons; in this case, three comparisons (p<0.05/3) were applied. Correlation is significant at p<0.016 (2-tailed).

** the adjusted Bonferroni correction for multiple comparisons; in this case, six comparisons (p<0.05/6) were applied. Correlation is significant at p<0.008 (2-tailed).

3.3.3.2 Correlation between light verb production and object naming

The correlation between light verb production in a sentence completion task and object naming in a confrontation task was examined using Spearman's rank correlation coefficient. The results of the correlation analysis are presented in Table 3.9 (below), it included a subset of the sample due to missing data for one participant with fluent aphasia. It shows that for the whole sample (i.e., *n*=28), there was a significant positive correlation between LVET scores and object naming scores (r_s =0.634, *n*=28, p=0.000, two-tailed).

Sub-group analyses required an adjusted Bonferroni correction for multiple comparisons, in this case two comparisons (p<0.05/2), which was p<0.025. Results showed non-significant correlations between LVET scores and object naming scores for the fluent group (*n*=15) (r_s =0.517 , *n*=15, p=0.048, two-tailed) and the non-fluent group (*n*=13) (r_s =0.384 , *n*=13, p=0.195, two-tailed). The whole group effect had, therefore, been carried by the fluent participants overall.

Table 3.9 Spearman's rank correlation analysis of the LVET scores with the object naming task of the BNT

LVET	Object naming BNT	
	r_s	p
Total *n*=28	**0.634***	0.000
Fluent *n*=15	0.517	0.048
Non-fluent *n*=13	0.384	0.195

* The adjusted Bonferroni correction for multiple comparisons; in this case, two comparisons (p<0.05/2) were applied. Correlation is significant at the p<0.025 level (2-tailed).

3.3.4 Research Question 4: How do sentence production skills vary according to measures of phonological and semantic processing skills and cognitive skills?

To explore the relationship between sentence production skills (as measured by the VAST sentence production task) and the participants' semantic, phonological, and cognitive skills, we conducted a multiple regression analysis using SPSS software (version 23.0.0.2). The analysis included a subset of 26 participants based on their background data availability at the time of running this test. However, as shown in Table 3.10, the participants' scores on the three tasks that represent those skills (non-word repetition immediate PALPA-8, 96 synonym judgment task, and Raven's Colored Progressive Matrices) were not significant predictors of sentence production skills.

Table 3.10 The unstandardized and standardized regression coefficients for the variables entered into the model

Variable	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error B	Beta	t	Sig.
Non word repetition (immediate) PALPA	0.462	0.184	0.45	2.518	0.2
96 synonym judgment	0.127	0.305	0.08	0.418	0.68
Raven's	0.490	0.354	0.284	1.384	0.18

*Dependent variable is sentence production skills

Given a lack of a strong and direct predictive relationship between these variables and sentence production, we next conducted Spearman's rank correlation coefficient tests to explore if there were correlational analyses between these measures. The results of the correlation analysis are presented in Table 3.11 (below). Similarly, the analysis included a subset of 26 participants based on their background data availability at the time of running this test. The adjusted Bonferroni correction for multiple comparisons, in this case three comparisons ($p<0.05/3$) was $p<0.016$. These results show that for the whole sample (i.e., n=26), there was a significant positive correlation between sentence production and scores on Raven's cognitive test (r_s =0.518, n=26, p=0.007, two-tailed), sentence production and scores on 96 synonym judgment task (i.e., tests receptive semantics) (r_s =0.719, n=26, p=0.000, two-tailed), and lastly, sentence production and PALPA8 non-word repetition immediate (i.e. tests phonology) (r_s=0.726, n=26, p=0.000, two-tailed). In broad terms, these findings implied a general effect of severity: participants who were severely impaired scored more poorly on all these measures, while those who were mildly impaired, showed higher performance across these skills.

Then, the sample was divided into two sub-groups, fluent and non-fluent, for comparison. The adjusted Bonferroni correction for multiple comparisons, in this case six comparisons (p<0.05/6) was p<0.008. The fluent group (i.e., n=14) showed no significant correlation between sentence production and scores on Raven's cognitive test (r_s=-0.085, n=14, p=0.773, two-tailed), nor between sentence production and scores on 96 synonym judgment task (r_s=0.467, n=14, p=0.092, two-tailed). Similarly, the non-fluent group (i.e., n=12) showed no significant correlation between: sentence production and scores on Ravens cognitive test (r_s=0.508 , n=12, p=0.092, two-tailed) nor between sentence production and scores on 96 synonym judgment task (i.e., tests receptive semantics) (r_s=0.537 , n=12, p=0.072, two-tailed). While the non-fluent group showed a significant correlation between scores on sentence production and PALPA8 non-word repetition immediate (r_s=0.799, n=12, p=0.002), for the fluent group, this comparison approached but did not reach significance (r_s=0.622, n=14, p=0.018, two-tailed). It suggests that the correlation between sentence production and phonological skills across the cohort and subgroups was the strongest of those we examined.

Table 3.11 Spearman's rank correlation analysis of the VAST sentence production score with the Raven's, 96 synonym judgment task, and auditory non-word repetition (immediate) PALPA 8 tasks.

Sentence production	Ravens		96 synonym judgment task		Auditory non-word repetition (immediate) PALPA 8	
	r_s	p	r_s	p	r_s	p
Total n=26	**0.518***	0.007	**0.719***	0.000	**0.726***	0.000
Fluent n=14	-0.085	0.773	0.467	0.092	0.622	0.018
Non-fluent n=12	0.508	.092	0.537	0.072	**0.799***	0.002

* The adjusted Bonferroni correction for multiple comparisons, in this case three comparisons (p<0.05/3), was applied. Correlation is significant at p<0.016 level (2-tailed).

** The adjusted Bonferroni correction for multiple comparisons, in this case six comparisons (p<0.05/6), was applied. Correlation is significant at p<0.008 level (2-tailed).

3.3.5 Research Question 5: What candidacy factors can we observe in relation to participant engagement in sentence production therapy?

The next chapter (Chapter 4) will aim to test the feasibility of a novel sentence therapy method in remediating sentence production deficits in aphasia. A final research question in this chapter focused on identifying candidates for the therapy study from the participants we screened (n=29). First, the language skills that are essential for successful engagement in the therapy program were defined. It included the presence of sentence or discourse deficits, the participant's awareness of the communication breakdown and the motivation to address them in therapy, a minimum level of auditory comprehension to understand multi-step verbal instructions. Moreover, since the therapy method will involve computer use, basic computer skills and some hand dexterity (either dominant or non-dominant hand) was required. Also, the therapy method involves recording and replaying the participant's verbal productions; therefore, a certain level of speech intelligibility is important to understand the recordings and self-assess the productions for any errors.

With this in mind, we first eliminated participants who had no concerns about their verbal communication and showed no difficulties in sentence or discourse production. The estimation was based on their high scores on standardised tests and supported by clinical observation (e.g., absence of hesitation or disfluency in spontaneous conversations). Those conditions applied to participants EBO, WC, Gha, PB, and AL (see Table 3.2). Next, we eliminated participants with severe apraxia or dysarthria who could not produce a single form of a phrase or a sentence, indicated by scoring zero in the sentence construction task. It included three participants: PW, JBO, and MD. For those participants, due to their severely unintelligible speech, Augmentative and Alternative Communication AAC systems would arguably better suit their immediate therapeutic needs. As explained, auditory comprehension is a prerequisite to benefit from the therapy trails. However, it is difficult to determine the minimum level needed to predict a successful engagement in the therapy program's tasks at this stage of the empirical work. An initial cut-off criterion was established to narrow down the sample of participants who would get invited to undergo a short therapy trial for stimulability assessment. We identified participants who performed poorly (below 50% accuracy) on the sentence comprehension task in addition to at least two more subtests within the VAST battery. It applied to two participants SH and JS. However, in JS's case, her poor performance on the comprehension tasks was strictly associated with implementing the

VAST test manual's scoring guidelines. When support was provided to the participant, such as repeating the verbal prompt, her scores reached ceiling in all subtests. Therefore, based on clinical observation, it was determined that she was likely to perform well in a more relaxed environment as in a therapy session, where prompts and assistance from the clinician will be available as needed. For that reason, participant JS was identified as a candidate for the therapy study. On the other hand, participant SH performed poorly, with less than 50% accuracy, on the sentence comprehension task, verb production, and sentence construction subtests. His performance did not improve with external support and prompts from the clinician; therefore, it was predicted that he would not be able to manage the demands of the therapy tasks.

In conclusion, a subset of 20 participants was identified as good candidates for the therapy study based on their performance on standardised tests as shown in Figure 19. This included the following participants: DP, CH, GP, WE, JK, DM, DF, JP, RH, AG, MH, PR, MO, ST, AD, DA, PM, AB, JS RR. They were invited to participate in an additional assessment session, baseline 2, which included a stimulability testing through a short demonstration of therapy tasks and observation of their response to cues and prompts.

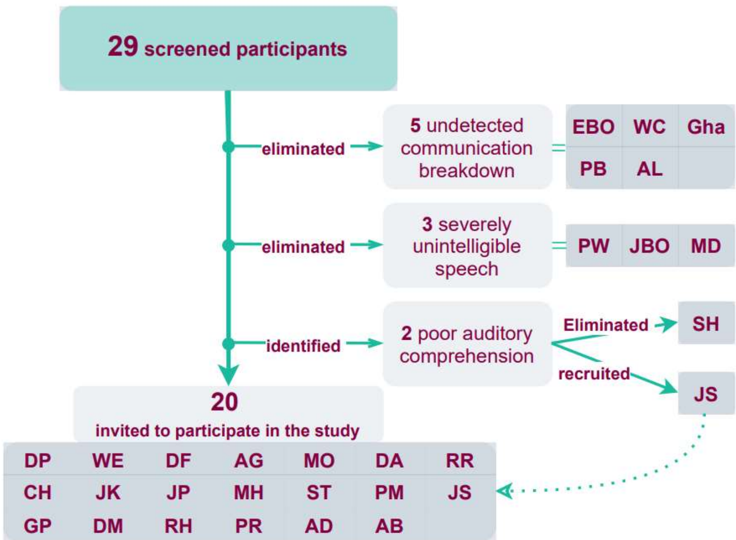


Figure 19 Candidate selection for therapy study

Results summary

Table 3.12 Summary of Table 3.4 Table 3.5 and Table 3.11 illustrating significant results using Spearman's rank correlation analysis

Correlation of sentence production skills with the following variables:	Total sample	Fluent	Non-fluent
Verb comprehension			
Sentence comprehension	✓	✓	
Grammaticality judgment task	✓	✓	
Action naming	✓	✓	✓
Fill-in verbs in sentences	✓	✓	✓
Ravens	✓		
96 synonym judgment task	✓		
Auditory non-word repetition (immediate) PALPA 8	✓		✓

Table 3.13 Summary of Table 3.8 and Table 3.9 illustrating significant results using Spearman's rank correlation analysis

Correlation of performance on LVET with the following variables:	Total sample	Fluent	Non-fluent
Action naming	✓		
Fill-in verbs in sentences	✓	✓	
sentence production	✓		
BNT	✓		

3.4 Discussion

The study aimed to examine sentence processing skills across a relatively large heterogeneous sample of participants. The findings presented a fine-grained snapshot of the variability of those skills across aphasia as a whole, and particularly, contrasting fluent and non-fluent variants. Having drawn participants from as wide a set of recruitment criteria as were feasible, there was a striking overall pattern in production skills towards high scoring performance across the cohort, which was largely driven by the participants with fluent aphasia.

To establish if our findings aligned with the characteristics reported in the fluent vs non-fluent language production literature, we analysed the scores of the total sample of participants, n=29, on the VAST test battery. The first analysis focused on what type of participant scored the highest on the sentence production task. A high score indicated that a participant was able to produce a complete, contentful, grammatically correct independently, and informative sentence in a simple picture description task. The informativeness aspect will negate the "empty speech" feature usually observed in fluent aphasia, and the grammaticality correctness will negate agrammatism usually associated with non-fluent aphasia.

Sentence production skills across the sample

The findings showed that 93% of those who scored above 80% accuracy level on the VAST sentence production task (a picture description task) were participants with fluent aphasia. However, it is worth mentioning that our sample of participants was mainly composed of participants with anomia classification (15 out of 29) and lacked participants with jargonistic symptoms of Wernicke's aphasia. Including participants with these symptoms would have bridged the severity gap between the two sub-groups. By definition, participants with fluent aphasia had stronger word retrieval skills and phonology scores. In addition to their superior performance, they were more reliable in generating acceptable synonyms in place of non-retrieved target words.

Moreover, the nature of the selected outcome measure, the VAST sentence production task, did not constrain participants' production across a range of sentence structures in response to the picture description task, although participants showed a preference for producing simple active sentences. More errors will likely be discovered upon increasing the task challenge and including constraints to prompt the production of more types of sentences (e.g. subject cleft, passives) with a range of complexity levels (Edwards and Bastiaanse, 1998; Hesketh and Bishop, 1996). A similar breakdown in PWA performance, induced by increased task complexity, has been reported in the literature (Lesser, 1989; Mitchum and Berndt, 1994; Weinrich et al., 1997). In those studies, the same participants who performed adequately in a picture description task (used full sentences) showed a discrepancy in performance by producing ill-formed and fragmentary utterances when the task required multi-sentence productions.

Nevertheless, this study's scoring guideline may have influenced our interpretation of the non-fluent participants' performance on the VAST sentence production task. The scoring rubric allocated points for each correctly produced content word and for each grammatical word that is typically needed to construct a grammatically correct sentence (see appendix 1). In our sample, several participants with non-fluent aphasia demonstrated frequent failure to produce content and function words; consequently, a large portion of the total score was deducted. For example, participant AD and RR constantly omitted subjects in a sentence and therefore produced incomplete sentences (e.g., AD "Pushing her", RR "Push away … (gesture) as well"). These characteristics are consistent with what has been reported in the literature of agrammatism in people with non-fluent aphasia, which includes the omission of the subject in a sentence, its main verb, function words, and inflections (Armstrong, 2000; Goodglass, 1993). On the other hand, participants with fluent aphasia in our sample also experienced word retrieval difficulties; however, they were more reliable in generating acceptable word substitutions. As a result, they would get the full score for the item even if it took many self-correction attempts and a long time to produce it. Those observed characteristic of fluent aphasia were in agreement with Bird and Franklin (1996) findings. The authors reported that the verbal output of individuals with fluent aphasia is mainly recognized for its frequent word substitution. Accordingly, the scoring system we implemented may have masked the deficit in fluent-aphasia and highlighted it in non-fluent fluent aphasia. A possible modification for better control of conditions would be to specify a fixed time limit for sentence production attempts that is equal for both subgroups, fluent and non-fluent.

In addition to profiling sentence production skills across a relatively large heterogeneous sample of PWA the data also allowed us to explore the correlation between specific linguistic skills. As seen in the summary Table 3.12 and Table 3.13, almost all tasks show a significant correlation with sentence production skills across the whole sample. Nevertheless, it is important to note that associations are not causal. Also, due to limitations in matching fluent and non-fluent subgroups, the contrast found between the two subgroups in statistical significance, of the correlation between specific linguistic skills and sentence production, should be interpreted with caution.

The correlation between verb production and sentence production skills

In the aphasia literature, it was hypothesized that the inability to access a target verb and the semantic and syntactic information contained within it (predicate-argument structure) is a cause for sentence production impairment, as the encoded information is essential in determining the structural frame of a sentence to be constructed (Grimshaw, 1990; Saffran, 1982). Our findings supported this notion (see Table 3.12). The non-fluent participants performed poorly on the action naming subtest, and the correlation analysis (Spearman's rank correlation coefficient) revealed a significant association with sentence production skills. The outcome was further supported by a significant correlation between performance on the fill-in verbs in a sentence task and the sentence production task. It showed that verb production in two contexts, confrontational naming and fill-in verbs in sentences, correlated with sentence production skills. Those findings were not restricted to the non-fluent subgroup but extended to include the fluent subgroup and the total sample.

The correlation between comprehension and production skills

The assessment of performance across the various subtests of the VAST enabled us to capture the patterns of relationship between related sentence processing skills, especially receptive versus expressive skills. The average performance of both aphasia subgroups in our sample on the comprehension tasks was below the level of neuro-typical control's performance reported in the literature (Bastiaanse et al., 2003). However, the fluent subgroup was generally more mildly impaired across both receptive and expressive skills, to the extent that mild-moderate impairments in expression were correlated with mild-moderate impairments in comprehension. For the non-fluent group, production skills were markedly poorer than comprehension, such that these skills did not correlate statistically. The fluent and non-fluent groups demonstrated a comparable performance on the comprehension tasks; however, there is a notable gap in their performance on the production tasks. Our findings align with reports in the literature which concluded that agrammatical errors in production do not necessarily originate from comprehension deficits (Goodglass, 1993).

However, the noted discrepancy in performance between receptive and expressive sentence processing skills could also be attributed to the design of each task. It is worth noting that the VAST comprehension tasks examined participants' comprehension of four types of sentence structures: active, passive, subject cleft, and object cleft sentences. On the other hand, the sentence production task did not restrict the participants' production to specific sentence structures; although most of the produced sentences were active simple sentences. It is reported in the literature that sentences produced by individuals with fluent aphasia lack complexity in structure (Bird and Franklin, 1996; Butterworth and Howard, 1987; Edwards, 1995). Therefore, a more sensitive measure that examines similar sentence types in both modalities, production and comprehension, could be more informative. Doing so could unmask a deficit in sentence production skills in fluent aphasia and reveal a gap between sentence comprehension and production skills similar to those observed in non-fluent aphasia.

<u>A novel scoring protocol for the VAST sentence production task</u>

This study's original contribution included the development of a scoring protocol for the VAST sentence production task. The VAST test manual did not propose a standard scoring system for this subtest. The approach used in standardization entailed "when a person produced a grammatical sentence that matched the picture, this was scored as correct, even when the response differed from the target sentence"; the author also explained, " the absolute score on this task is far less important than an analysis of the errors" (Bastiaanse et al., 2002, p.16). Accordingly, we designed a novel scoring rubric to achieve specificity in performance measurement across the key skills in sentence production. It accounted for lexical retrieval of content words, grammatical markings and deictic marking (definite/indefinite articles), tense morphology on verbs, and function words (see Appendix 1). In addition to demonstrating high levels of inter-rater reliability in this piloting, the sentence production task and the scoring protocol showed that relatively brief sampling of simple sentence level performance could convey a substantial amount of information about expressive processing skills, with efficiency in time and effort. It included information on morphology which may be a particular advantage in working within non-Indo-European language families, such as Arabic, where morphological processing tends to be richer and more meaning-bearing than English (Ryding, 2014). One limitation of the VAST sentence construction subtest in its current form was that many milder fluent participants did perform at or near the ceiling, which could limit its sensitivity in capturing therapy-induced changes. In that context, additional testing with non-picture based discourse tasks (e.g., procedural narratives or story retelling), which requires greater cognitive-linguistic demand, could be useful in revealing and measuring subtle expressive aphasic symptoms (Armstrong, 2000).

Light Verbs

In the literature, sentence deficits with relative impairment in verb retrieval are commonly associated with high reliance on the use of light verbs (e.g., 'go', 'get', 'have', 'do' and 'be') in sentence production (Berndt et al., 1997). Accordingly, we expected our analysis to reveal high scores on light verb retrieval skills when verb impairment was evident; however, the outcomes were mixed. The participants' average performance on verb retrieval tasks in three contexts (confrontational naming, fill-in verbs in sentences, and sentence construction in a picture description task) showed a clear deficit; nevertheless, their performance on the LVET task was markedly below that level, showing more severity (see Figure 17). While the fluent subgroup found light verb retrieval (i.e., LVET task) relatively difficult compared to other production tasks, their overall performance on those tasks was well aligned. On the other hand, in the non-fluent subgroup, light verb retrieval appeared to fall outside their range of production skills.

This discrepancy in performance between light verbs and full verbs may be attributed to the nature of the LVET task in our study (i.e., a constrained fill-in verbs in a sentence task). Also, in the sentence construction task (i.e., a picture description task) a participant is required to produce the target verb or a synonym that matches the stimulus in order to get the full score.

Therefore, the scoring guideline does not account for light verb production when produced to overcome word finding difficulty. It has been suggested that participants' reliance on light verb production is related to their intact sensitivity to the need for a verb to complete a sentence(Berndt et al., 1997). Therefore, this pattern of high reliance on light verbs associated with noticeable difficulties in producing full verbs may be more evident in less structured tasks such as story retell, narratives, and conversations.

Another possible reason for this discrepancy in performance could be related to the imageability effect for these high-frequency but very poorly imageable words (do, go, have etc.), which distinguishes light verbs from full verbs (tested in VAST action naming task) and nouns (tested in the BNT task) (Alyahya et al., 2018b; Conroy et al., 2009).

Given the linguistic characteristics of light verbs (i.e., high frequency, semantically empty words that generally covey nonspecific meanings), they are flexible and highly strategically useful if used to replace an unavailable verb due to a lexical retrieval deficit. With a limited number of light verbs in English, 10 according to Gordon and Dell (2003), the item-specific nature of lexical learning in aphasia therapy becomes an advantage rather than a disadvantage, in that specific learning and utilization of this narrow set of verbs is all that is functionally required (Whitworth et al., 2014). However, our current findings present a challenge therapeutically, in that light verbs could arguably be beyond any zone of 'learnability' or proximal development (Vygotsky, 1980) for non-fluent participants. On the other hand, it has been established that gesture therapy and semantic feature analysis can offer processing support in therapy to people with aphasia in relation to light verbs (Carragher et al., 2013).

Correlation between sentence production and background cognitive skills

The fact that sentence production skills did not appear to relate very closely to background measures representing non-linguistic cognitive skills and semantic association skills was perhaps not that surprising, given the emerging evidence for unconscious retrieval of sentence frame gestalts suggested by the construction grammar approach to expressive language analysis (Bruns et al., 2019). That said, assessment performance and therapy performance may differ here, if therapy shows itself to be more explicitly reliant on cognitive skills such as executive functioning. It is well established that attention-executive skills do play a predictive role in anomia therapy outcomes (Lambon Ralph et al., 2010). In moving from our assessment study presented in this chapter, to our therapy study presented in the next, we will be afforded the opportunity to test this and other hypotheses.

CHAPTER 4 Evaluating a novel hybrid sentence production therapy in a case series of English-speaking participants with Aphasia.

4.1 Introduction

Sentence production deficits are common and debilitating in aphasia, and many treatment approaches have been developed to address them (Boyle, 2017; Faroqi-Shah and Baker, 2017; Marshall, 2017). Despite the preponderance of lexical therapy studies in aphasia, it is well recognized that therapy focused at the single word level does not typically generate generalized changes in the language processing system (Best et al., 2013). Generally, item-specific therapy responses are noted, which though encouraging, can lack functional utility for PWA in their language use (Palmer et al., 2019). In contrast, sentence production therapy seems to provide the opportunity for more generalized gains in language processing, as suggested by the range of therapy approaches and therapy reports described in the literature review in Chapter 2. Furthermore, consistent with the approach taken in Chapter 3 of surveying sentence processing across the aphasia severity range and subtypes, many more people with aphasia may benefit from sentence production therapy than has been suggested by the frequent focus on non-fluent aphasia evident in the literature. In Chapter 4, we have piloted a novel, multilevel sentence production therapy that may lead to functional gains in participants' expressive skills. At the end of Chapter 3, we concluded that some of the participants who had been screened were good candidates for sentence production therapy. The current study aims to test these predictions by applying the same sentence production therapy template across a heterogeneous sample of participants with aphasia, a subset of the sample of participants in chapter 3. In doing so, we will compare the relative benefits of the novel sentence production therapy both within and between these participants.

The following is the list of criteria incorporated in the design of our novel therapy protocol for sentence therapy in aphasia. The therapy protocol should:

a. Deliver cost-effective means for a high dosage of therapy, with time efficiency.

The high intensity of therapy is commonly associated with better therapy gains (Basso, 2005; Bhogal et al., 2003; Cherney et al., 2011; Lee et al., 2009; Robey, 1998). However, the optimum amount, frequency, and duration to induce and maintain improvement remains a topic of debate in the literature (Dignam et al., 2016). Nevertheless, in our study, we will aim to explore the minimum amount of practice and time required to start noticing statistically significant gains in language skills across the participants.

b. Applicable to participants with a wide range of language and cognitive skills.

It includes participants with severe impairment in sentence production, those who communicate in 2-3 word combinations, and those who experience breakdowns in assigning thematic/grammatical roles within a sentence (Schwartz et al., 1987); also, participants with fluent aphasia who produce adequate simple sentences but may struggle with more complex structures or fail to combine them effectively in discourse.

c. Foster autonomy in language practice and promote self-monitoring and self-correction skills.

The aim is to promote the internalization of cueing strategies and, ultimately, the generalization of therapy gains to functional everyday communication.

d. Compatible with languages other than English.

As one of the wider goals of this thesis is to develop an Arabic version of the therapy approach and test its efficacy, the selected approaches and computer software needs to be language neutral.

In the current intervention design we have adopted the integrative model, described by Hinckley (2017), in combining three individual impairment-based approaches. Interventions that are categorised as "impairment-based" are those targeting the underlying cognitive linguistic functions ((De Bleser and Papathanasiou, 2003) rather than the consequences of the impairment on activities and life participation (i.e., classified as "participation-based" intervention (Hinckley, 2017)). Studies that have implemented a similar integrative model in aphasia intervention showed promising results. For example, Frederiksen and White (1989) demonstrated that integrating whole-task with part-task training achieved superior outcomes in terms of shorter skill acquisition time and greater skill transfer to untrained contexts. Accordingly, we combined the principles of VNeST (Edmonds et al., 2009), mapping therapy (Schwartz et al., 1994) and temporal window widening (i.e., through 'SentenceShaper', a cognitive processing prosthesis) (Linebarger et al., 2007) and delivered the therapy in a hierarchical, multileveled format. The aim was to gain maximum potential outcome, including direct and indirect therapy gains, within the least amount of time.

The multilevel structure of the therapy protocol involves targeting linguistic levels sequentially and in an increasing hierarchy of linguistic and cognitive demands. A growing body of evidence in the literature support the effectiveness of multilevel approaches targeting a number of language levels simultaneously in treating language production deficits in aphasia (Goral and Kempler, 2009; Milman et al., 2014; Whitworth et al., 2015). Moreover, the multiple assessment time-points research design will be implemented in our study, scheduled in between therapy levels and before increasing the complexity of the targeted sentence (e.g., Level-1 targets simple SVO sentence followed by Level-2 that targets sentence expansion, and then Level-3 that targets combining sentences in discourse) (see Table 4.4). The aim is to analyse the efficiency of the current protocol levels and determine if they equally added significant value to the outcome or one or more levels produced greater levels of change; the information is required for future refinement of the approach. The multilevel protocol structure with multiple testing time-points will also allow us to examine the interaction between levels and build predictions based on the interpretation of findings.

Considerations in selecting impairment-based methods

Poirier et al. (2021), conducted a systematic review of the literature on the efficacy of sentence production treatments in aphasia. The authors identified 11 different treatments presented in 25 studies with a combined sample of 84 PWA. The treatments were grouped into three main categories based on their therapy targets and the associated level of processing as illustrated in Bock and Levelt (1994b) language production model. The first group of treatments focus on verbs (functional level) and include: verb naming in isolation (e.g., semantic feature analysis, errorless training, and semantic cueing), verb naming with sentential complements (VNeST) (Edmonds et al., 2009), and multimodal approaches. The second group focus on sentence structures (functional and/or positional level- constituent assembly) and includes: Treatment of underlying forms (TUF) (Thompson and Shapiro, 2005), Mapping therapies, and conversational abilities (e.g., Intensive Language Action Therapy (ILAT) (Difrancesco et al., 2012), and Helm-Estabrooks Language Program for Syntax Stimulation (HELPSS) (Helm-Estabrooks and Ramsberger, 1986)). Finally, the third group focus on morphology (positional level- inflectional process) and includes: verb tense (e.g., Morphosemantic treatment (Faroqi-Shah, 2008) and Computerized Visual Communication (C-VIC) (Boser et al., 2000) and hierarchical approaches (Tree-Pruning Hypothesis TPH (Friedmann and Grodzinsky, 1997) with TUF like protocol). Overall, the results of the reviewed studies showed clear evidence of therapy-induced performance on trained items in 92% of the sample of participants. Also, generalization of therapy gains to untreated items (i.e., contains equivalent characteristics to the trained items in terms of number of arguments, frequency, type of sentence) was noted in 86% of the sample. Moreover, inter-modality generalization from oral sentence production to oral sentence comprehension was tested in a subgroup of 25 participants and 64% of the sample showed improvement. Transfer of therapy gains to discourse was analysed mainly through story retell tasks and the analysed variables included: mean length of utterances, number of grammatically correct sentences, number of correct information units, complexity of sentences, and number of open- and closed-class words. Accordingly, 73% of a sample of 59 participants showed therapy-induced improvements on those narrative discourse measures. Finally, retention of therapy gains was observed in the majority of participants after the discontinuation of direct therapy.

Of all the reviewed sentence therapy interventions above, four main treatments were classified with ‘strong’ methodological quality based on the Single-Case Experimental Design (SCED) ratings (L Tate et al., 2008): VNeST, TUF, MTs, and C-VIC (Poirier et al., 2021). The evidence further supported our selection of VNeST and MT as two of the main impairment-based approaches in our novel sentence therapy intervention. The reasons for eliminating the TUF and conversation-based therapies will be revisited shortly under the “considerations in selecting a computer-based method” subheading. Also, targeting morphology in our therapy design will be further explained under “novelty in adapting VNeST approach” subheading.

VNeST (Edmonds et al., 2009)

Edmonds (2016) conducted a literature review on all studies that implemented VNeST as an individual therapy approach. It included 5 studies and a pool of 19 English speaking participants with a range of aphasia types and severities. The findings showed preliminary evidence of the efficacy of this approach in improving lexical retrieval at the word, sentence, and discourse levels in English (Edmonds and Babb, 2011; Edmonds et al., 2014; Edmonds et al., 2009; Edmonds et al., 2015; Furnas and Edmonds, 2014) as well as across other languages (i.e., Korean) (Kwag et al., 2014). Therapy induced improvements were noted in both noun naming and verb naming in the majority of participants, 86% and 58% of the sample respectively, which was attributed to the semantic network activation induced by VNeST. In those studies, sentence production skills were measured by stimuli from the NAVS test (see Chapter 2) in a constrained sentence production task; however, the implemented scoring approach did not account for grammatical errors. With this criteria, 75% of the sample demonstrated improvements in sentence production skills on untrained items. On the other hand performance on discourse was measured by complete utterances CUs and revealed that 59% of the sample showed significant therapy gains (Edmonds, 2016).

It is worth noting that the VNeST therapy protocol presented in the literature, implemented a high dose of therapy with at least two sessions per week, totalling 3-3.5 hours per week (i.e., duration ranged between 4-15 weeks with the majority receiving 10 weeks of therapy) Edmonds (2016). Although a more recent single case study by Parkes (2017) examined a lower dose of delivery, 1.5 hours per week of clinician-direct therapy sessions (total number of hours was 4.5 hours), the outcomes did not replicate the positive findings published previously. Nevertheless, in the current study, we aim to test the outcome of incorporating VNeST in a lower dose and intensity, as a component of a specific hybrid multilevel therapy approach, to remediate sentence production deficits. We hypothesize that VNeST will complement other approaches when combined and delivered simultaneously.

Mapping therapy (Schwartz et al., 1994)

The mapping therapy studies encompass a group of highly diverse techniques, strategies, and outcome measures (Berndt and Mitchum, 1997; Byng, 1988; Byng et al., 1994; Dorze et al., 1991; Haendiges et al., 1996; Jones, 1986; Marshall et al., 1997; Mitchum et al., 1997b; Mitchum et al., 1995; Nickels et al., 1991; Rochon et al., 2005; Schwartz et al., 1994). A study by Rochon et al. (2005) targeted sentence production therapy is small sample of participants with chronic Broca's aphasia (n=3). In their study, the therapy dose was approximately 19 sessions of 1-hour biweekly sessions, delivered over the course of 6 months. The outcome measures encompassed: two constrained sentence production tests (the Caplan and Hanna's Sentence Production Test (Caplan and Hanna, 1998) and the Picture Description with Structure Modeling Test (Fink et al., 1995)) and the QPA discourse analysis method. The findings revealed improved production of trained canonical and non-canonical sentences as well as generalization of therapy gains to narrative production; however, improvements did not include untrained structures or cross-modality changes (e.g., sentence comprehension skills). Accordingly, we have chosen to include Mapping Therapy in our therapy design based on the long-standing evidence of its effectiveness in remediating sentence production and improving metalinguistic awareness which lasts even after the discontinuation of therapy (see Chapter 2 for further details).

<u>Considerations in selecting a computer-based method</u>

To achieve our aim of offering a method that can be implemented in high dosage, with cost-effectiveness, we have looked into computer-based aphasia therapy approaches, specifically those that target sentence and discourse production. Although several options existed at the time of planning the current study, few of them matched our criteria. SentenceShaper presented unique features that can be used as a platform to deliver our novel theory-driven impairment-based therapy protocol. The following is an overview of our justifications for choosing SentenceShaper over other methods, which we have reached after a thorough review of the literature (see Chapter 2).

AphasiaScripts™ contains useful elements that could serve the aims of our current study. These include the option of personalizing the scripts, customizing the level of cueing, and self-practice. Theoretically, with the author's permission, the program can be modified to include Arabic scripts and cues. However, a disadvantage of using it in our study would be the necessity of eliminating participants with more severe non-fluent aphasia. Those participants usually struggle to put two words together to form a phrase. Initiating a therapy program that requires them to formulate complete sentences in a conversational context without any preparation will be challenging. Also, since our sample is inclusive of PWA with a wide range of language skills, starting the therapy program at this linguistic level will pose different challenge levels across the sample. Doing so would indicate additional training sessions for some participants to prepare them for the task, but not the whole sample. This would contradict our purpose of delivering the same therapy program systematically across a sample of participants with a range of skills to examine the variation in their responses. Thus, to keep our inclusion criteria inclusive of participants with a wide range of language production skills, it was determined that the initial stage of the therapy program in our study should establish metalinguistic awareness. It can be achieved through tasks that focus on the basic form of sentences such as simple active sentences SVO, then gradually increasing the sentences' complexity in the next levels. Those tasks can be delivered through SentenceShaper® computer software which supported our selection.

Another intriguing option was Sentactics®, especially with the wealth of data in the literature supporting the Complexity Account of Treatment Efficacy CATE-Treatment of Underlying Forms TUF the software delivers (Thompson et al., 2010; Thompson et al., 2003). Nonetheless, applying TUF principles in Arabic will be linguistically challenging in light of the scarcity of literature investigating the spoken (colloquial) dialects. Adapting the TUF protocol to Arabic will require the efforts of linguists with a considerable background in Arabic linguistics and local dialects. The process of development would include identifying the sentence structure types that are most common in everyday verbal interactions and the rank of sentence structures in terms of complexity and relatedness. A consensus will also be needed to support the linguistic categorization of the sentence structures in Arabic colloquial dialects. Additionally, even if a permission was granted from the author to reprogramme the virtual clinician in the software to deliver cueing in Arabic, which is an integral element of the Sentactics®, the process will be complicated. Accordingly, it was concluded that Sentactics® would not be a practical option due to the project's time-frame restrictions.

SentenceShaper® was our choice for a core computer-based platform to deliver the impairment-based therapy protocol. It fosters autonomy, offers flexibility in editing and customising therapy tasks, and is compatible with languages other than English, including Arabic. Moreover, it adds a unique feature that will address verbal working memory limitations, one of the proposed causes for sentence production breakdown in aphasia (Linebarger et al., 2001), by alleviating time pressure in language practice; and providing optimal conditions for self-monitoring and self-correction skills. We anticipate that those skills will promote internalization of cueing strategies; and, therefore, the generalization of therapy skills to everyday functional communication.

A number of studies investigated the feasibility of SentenceShaper as a therapy tool (Linebarger et al., 2004; Linebarger et al., 2007; Linebarger et al., 2001; McCall et al., 2009). The intervention did not include direct instruction from a clinician but involved independent practice of story retelling and narrative production of personally relevant topics, movies, or television shows. The outcome measures included analysing the participants' spontaneous unaided verbal narratives using the Quantitative Production Analysis system QPA (Saffran et al., 1989). After at least 15 hours of home use, the following characteristics were observed: increased MLU, more structured utterances, improved grammatical well-formedness, and an increased proportion of words in sentences. Also, the produced narratives scored higher on informativeness measures. In 2009, a study by (McCall et al., 2009) explored the outcomes of implementing a more structured approach in targeting complex sentence production (subordinate clauses) through SentenceShaper. They found that this structured approach further advanced therapy gains as measured by words per sentence (e.g., words per sentence increased from 3.6 to 8.12 following the general therapy approach, then increased further to 11.56 words per sentence when syntactic structures were targeted in practice).

Although the majority of studies that tested the efficacy of SentenceShaper examined participants with non-fluent aphasia, not all participants presented with agrammatism. Linebarger and colleagues (2007) explained, "because the system works in such a general way, by supporting the retention and monitoring of speech, its impact may be interestingly heterogeneous across different kinds of language impairments" (Linebarger et al., 2007, p.55). Accordingly, incorporating SentenceShaper will fulfil one of the fundamental aims in our therapy design and create an approach that applies to a wide range of aphasia types and severities.

In our therapy design, we strived to maximize all possible therapy outcomes and target all areas that could benefit from the intervention. Therefore, with the cognitive processing support offered by SentenceShaper, we incorporated explicit language therapy methods to remediate sentence production impairments in PWA. It included VNeST and mapping therapy to increase metalinguistic awareness, activate semantic networks, and repair the mapping operations between thematic and grammatical roles in sentence processing in PWA. Also, we cost-effectively increased the dose of therapy through the use of technology.

The task presentation hierarchy in the therapy protocol starts with mapping therapy tasks and then followed by VNeST tasks. The mapping therapy tasks include picture prompts which could serve as a visual cue. On the other hand, VNeST tasks prompt the participants to produce sentences by generating scenarios from their own memories and experiences and use their own vocabulary to formulate them. The rationale here was that a task that requires self-generated responses without a visual aid would typically recruit different cognitive processes and be more challenging. Accordingly, we chose to first establish metalinguistic awareness with the mapping therapy approach, including pictures' support and then move to a more challenging task that requires additional independence and less need for cues.

Novelty in adapting VNeST approach

A computer-based delivery of VNeST method is available through the *Create* activity in the Advanced Naming Therapy app by Tactus Therapy Solutions Ltd (Tactus Therapy Solutions Ltd., 2021), which implements an adapted protocol of VNeST. However, to the best of our knowledge, all VNeST protocol versions do not include morphology/ verb inflections in the tasks of creating and reading the generated agent-verb-patient triads (Edmonds, 2014). In VNeST tutorial (Edmonds 2014, P.84), it was explained that the aim of intervention was to stimulate sufficient activation of content words to be included in a sentence structure; therefore, the addition of morphology/functors in training may detract the participant's focus from that aim. Nevertheless, the aim of our current therapy was not limited to the lexical level but extends to include repairing the processes involved in sentence production as described in language models (e.g., the morphophonolgical encoding stage in Levelt (1999) model). Although there is some variation between language models on when morphological encoding occurs, there is an agreement that those processes are to some extent distinct from lexical retrieval processes (Thompson et al., 2015). The design of our intervention includes training the production of morphology and function words in a sentence context, although it is not targeted directly. The main aim was to increase the participants' metalinguistic awareness (i.e., making linguistic rules conscious) through external support (e.g., a written scaffold, verbal instruction, non-verbal prompts to self-correct, etc.) and as needed to reach independence. The metalinguistic awareness approach was supported by evidence from studies on children with specific language impairment SLI which found that language repair, including the facilitation of complex sentence production, can be achieved through metalinguistic awareness (Hirschman, 2000).

Overall, we anticipate that the integration of the three (above described) impairment-based therapy approaches will be complementary, and each method will add its unique value to the combination. For example, mapping therapy is reported to repair and strengthen the mapping procedures between the thematic roles and grammatical roles; however, there is no evidence that it might induce semantic network strengthening that is linked to VNeST in the literature. On the other hand, unlike the mapping therapy, VNeST does not include grammar in training nor it is monitored by the outcome measures; therefore, it is unexpected to yield improvements in sentence grammaticality and well-formedness. Finally, SentenceShaper as a single approach does not target metalinguistic awareness that is offered by Mapping Therapy and VNeST and therefore cannot replace their benefits.

4.2 Research Questions/Study Aims

To this date, the outcomes of combining three principles in aphasia therapy, mapping therapy, Verb Network Strengthening Therapy, and SentenceShaper, in a tactically designed multilevel framework of delivery has not been investigated or reported in the literature. This study will examine the outcomes of a systematic application of a novel theory-driven computer-based approach to sentence production therapy across a varied and heterogeneous range of participants to answer the following research questions:

1. How feasible is this new approach in remediating lexical and sentence processing deficits in PWA?
2. Were treatment gains maintained once direct treatment was ended?
3. What is the minimum dose of therapy required to achieve noticeable improvements on constrained sentence production tasks?
4. How do patterns of treatment response vary across participants with varying baseline language and cognitive skills?
5. How do treatment outcomes compare across sentence level processing and other language and psycho-social measures?

4.3 Methods

The selected research design for this study was an experimental case-series with multiple assessment points.

4.3.1 Participants

A convenient non-selective sample was chosen for this study. Participants were recruited from the Neuroscience and Aphasia Research Unit NARU participant database at the University of Manchester (Butler et al., 2014) and Speakeasy specialist aphasia charity based in Ramsbottom, Manchester. The pool of potential participants encompassed most aphasia types, severities and varying levels of cognitive skills. We have adopted a pragmatic approach in recruitment by extending the invitations first to those who lived within close geographical proximity to the University of Manchester, the research site. Thirty-one participants responded to the participation invitation and underwent the initial language and cognitive screening to assess their candidacy. From this sample, 22 participants met the candidacy criteria; however, only 17 participants were interested in enrolling in the therapy program. Ultimately, 12 participants committed to the three-month course of therapy and completed the program as well as the maintenance testing (see Figure 20).

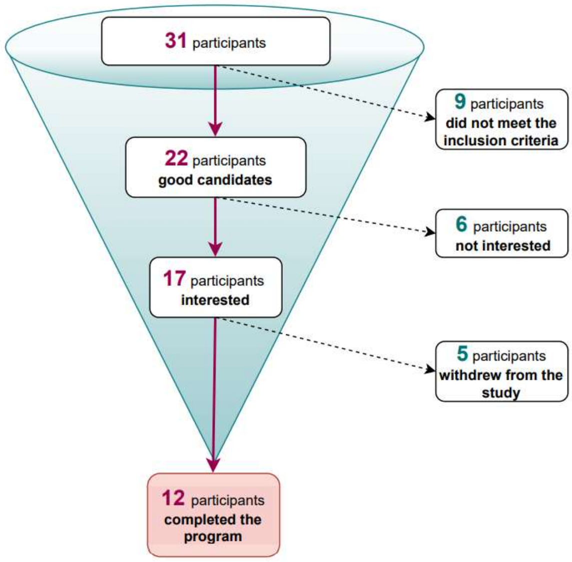


Figure 20 Recruitment process for the English therapy study

As outlined in Chapter 3, participants were recruited if they reported a history of left-hemisphere stroke and demonstrated characteristics of aphasia on the sentence construction task of the Verb and Sentence Test VAST (Bastiaanse et al., 2003) and/or the narrative construction task (i.e., cookie theft picture scene, and dinner party picture sequence). Due to the nature of the computerized therapy program used in the current study, basic computer skills and simple sentence reading comprehension (tested informally) were required for inclusion. The participant had to present an acceptable level of dexterity to independently operate and navigate the computer program using a computer mouse. Most importantly, the participants had to indicate their agreement to commit to 13 consecutive weekly visits to the university and adhere to the home practice program.

The exclusion criteria included signs of poor auditory comprehension of simple active SVO sentences determined by scoring below 45% accuracy on the VAST sentence comprehension task (Bastiaanse et al., 2003) and subjective clinical observation through informal conversations. The presence of apraxia, which is frequently associated with aphasia (McNeil et al., 2009), or dysarthria was not an indication for exclusion in this study unless it prevented the participant from producing any intelligible words in confrontational naming task (e.g., BNT) or informal conversations. Moreover, participants with a history of neurological disorders such as brain tumors, traumatic brain injuries, dementia, uncontrolled seizures etc., were excluded from the study. Similarly, poor corrected vision, and poor corrected hearing were indications for exclusion.

The sample of 12 participants who completed the program (5 females and 7 males) included 6 participants with fluent aphasia and 6 with non-fluent aphasia. It encompassed four aphasia subtypes across a range of severity levels: Anomia, Broca's Aphasia, Transcortical Motor Aphasia (TMA) and Conduction Aphasia, which was determined by the Boston Diagnostic Aphasia Examination BDAE (Goodglass and Kaplan, 1983) and the Western Aphasia Battery WAB (Kertesz, 1982). Time post-onset varied between participants, ranging from 31 to 140 months (Mean= 90.25, SD= 32.39). The age of participants ranged from 51 to 81 years old (Mean=66.08, SD= 8.82). Years of education ranged from 10 to 17 years (Mean=12.36, SD= 2.46). All participants were native English speakers. They reported normal or corrected-to-normal hearing and/or vision at the time of testing. Informed consent was taken from all contributing participants with the National Research Ethics Committee's approval, REC reference: 01/8/094.

Table 4.1 Participants' demographic information arranged according to their performance on BNT test.

Participant No.	Initials	Age (years)	Gender	Years of education	Time-post onset (months)	BDAE + WAB aphasia classification	BNT score (%)*
1	MH	71	M	11	56	Conduction	8
2	RH	69	M	17	51	Conduction	10
3	DA	51	F	11	88	Broca's	25
4	RR	63	M	13	94	Broca's	33
5	AB	54	M	13	127	Anomia	52
6	AD	81	F	11	113	Broca's	55
7	ST	68	F	11	78	Broca's	57
8	PR	76	F	11	96	TMA	58
9	JP	71	F	11	31	Anomia	58
10	GP	64	M	11	91	Anomia	67
11	WE	68	M	10	118	Anomia	70
12	DM	57	M	17	140	Broca's	92
	Mean	66		12	90		49
	SD	9		2	32		25

*The average score of neuro-typical adults is 54/60 (Nicholas and Brookshire, 1993; Van Gorp et al., 1986). The cutoff scores for unimpaired performance (two standard deviations below the mean) based on a sample of 60 subjects was 48 (Nicholas and Brookshire, 1993).

4.3.2 Assessment

4.3.2.1 Background neuropsychological assessment Data

The background data were retrieved from the NARU aphasia database, which included an extensive neuropsychological battery that examined language and cognitive skills of a wide range of participants with history of a single left hemisphere stroke at the chronic stage (described in Butler et al., 2014). For the current study, a number of measures were selected to represent an overview of the participants' language and cognitive skills before the intervention. It included data of their performance on the Boston Diagnostic Aphasia Examination BDAE short version (Goodglass and Kaplan, 1983), Light Verb Elicitation Test LVET (Carragher et al., 2013), the Raven's Coloured Progressive Matrices (Raven et al., 1962), Psycholinguistic assessments of language processing in aphasia PALPA (Kay et al., 1996), 96 trial synonym judgment test (Jefferies et al., 2009), the spoken picture description subtest (i.e., a picture description task) of the Comprehensive Aphasia Test CAT (Swinburn et al., 2004), and the Wechsler digit span ‘forward and backward digit span’ (Wechsler, 1987) (see Appendix 6 for an example of the task).

Table 4.2 Background data retrieved from NARU aphasia database of neuropsychological battery (percentage score)

Participant's No. and initials		BNT	LVET	Ravens	PALPA 8*	96 synonym**	CAT spoken	digit span forwards	digit span backwards
1	GP	57	57	97	43	90	78	38	29
2	WE	55	80	92	47	88	84	63	43
3	DM	72	0	92	60	96	56	38	0
4	JP	57	57	83	70	83	N/A	50	29
5	RH	2	40	83	3	90	63	25	29
6	MH	5	0	81	0	70	59	25	29
7	PR	38	27	81	57	83	88	75	0
8	ST	50	23	47	53	-	72	25	0
9	AD	50	70	64	23	82	69	75	43
10	AB	42	0	89	27	75	75	38	29
11	RR	23	0	89	10	82	56	25	29
Average		41.0	32.2	81.6	35.8	83.9	70.0	43.2	23.4
	SD	22.3	30.4	14.4	24.3	7.6	11.3	19.7	16.0

* Auditory word repetition ** 96-trial synonym judgment task

4.3.2.2 Baseline and outcome measures

Primary outcome measures

The therapy protocol in the current study is composed of tasks that directly target sentence construction and sentence expansion skills in addition to tasks that encourage combining sentences in discourse. Accordingly, the main assessment tools we have selected to measure therapy-induced changes in performance includes a constrained sentence production task (VAST sentence construction subtest) and a discourse elicitation task (cookie theft and dinner party picture stimulus). We were also interested in capturing any trends in performance on the primary therapy goals over the course of treatment. Therefore, the primary outcome measures were repeated at several time-points on pre-determined intervals (see Table 4.3 for more details). Since the outcomes of the VAST sentence production subtest will be key to our assessment of the feasibility of this novel therapy approach, we modified the scoring approach to include more in-depth systematic analysis than what has been offered in the original scoring guideline. It involved developing purpose-designed novel scoring protocol, presented in Appendix 1.

Within the assessment tools we incorporated 6 subtests from the VAST test battery. However, only one subtest, the VAST sentence production task, measured a primary outcome target in our study. For the purpose of continuity, we will be introducing the complete VAST test battery first under the primary outcome measures and then refer back to it when the 5 subtests are reintroduced as components of secondary outcome measures.

a. Verb and Sentence Test VAST (Bastiaanse et al., 2003)

The VAST is an adaptation of the Dutch test Werkwoorden-en Zinnentest: WEZT (Bastiaanse et al., 2000). The modification included the omission and addition of some items to the English version. Both versions have been standardized and checked for reliability and validity. Bastiaanse and colleagues (2003) reported high internal consistency for tasks that included 20 items and more, which supported the reliability of the test. The VAST validity has been demonstrated by comparing the Token Test from the Aachen Aphasia Test on the measure of severity of aphasia (Miller et al., 2000). Overall, the VAST was designed to supplement clinicians with a linguistically motivated diagnostic tool to investigate verb and sentence processing.

The VAST assessment contains ten subtests in total, distributed into three independent scoring sheets. Scoring sheet A encompass three comprehension subtests: verb comprehension, sentence comprehension, and grammaticality judgment task. Scoring sheet B focuses on testing production skills in four subtests: action naming, filling-in finite verbs in sentences, filling-in infinitives in sentences, and sentence construction. Scoring sheet C examines grammatical abilities in two sentence-anagrams subtests, one with picture stimuli and the other without picture stimuli.

However, to focus on outcomes related to the current intervention and maintain a tolerable duration of testing, only seven subtests were selected from the VAST. In each trial, a correct response scores 1 point, and the maximum potential score for each subtest equals the number of items. The VAST test norms were acquired from 80 neuro-typical native English speakers, and the majority of their scores reached ceiling levels (see Table i) (Bastiaanse et al., 2003).

Table i The control's (n=80) range of scores on each VAST subtest *

VAST scoring sheet	subtest	Control's range of scores	Maximum score
A	verb comprehension	38-40	40
	sentence comprehension	35-40	40
	grammaticality judgment	37-40	40
B	action naming	37-40	40
	filling-in finite verbs in sentences	8-10	10
	filling-in infinitive verbs in sentences	8-10	10
	sentence construction	16-20	20

*(Bastiaanse et al., 2003)

The sentence construction subtest, central to our study, contains a picture description task that examines sentence production skills in 20 trials with 20 different verbs. The authors explained, "The performance on this test reflects the ability to produce sentences in daily life.

The results on the other tests will help evaluate the nature of the difficulties that people with aphasia may experience when they are producing sentences" (Bastiaanse et al., 2002, p.14).

All VAST subtests start with verbal instruction and a demonstration by the clinician, in addition to 1-4 practice exercises (i.e., the number of practice items varies in each subtest) to confirm understanding of the task. To ensure consistency, the examiner adhered to the scripts listed in the VAST manual in delivering the task instructions. All trials included presenting a picture stimulus except the grammaticality judgment task, which relied only on a verbal stimulus. The picture stimulus was presented in a field of 4 for the comprehension tasks and in a field of 1 for the production tasks. The participants were asked to indicate their responses by pointing to the target picture in the comprehension tasks and verbally produce their answers in the production tasks.

The primary researcher administered all the baseline and interim testing as well as the scoring. Instrumental bias was avoided by adhering to the test manual's scoring instructions (Bastiaanse et al., 2002) except for two subtests. A novel scoring rubric was developed to closely examine the VAST sentence construction subtest findings (See Appendix 1). Also, since the assessment of morphology production in a cloze-sentence context was not within the current study's aims, the scoring guidelines for the two fill-in verbs in sentences subtests (i.e., fill-in finite verbs and fill-in infinitives) were modified. The two subtests were collapsed into one to increase the number of trials, from 10 items each to 20 items. For this subtest, any accurate verb production would receive a full score (i.e., 1 point per item) regardless of the accuracy of the morphology produced. For example, if the participant produced 'crawling' instead of the target 'crawl' in the following sentence: ' the child want to….to his mum', it would get the full mark for this item(i.e., 1 point).

Finally, the participants declared no prior knowledge of the tests; therefore, learning effects were eliminated. Also, to avoid experimental fatigue, the testing duration was limited to a maximum of two hours per session.

- Data collection procedure for the sentence construction subtest (VAST) main outcome measure

The VAST sentence construction subtest was the main outcome measure for monitoring therapy-induced changes throughout the therapy course, which included a 12-week course of therapy (60-90 minutes/week clinician delivered therapy sessions and 3-4 hours/week independent home practice). Before starting therapy, the participants' performance was determined by averaging two baseline scores collected at least one month apart and one-week before starting therapy. On the other hand, post-therapy testing was conducted no later than one week after completing the therapy program. Interim assessments at two intervals were also conducted at the 4th and 8th week of therapy. The same VAST sentence construction subtest items (i.e., a fixed set of 20 different verbs) were used in all assessments, which targeted sentence production in a picture description task. The pictures and sentences in the test were reserved for assessment and not included in therapy. Only independent, verbal, and intelligible productions were recorded, and no cues were given. A novel scoring protocol was implemented to score and analyse the productions (see Appendix 1). This guideline accounted for the minimum required sentence constituents (i.e., subject, verb, object, function words) to produce complete sentences describing the picture stimuli in the test (e.g., the girl throws a stick). Nevertheless, the therapy program targeted sentence expansion and combining sentences in a discourse, which goes beyond the minimum required production accounted for in the current scoring protocol (e.g., the girl throws the stick into the river). To capture those changes, we included an additional outcome measure to analyse discourse samples.

b. Discourse samples

Murray and Coppens (2013) highlighted the fact that structured tasks in most aphasia assessments examine language skills in a decontextualized manner (e.g., providing discreet levels of linguistic analysis at the level of phonology, lexical retrieval or constrained sentence production), which is not necessarily representative of the participant's skills in discourse production. Moreover, the discourse sampling tasks in most test batteries are either brief or absent. Accordingly, we supplemented the selected outcome measures with discourse tasks to provide a more comprehensive overview of language skills in PWA. It is reported in the literature that discourse analysis is sensitive enough to distinguish even the subtle changes in participants' performance over time (Hussmann et al., 2012), which may be missed by formal language assessments. Therefore, it represents a useful measure to our current study to track therapy-induced changes throughout the course of therapy, especially in people with milder forms of aphasia that scored near ceiling on structured tasks but still demonstrated signs of communication breakdown in unstructured conversations.

Two discourse samples were collected from the participants at each testing time-point (see Table 4.3) using two different elicitation tasks: the 'Cookie Theft' complex scene picture description (Goodglass and Kaplan, 1983), which prompts descriptive discourse production, and the 'Dinner Party' picture sequence storytelling (Mark et al., 1983) which prompts a narrative production.

The quantitative analysis of the discourse samples has been conducted using the Systematic Analysis of Language Transcripts SALT computer software program (Miller and Chapman, 1983). Holland and colleagues (1985) reported a successful application of SALT in analyzing discourse samples of PWA. We anticipated that using computer software for the analysis will provide consistency when the elicitation protocols and transcription format are implemented; therefore, allowing us to conduct within-participant and across-participant comparisons. Furthermore, the SALT procedures' reliability has been documented in the literature (Heilmann et al., 2008).

The speech samples' coding adhered to the SALT coding guideline (Salt Software LLC., 2020) and was conducted by an undergraduate student majoring in Speech and Language Therapy with experience working as an interpreter. The coded transcripts were then reviewed and edited by a trained specialist from the SALT Software company. Next, all SALT default variables were calculated for each transcript independently and results were displayed as a rectangular data file RDF. This data was then imported into an excel spreadsheet for review and interpretation (See Table 4.7 and Table 4.8). Finally, a subset of variables were selected for interpretation as can be seen in the results section. One informative variable that is calculated by SALT software is the "analysis set utterances" which reflects all "complete (not abandoned or interrupted), intelligible (do not contain any unintelligible segments), and verbal (do not consist entirely of non-verbal elements such as nods and shrugs)" (Salt Software LLC., 2015a) in the discourse. Nevertheless, the software does not distinguish between relevant and non-relevant utterances such as circumlocutions/non-topic related comments. Accordingly, as a final step to identify utterances that were off topic and excluding them from the total score, a code was created and implemented throughout the transcripts. This step was accomplished by selecting the "insert code" tab from the "edit" list. Then, from the pop-up window a code was selected from the list and labelled [OT] and described as "off target utterance". The [OT] code was then inserted at the end of each "off target utterance" by the primary investigator. Only utterances that were irrelevant to the events of the picture scene (i.e., cookie theft and dinner party) were identified as off topic (e.g., she has got a penny). However, utterances that were leading to a relevant information were not marked. Also, the code [OT] identified utterances, not words. A similar analysis that would identify relevant vs. irrelevant content words from the total number of words produced in a discourse was not available. Creating this special feature within SALT would require a much more complex process which will include reaching a consensus on the core content words in a given discourse and teaching the software to identify those words in a discourse. Nevertheless, with the current [OT] code, the irrelevant words within an utterance will warrant the exclusion of the whole utterance. Therefore, the analysis set of utterances will not reflect those off-topics words and utterances.

In SALT's default software settings, both "analysis set utterances" and "total number of words" variables do not distinguish between relevant and non-relevant words in a given discourse. Although we customised the "analysis set utterances" variable setting for this study to include only relevant sentences, a similar modification to the "total number of words" variable is a much more complex task. We excluded that option as it would require manual assessment of all words produced by the participants in our sample across multiple time points and compare the data to controls to reach a consensus on the relevance of each word, which was not available. Nevertheless, we have chosen to still use the data generated by the default software settings for the "total number of words" to get an insight on possible fluctuations in fluency associated with therapy-induced changes in sentence production skills. For example, in fluent aphasia a decrease in 'total number of words' that is associated with an increase or stability in 'analysis set utterances' scores may indicate better accuracy in word retrieval and diminishing of circumlocution and empty speech. On the other hand, in non-fluent aphasia an increase in 'total number of words' that is associated with an increase in 'analysis set utterances' scores may reflect improvements in content word retrieval and sentence production skills in discourse.

Likewise, the mean length of utterance in words MLU variable (i.e., within SALT default setting and calculated based on "analysis set utterances" only) was selected to get an estimate of the participants' progress across time in sentence construction and expansion which was directly targeted in therapy (i.e., Level 2 within the therapy protocol).

Secondary outcome measures

a. Verb and Sentence Test VAST (Bastiaanse et al., 2003)

The secondary outcome measures included 5 subtests from the VAST test battery (described above): verb comprehension, sentence comprehension, grammaticality judgment task, action naming and fill-in verbs in sentences (i.e., combines filling-in finite verbs in sentences and filling-in infinitives in sentences).The original scoring guidelines were followed in scoring 4 subtests, with the exception of the fill-in verbs in sentences subtest that required a slight modification as explained previously.

b. Light Verb Elicitation Test LVET (Carragher et al., 2013)

The light verb elicitation test examines the retrieval of light verbs in constrained fill-in verbs in a sentence task (see details outlined in Chapter 3).

It has been suggested in the literature that syntactic impairment in PWA is specifically associated with impaired light verb retrieval skills (Barde et al., 2006; Gordon and Dell, 2003; Thorne and Faroqi-Shah, 2016); however, the number of studies that investigated the topic remains small and further research is needed to reach solid conclusions. In the current study, we are interested in exploring any therapy-induced changes in performance on LVET task to find out whether a correlation can be found between performance on sentence production tasks and on light verb retrieval tasks.

c. BNT Boston Naming Test BNT (Goodglass and Kaplan, 1983)

The BNT is a confrontational object naming task (see the details outlined in Chapter 3). In the current study we were interested in exploring the predictors for therapy outcomes, a function that BNT is recognised for in number of studies (Conroy et al., 2009; Lambon Ralph et al., 2010).

A simple modification has been made to the scoring guideline implemented in Chapter 3. In the current study, the time-limit for producing a response has been extended from 10 seconds (implemented in Chapter 3) to 60 seconds. Through clinical observation in Chapter 3, we noticed that the participants' low scores were based on their word retrieval speed in confrontational naming, and when given the time their accuracy rate increased. In the current study, Chapter 4, the aim was to measure the participants' ability to retrieve a target word and not the speed of retrieval, since the therapy approach does not factor-in time restrictions. Moreover, the confrontational naming tasks in the VAST (action naming) and WAB (object naming) does not set a time-limit for responses. Therefore, this modification enabled us to test the participants' word retrieval ability in a context similar to the therapy tasks' context and to match the scoring criteria with the rest of the confrontational naming subtests within the selected outcome measures used in the study. An additional modification to the scoring guideline involved providing a single prompt to the participants to produce a more specific response or use another word to describe the object (for example, if a participant produced "plant" for "flower" or "boat" for "canoe").

d. Western Aphasia Battery WAB-R (Kertesz, 2007)

The WAB is a language assessment battery designed to identify the presence of aphasia and classify the symptoms into one of the classic aphasia syndromes (e.g., Broca's, Anomia, etc.). It examines several linguistic skills (e.g., lexical retrieval) and communication modalities (e.g., reading, listening, speaking, writing) through tasks and stimuli of various complexity levels.

In 2018, the Research Outcome Measurement in Aphasia (ROMA) consensus statement was issued, and it contained a recommendation for a core outcome set (COS) for use in aphasia treatment studies (Wallace et al., 2018b). The benefits include: " (1) increased transparency and reliability of research findings through the recommended reporting of a minimum set of outcomes; (2) production of compatible research data which can be efficiently synthesised in subsequent meta-analyses; and (3) reduced research wastage through the measurement of relevant outcomes which are more likely to inform treatment decision-making" (Wallace et al., 2018a, p.241). Accordingly, we decided to include the WAB as a secondary outcome measure. However, since the report was issued mid-trials, only half of our sample took part in the WAB testing. Additionally, the WAB contains subtests that overlap with other outcome measures in the current study, such as the WAB object naming subtest and the BNT. This would afford us the opportunity to check the stability of therapy gains in a specific domain, such as word retrieval skills, across tests and thus strengthen the reliability of the outcome.

The test battery can be divided into two parts with two independent record forms. Part one includes the following subtests: spontaneous speech assessment (tasks: a. conversational questions, b. picture scene description), auditory verbal comprehension (tasks: a. yes/no questions, b. auditory word recognition, c. sequential commands), repetition, naming and word finding (tasks: a. object naming, b. word fluency, c. sentence completion, d. responsive speech). The score summary sheet summarises the total score in each subtest and an overall Aphasia Quotient (AQ) that could be used to estimate aphasia severity. The summary also includes a table that illustrates the WAB-R aphasia classification criteria.

Part two of the test is labelled supplemental. It includes the following subtests: 1. reading (tasks: a. comprehension of sentences, b. reading commands, c. written word-object choice matching, d. written word-picture choice matching, e. picture-written word choice matching, f. spoken word-written word choice matching, g. letter discrimination, h. spelled word recognition, and i. spelling), 2. writing (tasks: a. writing upon request, b. writing output, c. writing to dictation, d. writing dictated words, e. alphabet and numbers, f. dictated letters and numbers, and g. copying a sentence), 3. apraxia, 4. constructional, visuospatial, and calculation (tasks: a. drawing, b. block design, c. calculation, d. Raven's Coloured Progressive Matrices (RCPM), 4. Supplemental writing and reading (tasks: a. writing irregular words to dictation, b. writing non-words to dictation, c. reading irregular words, d. reading non-words). Part two form also includes a score summary worksheet that highlights the total score of each subtest. It produces a Language Quotient (LQ) and a Cortical Quotient (CQ).

Shewan & Kertesz (1980) reported excellent reliability characteristics of the WAB for both internal consistency and temporal stability. Additionally, high inter-rater and intra-rater reliability score were reported. In terms of validity, "the WAB's content is similar to other aphasia tests and correlates highly with the NCCEA, satisfying content and construct-validity criteria"(Shewan and Kertesz, 1980, p.323), referring to Neurosensory Center Comprehensive Examination for Aphasia NCCEA (Spreen and Benton, 1969).

The scoring procedure has been modified for the current study. Each subtest item was scored either 1 point for a correct response or 0 points for a wrong response. Therefore, the total score for each subtest will be equal to the total number of items, as the following: yes/no questions (20 points), auditory word recognition(60 points), sequential commands (11 points), repetition (15 points), object naming (20 points), word fluency(20 points), sentence completion (5 points), and responsive speech (5 points).

e. The Communication Outcomes after Stroke (COAST) (Long et al., 2008)

There are several assessment tools available to measure communication effectiveness post-stroke, such as the Functional Assessment of Communication Skills for Adults ASHA-FACS (Frattali et al., 1995), Therapy Outcome Measure TOM (Enderby et al., 2013), the Stroke Impact Scale (Duncan et al., 1999), Burden of Stroke Scale BOSS (Doyle et al., 2007), the Communication Effectiveness Index CETI (Lomas et al., 1989), and the Stroke and Aphasia

Quality of Life Scale SAQOL-39 (Hilari et al., 2003). However, most of them share a number of limitations, such as the reliance on an observer's input instead of PWA's own perspective, a narrow focus on artificial communication tasks, or lengthy and time-consuming to complete.

The Communication Outcomes after Stroke (COAST) (Long et al., 2008) is distinguished as a patient-reported outcome measure (PROM) which is reasonably brief to administer. It is a scale that measures self-perceived communication effectiveness in people with communication problems after stroke. It contains 20 items presented individually as a written text at the top of the page with a picture presented in the page centre and a graded response scale at the bottom of the page. According to Long and colleagues (2008), this assessment showed evidence of "good reliability and validity with some supportive evidence of construct validity" (Long et al., 2008, p.1089).

The interviewer administered the assessment, in which task instructions are presented verbally and followed by one practice trial to ensure the participant's understanding. Thereafter, in each trial, the test administrator reads aloud each item written at the top of the page, presented to the particiant to follow, with the key words highlighted. The participant is then asked to indicate their rating by pointing to one of the boxes in the 5-point scale. According to the test manual, non-verbal support from the interviewer is permitted as well as rephrasing from an alternative phrase bank provided in the manual, to insure consistency.

The COAST was selected as an outcome measure for the current intervention study based on the assumption that all enrolled participants can comprehend simple sentences and questions with visual cues since it is one of the inclusion criteria. Therefore, an outcome measure that documents the PWAs' perspective on their everyday communication effectiveness before and after the intervention was found to be a valuable addition to the current study. The following is an example of Item 4 within the COAST.

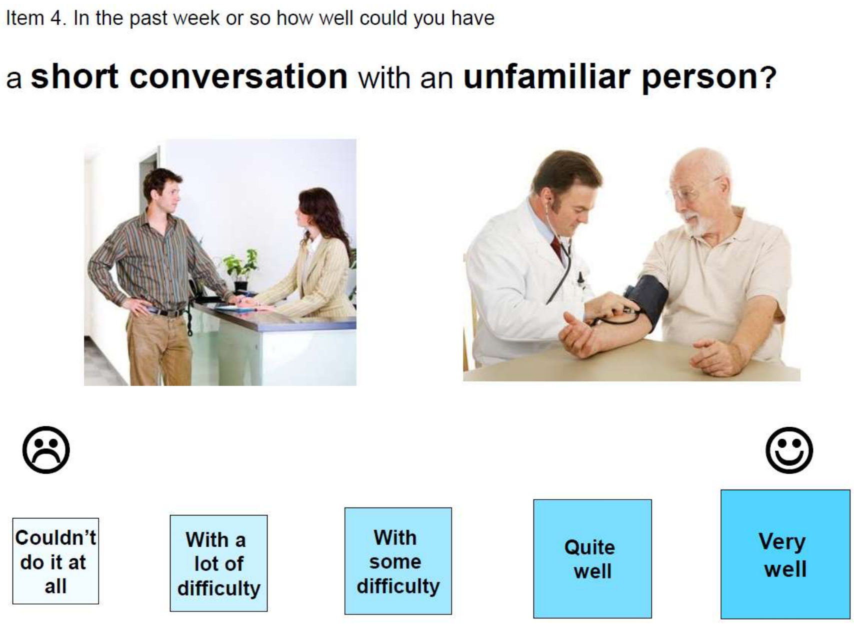


Figure 21 Example of an item in the COAST (Long et al., 2008)

Table 4.3 Testing frequency

	Test	Baseline-1 V1	Baseline-1 V2	Baseline-2	Interim 1	Interim 2	Interim 3	Maintenance
1	VAST (Sentence construction task)	√	√	√	√	√	√	√
2	VAST (scoring sheet A and B)	√	-	-	-	-	√	√
3	LVET	√	-	-	-	-	√	-
4	SentenceShaper performance data	-	√	-	√	√	√	-
5	Narrative samples	-	√	√	√	√	√	√
6	COAST	-	√	-	-	-	√	-
7	BNT	√	-	-	-	-	√	√
8	WAB	√	-	-	-	-	-	√

(√) administered at this time point (-) not administered at this time point

Baseline 1: testing was completed at least 2 weeks before the start of the therapy program.V1: first visit

V2: second visit. Baseline 2: beginning of the first therapy session and before the start of the therapy tasks. Interim 1: beginning of the fifth therapy session and before the start of the therapy tasks. Interim 2: beginning of the ninth therapy session and before the start of the therapy tasks. Interim 3: thirteenth session, no therapy tasks were administered. Maintenance: six weeks post-therapy.

4.3.3 Therapy protocol and procedure

The therapy protocol includes various tasks designed to address a range of therapy objectives and meet the therapeutic requirements of a diverse group of participants with aphasia. For clarity and to present the tasks in an increasing hierarchy of difficulty, the therapy tasks were distributed into three levels. Level 1 targets simple SVO sentence; Level 2 targets sentence expansion; and Level 3 targets combining sentences in discourse. Each level is further divided into two phases: Phase A involves a picture description task, and Phase B involves a task that prompts self-generated sentences without picture stimuli.

Level 1 targets the production of simple active SVO sentences in two phases (phase A and phase B). **Phase A** involves a picture description task and implements the mapping therapy approach to establish/enhance metalinguistic awareness. The initial stages of training involve presenting a picture scene on the SentenceShaper screen (see Figure 23) and asking the participant to answer Wh-questions, presented verbally by the clinician, to elicit the subject, verb, and object. The clinician presents a written scaffold (see Figure 22) to the participant and performs online transcription of each word the participant produces, and prompts the production of any missing words by pointing to the blank space in the scaffold until a complete sentence is accomplished. Then, it is followed with Wh-questions presented verbally by the clinician (e.g, who pushed the box? What is the dog doing? What is being pushed?) for the second time. The purpose is to emphasize the relation between of thematic roles and the grammatic roles (i.e., who did what, to whom). Next, using SentenceShaper, the participant is instructed to record each word and order them in the sentence row (see Figure 23) using the written scaffold as a guide. Then, the participant is asked to playback the whole sentence and pay attention to any possible errors. If errors were noted, the participant is prompted to edit them (e.g., delete, add, rerecord, etc.,). Lastly, the participant is asked to verbally produce the whole sentence three times (without recording). For further details on the recording and editing process on SentenceShaper and the level of support provided by the clinician see Appendix 3.

This high level of support will then be gradually faded until the participant reaches independence in verbally producing and recording the target sentences with no cues. At this stage, the participant is asked to record their complete sentence in one recording (one snippet instead of individual snippets for each word) and listen to the playback to self-assess the accuracy of their production before moving to the next item.

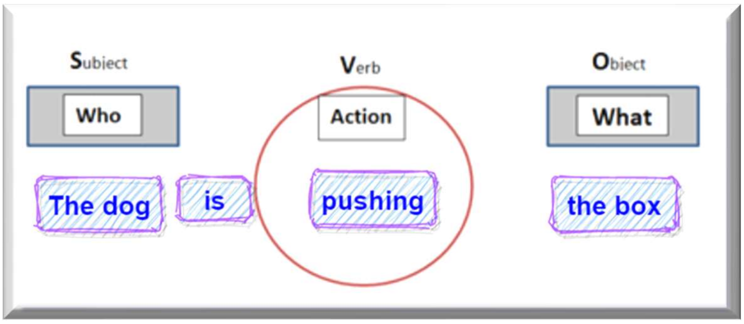


Figure 22 Example of the written scaffold used in Level 1

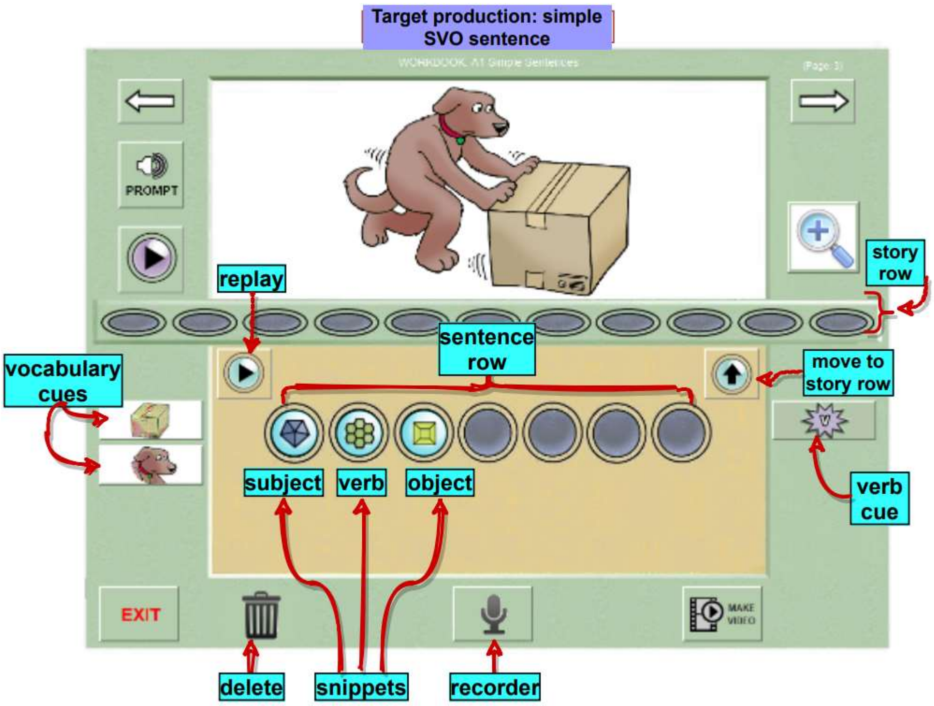


Figure 23 Screenshot of SentenceShaper page in Level 1 – Phase A

Phase B is based on VNeST principles. The scaffold in this phase is similar to the one presented in Phase A, but the task is different. In this task, the picture stimulus is removed and replaced with a written text of the target verb on the screen (see Figure 24). The goal is to prompt the participant to generate a subject and an object around the verb by relying on their own vocabulary repertoire and personal memory of events to generate the scenarios. For example, participant JP, whose hobby is gardening, when given the verb "push", produced the following sentence: "The gardener pushes a lawnmower in the field". In this phase, the participant is instructed to generate three different sentences (i.e., presenting three different scenarios) using three different subjects and objects of each target verb. For example, the verb "push", the three different sentences could be: the mother is pushing a pram, the mechanic is pushing the car, and the dog is pushing a box. The purpose is to stimulate a widespread semantic activation of verb networks.

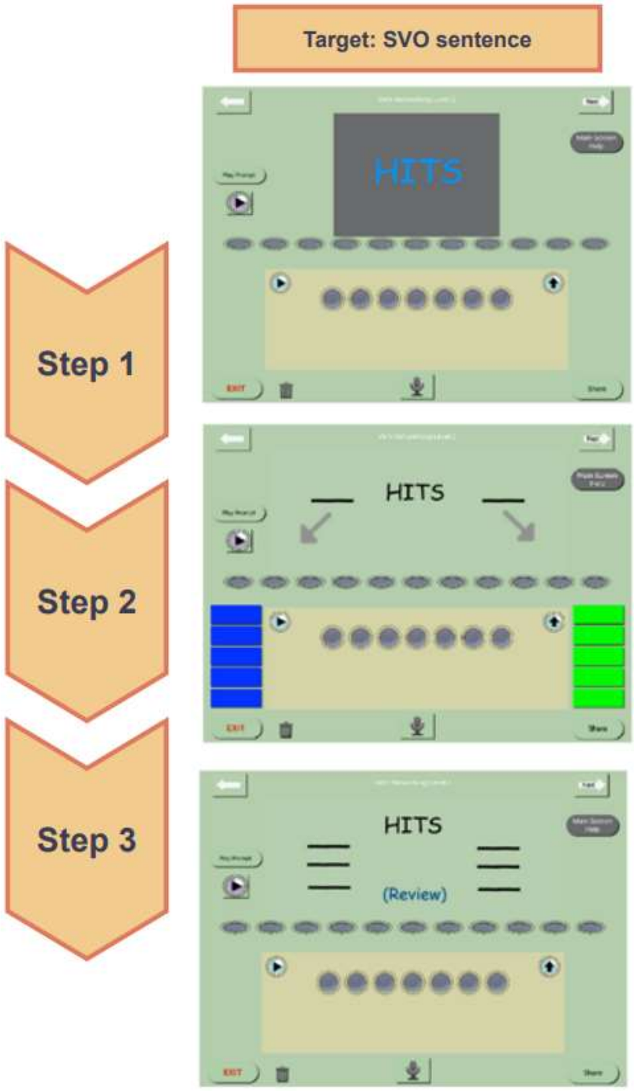


Figure 24 Example of VNeST task in Level-1 Phase-B,

Level 2 focus on promoting elaboration and sentence expansion. It refers to the process of adding one or more words, phrases, or clauses to the main clause. The same scaffold framework is used in this level, as in Level 1, but with some modification to add slots for adjectives and/or a prepositional phrase (see Figure 25). The same previously introduced steps were implemented to elicit the subject, verb, and object, with additional Wh-question to encourage sentence expansion and elaboration (e.g., how, why, when, where). Similar to Level 1, **Phase A** includes picture stimuli (see Figure 26), and **Phase B** includes a written text of the target verb with a visual scaffold to guide the sentence construction (see Figure 27).

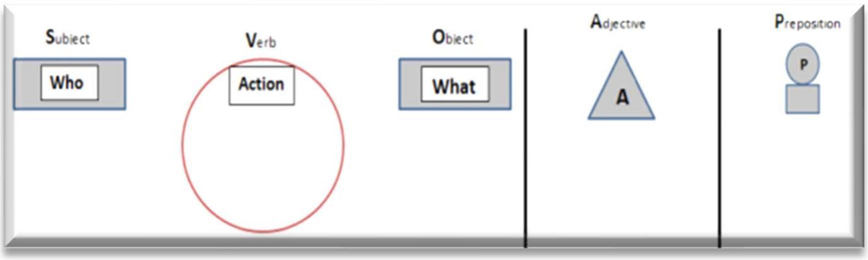


Figure 25 Example of the written scaffold used in Level 2

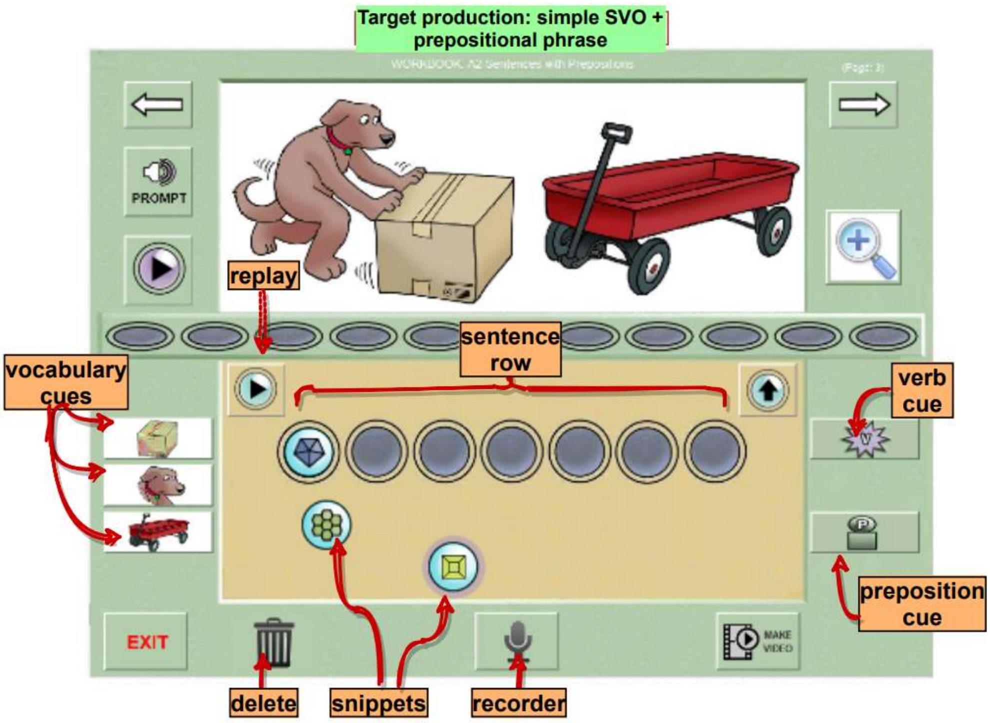


Figure 26 Example of the picture stimulus in level 2- phase A

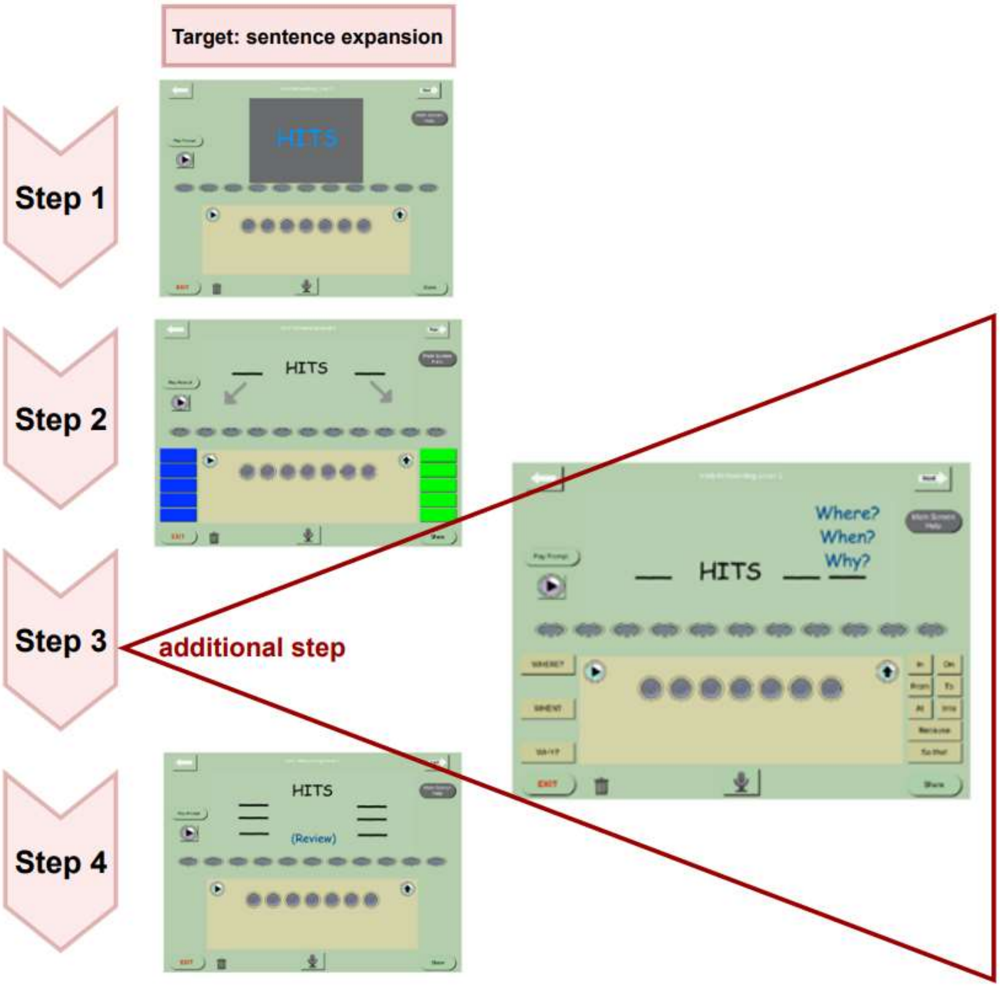


Figure 27 Example of VNeST task in Level-2 Phase-B,

Level 3 targets multi-sentence narrative construction. **Phase A** requires the participant to generate a narrative in a picture sequence task (i.e., workbook D in SentenceShaper). Phase B includes an open-ended questions task (e.g., retelling movies plots or talking about personal live events) with no visual prompts.

Table 4.4 Levels of therapy delivered over the time course of the program

ACTIVITY	ACTUAL START	ACTUAL DURATION	WEEKS 1	2	3	4	5	6	7	8	9	10	11	12
Level 1 - Phase A	1	2												
Level 1 - Phase B	3	2												
Level 2 - Phase A	5	2												
Level 2 - Phase B	7	2												
Level 3 - Phase A	9	2												
Level 3 - Phase B	11	2												

Time duration determined the participants' progress from one level to the next, as illustrated in Table 4.4 above. However, the success rate at each level will determine the amount of support provided to each participant. For example, participant RH required written cues (i.e., the initial letter of the target word) to overcome his word-finding difficulty and produce a simple SVO sentence (in Level-1 Phase-A), with a success rate of less than 50% accuracy for his independent productions. By the time two weeks of the therapy program had passed, he was automatically moved to the next level (Level-1 Phase-B), as had the rest of the participants. However, he was still provided with the same level of support he required to succeed (i.e., written cues) and as needed. The cueing type (i.e., phonemic, gesture, written, etc.) and level (i.e., immediate, delayed, etc.) was tailored to each participant and was faded gradually to foster internalization of these cueing strategies and reach independence (see Appendix 3 for more details).

In summary, the therapy program was comprised of three levels, with each level divided into two phases. The two phases in all three levels share a similar framework; the first phase always provides a structured task with picture stimuli, while the second phase prompts self-generated productions (i.e., sentence or discourse) with no picture stimuli (see Figure 28). However, the written scaffold and the Wh-questions were constant elements in all levels and phases, and unlike cueing, they do not fade according to the participants' progress or independence level.

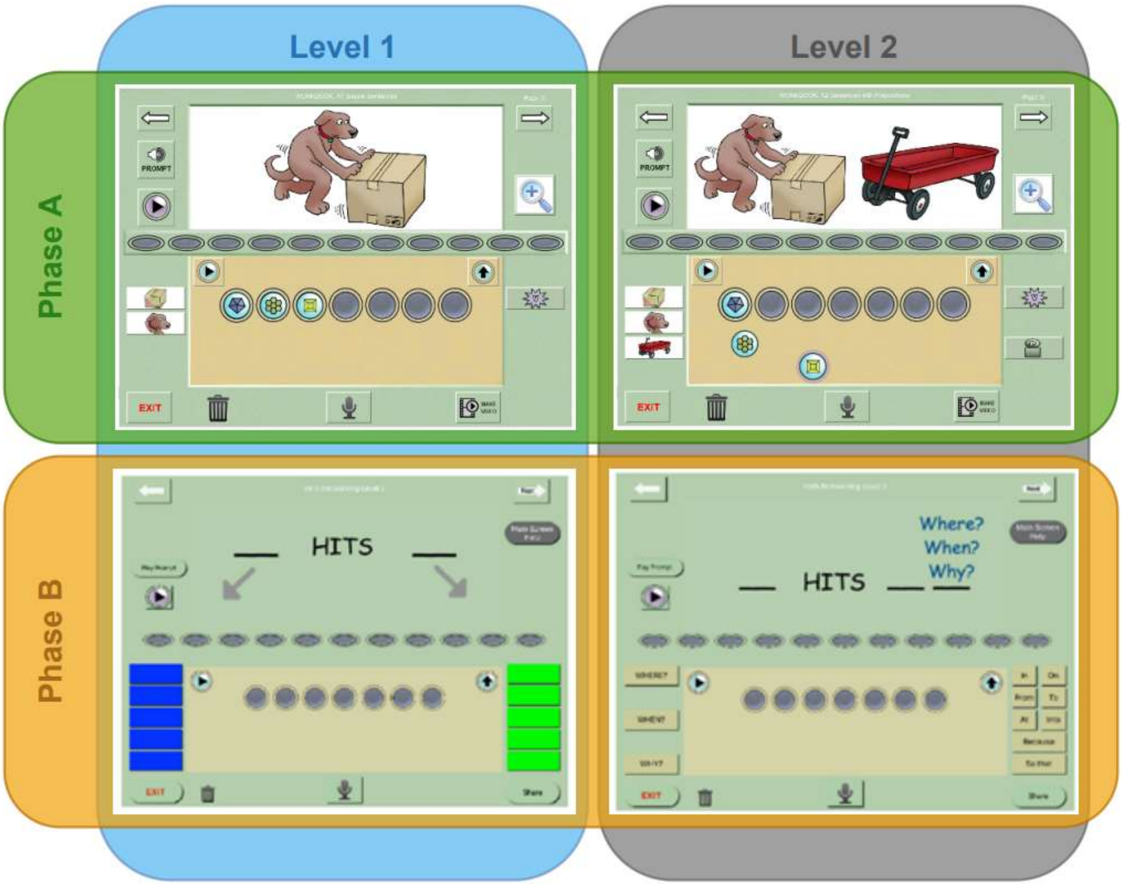


Figure 28 Examples of the display page on SentenceShaper screen

4.4 Results

The Results have been set out in order of the research questions identified in the study aims.

4.4.1 Research question 1: How feasible was this new approach in remediating lexical and sentence processing deficits in PWA?

In this context, we will measure the feasibility of the current approach according to its ability to: 1. induce statistically significant direct therapy gains in sentence production and discourse. 2. induce statistically significant indirect therapy gains in linguistically related (untargeted) skills. 3. maintain gains after the discontinuation of therapy. 4. improve self-perceived communication effectiveness.

Additional factors will be considered within the feasibility assessment, which will be further explained in the discussion section of this chapter. It include: 1. the usability and learnability of the therapy protocol. 2. participants' acceptance of the computer-based tool, their level of independence, and adherence to home practice requirements. 3. study retention rate.

a) Primary outcome measure – sentence production and discourse

I. Sentence production

- Description of individual change

With respect to therapy-related changes in outcome measures, comparing pre- versus post-treatment time points, Table 4.5 presents the participants' scores on the VAST battery production subtests. The participants were ordered in a descending hierarchy according to their baseline scores on the VAST sentence construction subtest.

As a group, the mean score on the VAST sentence construction subtest showed 20% therapy-induced increase. To evaluate the sample's sentence production scores compared to baseline, a cut-off score was set at 70% accuracy (represents the group's average score 14/20) on their baseline scores on the sentence construction subtest. This divided the sample into two subgroups, 6 participants that scored at or below 70% (ST, RH, AB, DA, AD, RR, with respective baseline scores of 14,13,11,11,10,9), and 6 participants with baseline scores above 70% (JP, WE, DM, GP, MH, PR, with respective scores of 18,18,17,17,16,15). The subgroup of participants who performed at or below 70% accuracy for sentence construction showed an average score increase of 5.3 points (raw score), 25% increase post-therapy. In comparison, the subgroup that performed above 70% at baseline showed an average score increase of 2 points (raw score),10% increase only post-therapy. Therefore, the subgroup of participants with a lower starting point in baseline performance demonstrated superior therapy gains. Noteworthy to mention and that the subgroup that scored above 70% at baseline and achieved more modest therapy gains were mainly composed of participants with fluent aphasia (4 participants with fluent aphasia: JP, WE, GP, MH; and 2 participants with non-fluent aphasia: DM, PR). On the other hand, the subgroup that performed at or below 70% at baseline and had superior therapy gains were mainly composed of participants with non-fluent aphasia (4 participants with non-fluent aphasia: ST, DA, AD, RR; and 2 participants with fluent aphasia: RH, AB). This may reflect a weak tendency for the non-fluent participants to be somewhat more severe for sentence construction at baseline (mean non-fluent baseline= 12.6; mean fluent baseline=15.5, though the range of scores overlapped).

Table 4.5 Participants' performance on VAST production subtests before and after therapy

Initials		Verb production		Fill-in verbs in sentences		Sentence construction		Total production	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Max. score		40	40	20	20	20	20	80	80
1	JP	22	25	13	15	18	20	53	60
2	WE	35	37	19	18	18	19	72	74
3	DM	39	39	18	18	17	19	74	76
4	GP	32	30	14	18	17	18	63	66
5	MH	8	8	6	10	16	19	30	37
6	PR	32	26	9	11	15	18	56	55
7	ST	31	30	11	15	14	19	56	64
8	RH	16	24	11	15	13	18	40	57
9	AB	20	20	5	10	11	15	36	45
10	DA	12	16	9	12	11	17	32	45
11	AD	23	31	16	17	10	14	49	62
12	RR	22	23	11	11	9	17	42	51
Mean (SD)		24(10)	26(9)	12(4)	14(3)	14(3)	18(2)	50(15)	58(12)

- Statistical analysis of individual participants

To examine within-subject statistical significance of results ($p< 0.05$) of therapy-induced changes in sentence production scores, obtained through the novel scoring system (see Appendix 1), we conducted Wilcoxon matched pairs test. As shown in

Table 4.6, 10 out of 12 participants (83% of the sample) presented statistically significant changes in their sentence production skills post therapy.

Table 4.6 Individual participant statistics for significance (p<0.05) of therapy gains on the VAST sentence construction subtest using Wilcoxon matched pairs test of ordinal data.

Participant No. and initials		Pre-therapy mean (SD)	Post-therapy mean (SD)	Wilcoxon Z	One tail *p*
1	JP	6.75 (1.52)	7.3 (1.34)	1.86	0.0311
2	WE	7 (1.31)	7.45 (1.15)	0.96	0.1696
3	DM	6.6 (0.84)	7.75 (1.07)	3.68	0.0001
4	GP	6.83 (1.33)	7.35(1.27)	1.82	0.0341
5	MH	6.2 (1.74)	7.15 (1.50)	1.51	0.0659
6	PR	5.75 (1.81)	7.3 (1.66)	2.75	0.0030
7	ST	5.5 (1.70)	7.4 (1.27)	3.46	0.0003
8	RH	5.38 (1.63)	7.1 (1.68)	3.31	0.0005
9	AB	4.4 (1.62)	5.8 (1.54)	2.44	0.0073
10	DA	4.2 (1.24)	6.9 (1.41)	3.91	0.0000
11	AD	3.98 (1.67)	5.75 (2.09)	2.71	0.0033
12	RR	3.58 (1.50)	6.5 (1.99)	3.80	0.0001

Note: data in bold are significant at the p<0.05 level. The maximum mean of scores for the sentence construction task is 8.1 with a standard deviation of 1.02 (i.e., due to the variety of maximum scores for each item).

- Statistical analysis of change across participants

To examine across-subjects statistical significance ($p < 0.05$) of therapy-induced changes in scores, Wilcoxon matched pair test of ordinal data was conducted. The results showed that changes in scores on the VAST sentence production subtest were statistically significant (z=-3.03, n=12, p=0.0012, one-tailed).

II. Discourse

In this section, we will assess therapy gains generalisation to discourse production. The scores were collected at three time-points using the cookie theft (Goodglass and Kaplan, 1983) and the dinner party (Kertesz, 2007) picture stimuli. The systematic analysis of language transcripts (SALT) software was used to analyse the speech samples. From the SALT analysis variables, the following 3 variables were selected based on being most clinically meaningful and relevant to the therapy delivered: analysis set utterances (i.e., SALT's default analysis set is a subset of the total utterances in which utterances are complete, intelligible, and verbal), MLU (mean length of utterance) in words, and the number of total words. The scores on the selected variables were calculated for each transcript independently, and results were displayed as a rectangular data file RDF. This data was then imported into an excel spreadsheet for review and interpretation (See Table 4.7 and Table 4.8)

In the current study, 50% of the sample (n=6) underwent two baseline assessments. For this group, their performance before therapy intervention was determined by the average score of baseline 1 and baseline 2.

We examined therapy-induced changes by analysing the participant' production of the cookie theft (CT) and dinner party (DP) discourse using three measures generated by the SALT software at three time-points (i.e., baseline, interim 3, and maintenance). The outcome measures included analysis set utterances (refers to a subset of the discourse that includes only complete grammatically correct utterances and excludes mazes, false starts and repetitions.), MLU, and the number of total words. Given the high variability both within and between neurotypical controls and participants with aphasia, we have presented these data in descriptive terms rather than statistical analyses. To identify substantial changes, we focused on the percentage of increase in scores compared to baseline. Accordingly, we implemented the following formula to calculate the percentage increase percentage: (Interim3 / Baseline) x 100 and (Maintenance/ Baseline) x 100. The criterion for substantial improvement was set at 20% and above increase in scores compared to baseline.

To analyse the groups' performance, we examined changes in the mean score. First, the analysis set utterances variable showed a reduction of 12% in the mean score on the cookie theft discourse. However, the MLU showed an increase of 25%, while the number of total words remained roughly the same. Those findings were detected immediately post-therapy (interim 3) and were roughly maintained 6 weeks post the discontinuation of direct therapy. On the dinner party discourse, more favourable results were noted on all three measures. The average score showed an increase of 10%, 20%, and 28% on the analysis set utterances, MLU, and the number of total words, respectively. Similarly, the scores were generally maintained 6 weeks post discontinuation of direct therapy, with minimal 2-8% variation.

Next, we examined individual performance on the same measures. On the cookie theft discourse, the analysis set utterances measure revealed that five participants showed substantial (>20%) improvement: AD (72%), DA (300%), PR (100%), RR (20%), and ST (58%). On the other hand, participant AB did not show such a substantial increase in score immediately post-therapy but still showed noticeable gains at maintenance (22%). Interestingly, 4 out of 5 participants of the same group showed consistent therapy gains across tests on the same variable. They showed substantial improvement on the dinner party discourse at interim 3: AD (58%), DA (68%), PR (100%), and RR (36%), all classified with non-fluent aphasia. On the other hand, participant MH showed improvement on the dinner party discourse (83%) but not the cookie theft discourse. Comparably, participant ST showed noticeable improvement on the cookie theft (57%) discourse but not on the dinner party discourse.

Next, we analysed the changes in Mean Length of Utterances MLU in words from the analysis set of utterances generated by SALT. The outcomes of the cookie theft discourse showed that 9 out of 12 participants demonstrated noticeable improvements at either interim 3 time-point (6 participants: AD 115%, JP 36%, MH 26%, PR %51, RH 34%, and RR 123%) or at maintenance (3 participants: DM 33%, GP 50%, and ST 23%). Five out of this group of nine showed consistent gains in MLU on the dinner party discourse (AD 67%, DM 52%, GP 34%, RH 21%, and ST 40%). On the contrary, 4 participants (JP, MH, PR, and RR) demonstrated improvements on a single discourse, the cookie theft discourse but not the dinner party. Likewise, 2 participants (AB 57% and WE 21%) demonstrated improvements in the dinner party discourse but not the cookie theft discourse.

Finally, we examined the number of total words from the analysis set of utterances. For the cookie theft discourse, 6 out of 12 participants showed noticeable gains at interim3 or Maintenance: AB 26%, AD 272%, DA 50%, PR 207%, RR 169%, ST 80%. Only 5 of those

6 participants showed consistent improvements on the dinner party discourse: AD 162%, DA 73%, PR 114%, RR 23%, and ST 48%. Interestingly, those participants were all classified with non-fluent aphasia. On the other hand, participant AB showed gains on the cookie theft discourse only (26%), while a group of 5 participants showed gains on the dinner party discourse only: DM 54%, MH 96%, RH 21%, and WE 26%.

Table 4.7 Cookie theft discourse analysis

Analysis Set Utterances	AB	AD	DA	DM	GP	JP	MH	PR	RH	RR	ST	WE	Mean	SD
Baseline	23	7	0.5	9	14	8	12.5	4	29	5	7	6	10.58	8.50
Interim 3	22	12	2	4	9	7	9	8	18	6	11	4	9.3	5.8
% change from baseline	-4	**+71**	**+300**	-56	-36	-13	-28	**+100**	-38	**+20**	**+57**	-33	-12	
Maintenance	28	9	6	5	7	9	11	6	19	7	8	6	10.1	6.8
% change from baseline	**+22**	**+29**	**+1100**	-44	-50	13	-12	**+50**	-34	**+40**	14	0	-5	

MLU in Words*	AB	AD	DA	DM	GP	JP	MH	PR	RH	RR	ST	WE	Mean	SD
Baseline	5.3	2.6	3.5	4.8	7.7	8.3	6.45	3.3	6.4	2.6	4.7	8	5.30	2.07
Interim 3	6	5.6	3	5	8.7	11.3	8.1	5	8.6	5.8	4.8	7	6.6	2.3
% change from baseline	13	**+115**	-14	4	13	**+36**	**+26**	**+52**	**+34**	**+123**	2	-13	**+25**	
Maintenance	6	2.6	3	6.4	11.6	8.2	6.6	6.2	6.8	5.1	5.8	8.8	6.4	2.4
% change from baseline	13	0	-14	**+33**	**+51**	-1	2	**+88**	6	**+96**	**+23**	10	**+21**	

Number Total Words*	AB	AD	DA	DM	GP	JP	MH	PR	RH	RR	ST	WE	Mean	SD
Baseline	133.5	18	4	43	108	66.5	81	13	185	13	29.5	48	61.88	55.97
Interim 3	132	67	6	20	78	79	73	40	154	35	53	28	63.8	44.1
	-1	**+272**	**+50**	-53	-28	19	-10	**+208**	-17	**+169**	**+80**	-42	3	
Maintenance	168	23	18	32	81	74	72	37	130	36	46	53	64.2	45.1
	+26	**+28**	**+350**	-26	-25	11	-11	**+185**	-30	**+177**	**+56**	10	4	

*of analysis set utterances

Table 4.8 Dinner party discourse analysis

Analysis Set Utterances	AB	AD	DA	DM	GP	JP	MH	PR	RH	RR	ST	WE	Mean	SD
Baseline	41	12	12.5	16	23	28	18	8	58	12.5	16	21	22.5	14.8
Interim 3	34	19	21	16	21	26	33	16	58	17	17	18	24.7	12.2
% change from baseline	-17	**+58**	**+68**	0	-9	-7	**+83**	**+100**	0	**+36**	6	-14	10	-17
Maintenance	40	21	30	13	20	22	26	17	60	17	10	23	25.3	14
% change from baseline	-2	**+75**	**+140**	-19	-13	-21	**+44**	**+113**	3	**+36**	-38	10	12	-2

MLU in Words*	AB	AD	DA	DM	GP	JP	MH	PR	RH	RR	ST	WE	Mean	SD
Baseline	5.35	2.4	3.4	3.3	7.5	6.9	7.25	3.6	5.7	3.85	3	7.1	4.9	1.9
Interim 3	8.4	4	3.4	5	7.7	7.7	7.6	3.9	6.9	3.5	4.2	8.6	5.9	2.1
% change from baseline	**+57**	**+67**	0	**+52**	3	12	5	8	**+21**	-9	**+40**	**+21**	**+20**	
Maintenance	5.7	3.4	2.2	6.2	10	7	6.8	4.2	6.4	3.4	5.8	8.2	5.8	2.2
% change from baseline	7	**+42**	-35	**+88**	**+33**	1	-6	17	12	-12	**+93**	15	18	

Number Total Words*	AB	AD	DA	DM	GP	JP	MH	PR	RH	RR	ST	WE	Mean	SD
Baseline	239.5	29	41	52	172	192.5	128.5	29	331	48	48	150	121.7	97.9
Interim 3	284	76	71	80	161	199	252	62	400	59	71	155	155.8	109.5
% change from baseline	19	**+162**	**+73**	**+54**	-6	3	**+96**	**+114**	**+21**	**+23**	**+48**	3	**+28**	
Maintenance	249	71	65	80	199	154	176	71	382	58	58	189	146	99.8
% change from baseline	4	**+145**	**+59**	**+54**	16	-20	**+37**	**+145**	15	**+21**	**+21**	**+26**	**+20**	

*of analysis set utterances

b) Secondary outcome measures – other VAST subtests, BNT, LVET and WAB

Within the secondary outcome measures, 5 subtests of the VAST were administered to collect data from a sample of 12 participants, while the WAB was used to collect data from 6 participants only. On the other hand, both BNT and LVET tests are composed of a single task. Therefore, to simplify the presentation of outcomes, we grouped the data obtained by the secondary outcome measures into two sections: I. other VAST subtests. II. BNT, LVET and WAB.

I. Other VAST subtests

- Description of individual change

As can been seen from Table 4.5, with respect to 'Total Production' scores and descriptive statistics, 11 of the 12 participants showed numerical improvement in performance post-therapy with an average increase of 8 points (raw score) (50/80 pre-therapy to 58/80 post-therapy) and a range of 2-17 points (raw score), while 1 participant (PR) showed a decrement of 1 point (56 – 55/80). Across the three sub-tests which contributed to the sum score of 'Total Production' (verb production, fill-in verbs in sentences, sentence construction), there was a mean of 5% gain for verb production (24-26/40), a mean of 10% gain in fill-in verbs in sentences (12-14/20), and a mean of 20% gain in sentence construction (14-18/20). This may have reflected the majority of treatment focus on the sentence rather than lexical therapy in the multilevel treatment program.

Looking next at the VAST comprehension scores, Table 4.9 presents the raw scores for the three comprehension subtests (verb comprehension, sentence comprehension, and grammaticality judgement) and the sum of these subtests 'Total comprehension'. The mean subtest scores showed group increases in performance across 2 of the 3 subtests (verb comprehension, sentence comprehension) of modest proportions (5% and 2.5% post-therapy gains, respectively), while the grammaticality judgement task showed stable post-therapy performance. The total comprehension scores showed an overall modest change of 2.5%.

Table 4.9 Participants' performance on VAST comprehension subtests before and after therapy:

Participant No. and Initials		Verb comprehension		Sentence comprehension		Grammaticality judgment		Total comprehension	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Maximum score		40	40	40	40	40	40	120	120
1	JP	37	36	21	29	30	29	88	94
2	WE	35	40	37	37	36	38	108	115
3	DM	39	40	17	17	25	28	81	85
4	GP	39	39	22	29	32	31	93	99
5	MH	32	35	27	29	35	32	94	96
6	PR	27	35	37	34	34	37	98	106
7	ST	38	38	25	21	30	30	93	89
8	RH	37	40	20	21	33	32	90	93
9	AB	40	38	27	29	28	28	95	95
10	DA	27	34	22	18	30	24	79	76
11	AD	33	35	25	30	36	35	94	100
12	RR	40	37	20	22	20	22	80	81
Mean(SD)		35(5)	37(2)	25(6)	26(6)	31(5)	31(5)	91(8)	94(11)

- Statistical analysis of individual participants

To examine the statistical significance of the results ($p< 0.05$), we conducted McNemar tests of nominal data (correct/incorrect). As shown in Table 4.10, the most frequent indirect therapy gain is noted on the verb comprehension subtest, with 3 out of 12 participants (WE, PR, and DA) performing significantly better post-therapy. The next frequent subtest to show significant indirect therapy gains was the verb production subtest, with 2 participants (RH and AD) scoring significantly higher post-therapy. On the other hand, a significant improvement on the sentence comprehension subtest was limited to one participant (JP).

Table 4.10 Individual participant statistics details for significance (p<0.05) of therapy gains on the VAST subtests using McNemar's test.

#	Initials	Verb comprehension	Sentence comprehension	Grammaticality judgment	Verb production	Fill-in verbs in sentences
	p=	one-tailed	one-tailed	one-tailed	one-tailed	one-tailed
1	JP	0.5000	**0.0384**	0.5000	0.3036	0.3125
2	WE	**0.0313**	0.6875	0.3125	0.2500	0.5000
3	DM	0.5000	0.6230	0.3036	0.7500	0.7500
4	GP	0.7500	0.0592	0.5000	0.3438	**0.0625**
5	MH	0.2266	0.4018	0.2266	0.6367	0.1094
6	PR	**0.0039**	0.2266	0.1875	**0.0898**	0.3770
7	ST	0.6875	0.1938	0.6875	0.5000	0.1445
8	RH	0.1250	0.5000	0.5000	**0.0107**	0.1094
9	AB	0.2500	0.3770	0.6230	0.5982	**0.0625**
10	DA	**0.0592**	0.2120	**0.0730**	0.1719	0.2539
11	AD	0.3125	0.1133	0.5000	**0.0384**	0.5000
12	RR	0.1250	0.4073	0.3872	0.5000	0.6875

Note: data in bold and highlighted in grey were significant at the p<0.05 level. Data in bold only approached significance (<0.1).

- Statistical analysis of change across participants

To examine the statistical significance (p< 0.05) of therapy gains across participants on VAST subtests we administered Wilcoxon matched pairs test of ordinal data. The findings showed that significant results were limited to fill-in verbs in sentences subtest only (z=2.62, n=12, p=0.004, one-tailed). However, it approached significance on the verb comprehension subtest (z=1.64, n=12, p=0.051, one-tailed). The other three subtests did not show statistically significant changes in scores: sentence comprehension (z=0.92, n=12, p=0.1787, one-tailed), grammaticality judgment task (z=0.00, n=12, p=0.5000, one-tailed) and verb production (z=1.19, n=12, p=0.1175, one-tailed).

II. BNT, LVET and WAB

- Description of individual change

As can been seen from Table 4.11 (below), 83% of the sample of participants (10 of the 12) showed numerical improvement in performance on BNT post-therapy with an average increase of 3 points (raw score) and a range of 1-8 points (raw score), while two participants (RH and DA) showed no change in performance. On the other hand, only 50% of the sample of participants (6 out of 12) showed improvement on the LVET test that ranged between 1 to 6 points (raw score), while 3 participants showed no change (DM,DA and RR) and 3 participants showed a decrement of 1 point (JP and RH)or 3 points (AD)

On the Western Aphasia Battery (WAB) only 6 participants underwent pre- and post-therapy testing as it was added to the testing battery mid-trials (see methods section for further details); Table 4.12 displays their raw scores. The samples' average score on each comprehension subtest, showed a modest decline that ranged between 0.2-1.0 points: yes/no questions (18.5 to 18.3, SD= 1.9 to 1.2), auditory word recognition(54.5 to 54.3, SD= 5.8 to 5.9), sequential commands (7.2 to 6.2, SD= 2.5 to 0.8). On the other hand, the samples' average score on the production subtests showed an increase that ranged between 0.4 and 1.2 points: repetition (7.8 to 8.2, SD= 2.9 to 2.9), object naming (14.0 to 15.2, SD= 6.6 to 4.6), word fluency (6.3 to 7.3, SD= 3.9 to 4.3), and responsive speech (3.2 to 3.8, SD= 1.3 to 1.2). Nonetheless, a single subtest showed a decline of 0.5 points post-therapy compared to baseline: sentence completion (3.5 to 3.0, SD= 1.5 to 1.4). As a group, the participants' performance on the WAB subtests (both comprehension and production) showed stability in performance, or a slight change limited to 1 or 2 points, increase or decrease.

However, more noticeable changes were observed in individual performances. For example, participant MH showed an increase of 5 points in 2 out of 8 subtests (auditory word recognition and object naming); DA showed an increase of 3 points in performance on 1 out of 8 subtests (yes/no questions); and RR score's increased 3 and 4 points on 2 out of 8 subtests (object naming and word fluency, respectively). On the other hand, ST showed a decline of 4 points in 2 out of 8 subtests (sequential commands and object naming). Likewise, AB's scores declined by 4 points on 1 out of 8 subtests (sequential commands).

Table 4.11 Participants' performance on BNT and LVET tests (raw scores) before and after therapy

Participant No. and Initials		BNT		LVET	
		Pre-	Post-	Pre-	Post-
Maximum score		60	60	30	30
1	JP	35	37	17	16
2	WE	42	46	24	27
3	DM	55	57	0	0
4	GP	40	43	17	19
5	MH	5	13	0	6
6	PR	35	42	9	13
7	ST	34	35	7	10
8	RH	6	6	12	11
9	AB	31	33	0	1
10	DA	15	15	1	1
11	AD	33	37	21	18
12	RR	20	25	0	0
Mean(SD)		29(15)	32(15)	9(9)	10(9)

Table 4.12 Participants' performance on WAB subtests (raw scores) before and after therapy

Participant No. and initials		Yes/No Qs		auditory word recognition		sequential commands		repetition		Object naming		word fluency		sentence completion		responsive speech	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Maximum score		20	20	60	60	11	11	15	15	20	20	20	20	5	5	5	5
1	JP	19	20	60	60	9	7	11	12	19	19	13	15	5	5	5	5
2	MH	20	19	50	55	6	7	8	8	2	7	5	3	4	4	2	3
3	ST	19	17	56	55	9	5	10	10	20	16	6	5	5	3	4	4
4	AB	20	19	57	55	10	6	9	9	17	19	7	9	3	3	4	5
5	DA	15	18	45	43	4	6	3	4	12	13	6	7	1	1	2	2
6	RR	18	17	59	58	5	6	6	6	14	17	1	5	3	2	2	4
mean		18.5	18.3	54.5	54.3	7.2	6.2	7.8	8.2	14.0	15.2	6.3	7.3	3.5	3.0	3.2	3.8
SD		1.9	1.2	5.8	5.9	2.5	0.8	2.9	2.9	6.6	4.6	3.9	4.3	1.5	1.4	1.3	1.2

- Statistical analysis of individual participants

To examine within-subject statistical significance of BNT and LVET results ($p < 0.05$), we conducted McNemar tests of nominal data (correct/incorrect). As shown in Table **4.13** (below), improvements on the BNT and LVET tests was exclusive to one participant (MH).

Likewise, we used McNemar test to examine the statistical significance of the results ($p < 0.05$) on WAB subtests. As shown in Table 4.14, participant MH's gains in auditory word recognition and object naming were statistically significant ($p = 0.0313$, one-tailed). Participant MH (classified with conduction aphasia) demonstrated statistically significant improvement in lexical retrieval that was consistent across different outcomes measures: WAB, BNT and the LVET. Nevertheless, participant RR's (classified with Broca's aphasia) scores on the word fluency subtest of the WAB were borderline to significance ($p=0.065$, one-tailed).

Table 4.13 Individual participant statistics details for significance (p<0.05) of therapy gains on BNT, and LVET using McNemar's test.

#	Initials	BNT	LVET
	p=	one-tailed	one-tailed
1	JP	0.3770	0.5000
2	WE	0.1938	0.1875
3	DM	0.3125	1.0000
4	GP	0.2744	0.3438
5	MH	**0.0107**	**0.0156**
6	PR	0.0592	0.0898
7	ST	0.5000	0.2539
8	RH	0.6875	0.5000
9	AB	0.3953	0.5000
10	DA	0.6875	1.0000
11	AD	0.2120	0.2539
12	RR	**0.0898**	1.0000

Note: data in bold and highlighted in grey were significant at the p<0.05 level. Data in bold only approached significance (<0.1).

Table 4.14 Individual participant statistics details for significance (p<0.05) of therapy gains on the WAB subtest using McNemar's test.

#	Initials	Yes/No Qs	auditory word recognition	sequential commands	Repetition	Object naming	word fluency	sentence completion	responsive speech
	p=	one-tailed	one-tailed	one-tailed	one-tailed	one-tailed	one-tailed	one-tailed	one-tailed
1	JP	0.5000	1.0000	0.3125	0.5000	1.0000	0.2500	1.0000	1.0000
2	MH	0.5000	**0.0313**	0.5000	0.7500	**0.0313**	0.2500	1.0000	0.5000
3	ST	0.2500	0.5000	0.0625	1.0000	0.0625	0.5000	0.2500	1.0000
4	AB	0.5000	0.3633	0.1094	0.7500	0.2500	0.2500	0.7500	0.5000
5	DA	0.1875	0.3770	0.2500	0.5000	0.5000	0.5000	1.0000	1.0000
6	RR	0.5000	0.5000	0.5000	1.0000	0.1875	**0.0625**	0.5000	0.2500

Note: the data in bold and highlighted in grey were significant at the p<0.05 level. The data in bold only approached significance.

- Statistical analysis of change across participants

More broadly, statistical significance (p< 0.05) of therapy gains across the group on BNT, LVET and WAB subtests were examined using Wilcoxon matched pairs tests. The scores on BNT test approached significance (z=2.76, n=12, p=0.0058, two-tailed); however, the LVET (z=1.43, n=12, p=0.15, two-tailed) and WAB subtests (Table 4.15), were not statistically significant.

Table 4.15 Statistics details for significance (p<0/.05) of participants' performance on WAB subtest using Wilcoxon signed ranks test (n=6)

WAB subtest	Z	Asymp. Sig. (2-tailed)
Post Yes/No Qs – Pre Yes/No Qs	-.431	0.666
Post Auditory word recognition – Pre Auditory word recognition	-.680	0.496
Post sequential commands – Pre sequential commands	-.846	0.398
Post repetition – Pre repetition	-1.414	0.157
Post object naming – Pre object naming	-.944	0.345
Post Word fluency – Pre Word fluency	-1.236	0.216
Post Sentence completion – Pre Sentence completion	-1.342	0.18
Post responsive speech – Pre responsive speech	-1.633	0.102

4.4.2 Research Question 2: Were treatment gains maintained once direct treatment was ended?

At the end of the therapy program, a period of 6 weeks of self-directed treatment sessions followed, with no clinician input or guidance. The participants were free to use the software to reinforce previous treatment tasks or goals, or to attempt new forms of language production. During that period, self-directed home practice was encouraged; nevertheless, the participants reported inconsistent adherence to the recommendation.

Table 4.16 presents the participants' raw scores on the VAST sentence construction task (maximum potential score = 20) at five testing time-points: baseline (the average score of two assessments at least two weeks apart), interim 1, interim 2, interim 3 (immediately post-therapy), and maintenance (i.e., 6 weeks post the discontinuation of therapy). The group's average score showed an increase of 3 points at interim 1 (4^{th}-5^{th} week of therapy) and 1 point at interim 2 (8^{th} and 9^{th} week of therapy); after that, the group's score was maintained at 18 points at interim3 and maintenance testing time-points.

Individual performance assessment relative to the maintenance of therapy gains was conducted by comparing individual scores at two testing time-points, interim 3 and maintenance. It revealed that 33% of the sample of participants (4/12) showed a drop of an average of 1.75 points at maintenance compared to post-therapy, while 42% (5/12) showed an increase by an average of 1.2 points, and 25% (3/12) showed stability in performance by scoring 19/20 at both time-points.

As illustrated in Table 4.16 and Figure 29, PR and RR demonstrated maximum gains at maintenance which exceeded any score they have achieved throughout the course of therapy. On the other hand, WE, DM, GP, and AD were able to maintain the maximum score they have achieved at any given time-point during the course of therapy (i.e., interim 1, 2, or 3). However, MH, ST, RH, and AB score at maintenance showed a drop of 1 point, while DA showed a drop of 2 points, and JP showed a drop of 3 points compared to their maximum achieved score at any of the three interim assessment time-points.

Table 4.16 Participants' performance on VAST sentence construction task at 5 testing time-points throughout the course of therapy

#	Initials	Baseline	Interim 1	Interim 2	Interim 3 (post-therapy)	Maintenance (6 weeks post-therapy)
1	JP	18	18	18	20	17
2	WE	18	19	19	19	19
3	DM	17	19	19	19	19
4	GP	17	17	19	18	19
5	MH	16	19	20	19	19
6	PR	15	17	18	18	19
7	ST	14	15	19	19	18
8	RH	13	18	18	18	17
9	AB	11	17	16	15	16
10	DA	11	16	17	17	15
11	AD	10	14	16	14	16
12	RR	9	17	15	17	18
	mean	14	17	18	18	18
	SD	3	1.5	1.5	1.8	1.6

*note: all scores are raw data. The maximum potential score is 20 points.

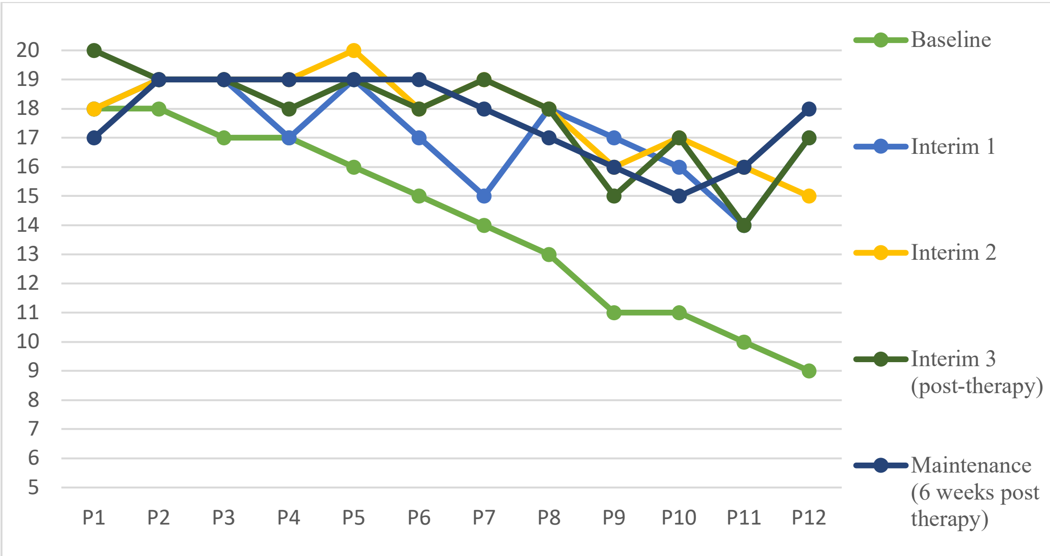


Figure 29 Participants' performance on VAST sentence construction task at 5 testing time-points throughout therapy.

Table 4.17 (below) presents the participants' raw scores on each VAST subtest at two time-points: immediately after therapy, and 6-weeks post the end of therapy with no direct intervention. To statistically examine the scores' stability at those two time-points, Wilcoxon matched pairs test was conducted. As shown in Table 4.18, no significant change has been detected in all except for one subtest, the grammaticality judgment subtest. Accordingly, participants demonstrated stability in performance on verb comprehension, sentence comprehension, verb production, fill-in verbs in sentences, and sentence construction at two testing time-points (i.e., post-therapy and maintenance). On the other hand, the grammaticality judgment task did show a significant change from post therapy to maintenance testing. The group's average raw score increased from 30.5 (SD 4.8) to 31.8 (SD 5.3), in which 9 out of 12 participants showed an increase in scores that ranged between 1-4 points. These results indicate an improvement in performance at maintenance compared to post-therapy.

Table 4.17 Participants' performance on VAST subtests immediately after completing the therapy program and 6-weeks post after the discontinuation of therapy.

Participant No. and initials		Verb comprehension		Sentence comprehension		Grammaticality judgment		Verb production		Fill-in verbs in a sentence		Sentence construction	
		Interim 3	maintenance	Interim 3	maintenance	Interim 3	maintenance	Interim 3	maintenance	Interim 3	maintenance	Interim 3	maintenance
Maximum score		40	40	40	40	40	40	40	40	20	20	20	20
1	JP	36	36	29	24	29	33	25	27	15	13	19.8	17
2	WE	40	40	37	36	38	39	37	35	18	18	18.6	18.9
3	DM	40	39	17	22	28	27	39	38	18	19	19.4	19.4
4	GP	39	40	29	21	31	35	30	34	18	18	18.4	19.1
5	MH	35	36	29	29	32	33	8	10	10	11	18.8	19.1
6	PR	35	40	34	35	37	37	26	32	11	13	18.3	18.9
7	ST	38	34	21	20	30	31	30	29	15	12	18.5	18
8	RH	40	38	21	20	32	34	24	24	15	14	17.8	16.6
9	AB	38	37	29	24	28	30	20	20	10	12	14.5	15.8
10	DA	34	34	18	20	24	27	16	13	12	13	17.3	15
11	AD	35	36	30	22	35	36	31	29	15	12	14.4	15.6
12	RR	37	39	22	16	22	20	23	25	11	11	17.1	18
	mean	37.3	37.4	26.3	24.1	30.5	31.8	25.8	26.3	14.0	13.8	17.7	17.6
	SD	2.2	2.2	6.4	6.2	4.8	5.3	8.6	8.6	3.1	2.9	1.7	1.6

Table 4.18 VAST Task-specific statistics details for significance ($p<0.05$) to test the stability of therapy gains at maintenance using Wilcoxon matched-pairs test.

Wilcoxon matched-pairs	Z	Asymp. Sig. (2-tailed)
Verb Comprehension (Maintenance) – Verb Comprehension (Interim3)	-.241	0.809
Sentence Comp (Maintenance) – Sentence Comp (Interim3)	-1.657	0.097
Grammaticality Judgment (Maintenance) – Grammaticality Judgment (Interim3)	-2.071	**0.038**
Total Comp (Maintenance) – Total Comp (Interim)	-.669	0.503
Verb Production (Maintenance) – Verb Production (Interim3)	-.672	0.502
Fill-in Verbs in Sentences (Maintenance) – Fill-in Verbs in Sentences (Interim 3)	-.360	0.719
Sentence Construction (Maintenance) – Sentence Construction (Interim3)	-.133	0.894
Total Production (Maintenance) – Total Production (Interim3)	-.225	0.822

4.4.3 Research Question 3: What is the minimum dose of therapy required to achieve maximum improvements on constrained sentence production tasks?

Table 4.16 above presented the participants' scores on the VAST sentence construction subtest at multiple testing time-points (see Table 4.3 for more details on testing frequency and time-points).The VAST sentence construction subtest items were reserved for the assessment and were not included in any therapy task; also, feedback on performance on the test was withheld throughout the multiple assessment time-points to minimize learning effects. Accordingly, the subtest examines the participants' generalisation of sentence construction skills. The multiple testing time-points allowed us to identify the minimum dose of therapy and duration required to notice evidence of therapy gains. As shown in Table 4.16 (above), the mean score of the group raised from 14 points (SD=3) at baseline, to 17 points (SD=1.5) at interim 1, and then to 18 points (SD=1.5) at interim 2. After that, the average score stabilized at 18 points at interim 3 and maintenance (SD= 1.8 and 1.6 respectively). A closer inspection of individual scores revealed a peak in performance at interim 2 time-point for 50% (6/12) of the sample of participants (GP, MH, PR, ST, DA, and AD). Of this group of 6 participants, 50% (3/6) (PR, ST, and DA) maintained their score until the end of the therapy, while 33% (2/6) of the remaining participants (GP and AD) showed a slight dip of 1 point at interim-3. On the other hand, 42% of the sample (5/12) showed improvements as early as interim 1 time-point. Of this group, 60% (3/5) of the sample of participants (WE, DM, and RH) maintained their performance throughout the course of therapy. In conclusion, for 75% of the sample of participants (9/12), 8 weeks of therapy was sufficient to show evidence of a positive response to intervention and therapy-induced gains in sentence production skills.

4.4.4 Research Question 4: How do patterns of treatment response vary across participants with varying baseline language and cognitive skills?

The sample's performance on background language and cognitive tests was retrieved from the NARU database. Table 4.19 shows background data across four tests: the BNT (taken to be an index of baseline naming accuracy), PALPA8 (taken to be an index of baseline phonological skill), the Raven's (index of general cognitive ability), and 96 synonym judgment test (representing semantic knowledge). The table also includes the participants' scores on the VAST sentence construction task, collected by the primary investigator for the current study. In the Table, the participants have been ordered, highest to lowest, according to their baseline scores on the VAST sentence construction subtest.

To investigate the extent of any correlations between background performance on language and cognitive tests and therapy gains in sentence production, we conducted Spearman's correlation test for non-parametric data. The statistical analysis of data involved 10 participants from the sample due to missing background data for 2 participants, ST and DA. The results of the correlation analyses are presented in Table 4.20 (below). For this sample (n=10), a significant positive correlation between gains in sentence construction scores and background score on BNT (r_s =-.708, n=10, p=0.011, one-tailed), and between gains in sentence construction scores and 96 synonym judgment task (r_s =-.653, n=10, p=0.020, one-tailed) was identified. However, no significant correlation was found between gains in sentence construction skills and scores on Raven's test or the Auditory non-word repetition PALPA8.

Table 4.19 Participants' raw scores on language and cognitive tests

Participant No. and initials		BNT	Ravens	96 synonym judgment task	Auditory non-word repetition *	VAST sentence production		
		baseline	background	background	background	baseline	Post therapy	gains *
Maximum score (per cent)		100	100	100	100	100	100	(post-pre)
1	WE	70	92	88	47	90	100	10
2	GP	67	97	90	43	90	95	5
3	DM	92	92	96	60	85	95	10
4	PR	58	81	83	57	85	90	5
5	ST	57	47	-	53	80	95	15
6	AD	55	64	82	23	75	90	15
7	JP	58	83	83	70	70	95	25
8	AB	52	89	75	27	65	90	25
9	RH	10	83	90	3	55	75	20
10	MH	8	81	70	0	55	85	30
11	RR	33	89	82	10	50	70	20
12	DA	25	-	-	-	45	85	40
Average		48.8	77	79	43.2	70.4	88.8	18.3
	SD	25.1	17.6	11	30.2	17.5	8.8	10.5

*PALPA8 ** scores post therapy (minus) scores at baseline

Note: the results of participants ST and DA were excluded from the analysis due to incomplete background data

Table 4.20 Spearman's correlation

Gains in sentence production scores	BNT		Ravens		96 synonym judgment task		Auditory non-word repetition*	
	r_s	p	r_s	p	r_s	p	r_s	p
Total n=10	-.708*	.011	-.363	.151	-.653*	.020	-.436	.104

*Correlation is significant at the p<0.05 level (1-tailed)

4.4.5 Research Question 5: How do treatment outcomes compare across sentence level processing and other language and psycho-social measures?

Table 4.21 shows the participants' scores on the COAST, a scale that measures self-perceived communication effectiveness, before and after therapy. An increase in scores after therapy was noted in 7 out of 12 participants, ranging between 5 and 35 points. In contrast, 4 out of 12 participants showed a drop in their scores with a range between -1 and -9. On the other hand, one participant, PR, showed stability in her scores.

To further investigate the statistical significance of the groups' score changes on the COAST after therapy, we ran a Wilcoxon matched pairs test. It revealed that, although the average score increased by 6 points post-therapy (50.2 – 56.2), the overall gain did not reach statistical significance for participant self-rating (Z=-0.275, n=12, p=0.783, two-tailed). We also explored the correlation between gains on the COAST scale and therapy gains on the VAST sentence production subtest. Similarly, Spearman's correlation test did not reveal any statistical significance ($r_{s=}$ 0.318, n=12, p=0.314, two-tailed).

Table 4.21 Participants' performance on COAST (raw scores) before therapy, after therapy, the difference between them (therapy gains), and therapy gains as measured by performance on the VAST sentence production subtest.

Participant No. and initials		COAST		COAST gains	VAST sentence production gains
		Pre-therapy	Post-therapy		
Maximum score		80	80	80	20
1	JP	68	59	-9	2
2	WE	64	58	-6	1
3	DM	42	50	8	2
4	GP	30	65	35	1
5	MH	54	53	-1	3
6	PR	53	53	0	3
7	ST	36	49	13	5
8	RH	48	56	8	5
9	AB	55	53	-2	4
10	DA	48	62	14	6
11	AD	57	62	5	4
12	RR	47	54	7	8
	Mean	50.2	56.2	6	3.7
	SD	10.9	5.1	11.6	2.1

Table 4.22 Summary of therapy gains per participant across different language modalities and linguistic levels

Linguistic level			Skills (outcome measure)	JP	WE	DM	GP	MH	PR	ST	RH	AB	DA	AD	RR
Production	1	Word	Action naming (VAST)								✓			✓	
			Object naming (BNT + WAB)					✓							
	2	Sentence	Picture description (VAST)	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓
			Fill-in verbs in sentences (VAST)				AS					AS			
			LVET					✓							
	3	Discourse (either cookie theft or dinner party)*	Analysis set utterance**					✓	✓	✓		✓	✓	✓	✓
			MLU**	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
			Number total Words**	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
			**measured by SALT software												
Comprehension	4	Word	Verbs (VAST)		✓				✓				✓		
			Nouns (WAB)					✓							
	5	Sentence	Sentence picture matching (VAST)	✓											
			Grammaticality judgment (VAST)												
			Following sequential commands (WAB)												

✓= statistically significant, AS= approaching significant at 0.0625 *The criterion for substantial improvement was set at 20% and above increase in scores compared to baseline

4.5 Discussion

This study's central aim was to integrate various effective impairment-based sentence therapy approaches, supported by evidence in the literature, into a hybrid, multilevel and hierarchical method of treatment that applies to PWA with a range of language deficit profiles. Integrating these approaches was to optimize their feasibility by reducing the time-frame that would have been required to implement each one of them individually and sequentially. The core components of this hybrid method include mapping therapy, VNeST, and SentenceShaper. Based on the VNeST literature, it was anticipated that VNeST, which targets the lexical variant of the mapping deficit hypothesis as well as semantic activation, would stimulate a widespread activation of semantically related verb networks and improve word retrieval within sentences and discourse production (Edmonds and Babb, 2011; Edmonds et al., 2009). On the other hand, the mapping therapy, which targets the procedural variant of the mapping deficit hypothesis, was expected to increase metalinguistic awareness of the relationship between the verb and its thematic roles (Schwartz et al., 1987) and thus facilitate more frequent and accurate expressions of verb argument structures. Also, improving those skills would promote more successful attempts of self-correction. Additionally, SentenceShaper adds a novel element to our method by offering cognitive processing support to the participant. It relieves the PWA of time pressure in language production and the processing overload associated with it that interferes with language production and often leads to sentence production breakdown in (Linebarger et al., 2004). Accordingly, self-monitoring opportunities will be more frequent, with the availability of supportive tools to practice self-correction; it includes features such as the record, reorder, edit, delete, insert, and playback within SentenceShaper. Lastly, the discourse-focused tasks presented in the therapy program's final level (Level 3) were designed to gradually bridge from enhanced sentence production within constrained and limited output strings to increasingly complex event sequences providing a scaffold for longer and more linguistically rich narrative monologue samples. It was hypothesized that these bridging tasks would promote greater resilience for enhanced sentence production skills, becoming automatized within more cognitively demanding discourse levels of communication.

To determine the participants' therapy gains in sentence production skills, we selected the VAST sentence construction subtest as a primary outcome measure. Accordingly, the test items were reserved for the assessment purpose and were not included in therapy to eliminate the learning effect. Therefore, the participants' scores on this subtest will reflect their generalised sentence production skills on untargeted items. As mentioned in the method's section, the therapy items included a range of preselected verbs (within the built-in workbooks in SentenceShaper) and verbs selected by individual participants, based on their personal preference, for the self-generated sentences task. Due to the undefined set of treated verbs in our study, the outcome analysis will involve assessing the generalisation of language skills as measured by the selected outcome measures.

The feasibility of the current therapy program in remediating lexical and sentence processing deficits in PWA

As measured by standardised assessment tools, the findings revealed robust improvements in sentence and discourse production skills. Also, generalisation of therapy gains to untargeted skills such as lexical retrieval and comprehension at the word level and sentence comprehension was noted in some participants. Therapy gains were maintained and no signs of deterioration were detected after the discontinuation of therapy. Moreover, the outcomes of the current approach were comparable to the results reported in the literature of the implementation of each original therapy protocol (that composed our hybrid approach) in isolation. Overall, the therapy outcomes supported the feasibility of the current approach; It was further emphasized by clinical observation such as the high retention rate and good adherence to the therapy program. Also, by including technology, the program was successful in inducing statistically significant therapy gains in a cost-effective way, through increasing independent home practice and minimizing clinician-delivered therapy sessions. The topic will be revisited shortly in the conclusion.

Lexical processing gains at the word level

Noun processing gains at the word level were tested in three subtests: WAB auditory word recognition, WAB object naming, and BNT object naming. As shown in Table 4.22, a single participant within the sample, MH, demonstrated statistically significant noun processing across all three subtests. Verb processing gains at the word level were tested in two subtests: VAST action naming and VAST verb comprehension. The assessment revealed that 42% of the sample showed verb lexical gains on either comprehension or production tasks. Overall, the outcomes of our novel therapy method showed more consistency in inducing verb processing gains (42% of the sample) than noun processing gains (8.3% of the sample) at the word level.

The statistical analysis of scores across participants on the above lexical processing tasks showed that results approached significance on the VAST verb comprehension and BNT. However, it did not reach significance on the VAST verb production or the WAB auditory word comprehension and object naming subtests.

Association between lexical processing gains at the word level and gains in sentence production skills

As shown in Table 4.22, 33% of the sample of participants demonstrated therapy-induced improvements in verb processing at the word level, production or comprehension, that was associated with improvements in sentence production. However, this pattern did not apply to noun processing gains. For example, participant MH demonstrated statistically significant improvements in noun processing at the word level, in both modalities production and comprehension; those gains were not associated with improved sentence production skills. On the other hand, 50% of the sample showed statistically significant therapy gains in sentence production skills that were not associated with lexical gains at the word level.

Assessing this discrepancy in the frequent association of verb processing gains vs noun processing gains with sentence production gains will only be possible if nouns and verbs in each test were matched for their psycholinguistic features (Alyahya et al., 2018b). This requirement is not fulfilled in the current study; therefore, the data is insufficient to draw conclusions about the observed performance pattern.

Verb processing at the sentence level

Participants' gains in verb processing at the sentence level were examined in two production tasks: VAST fill-in verbs in a sentence and light verb elicitation test LVET. The LVET was the only test that detected a significant increase in scores in a single participant – this was MH who had also demonstrated significant improvement in 3 noun processing tasks (at the word level). Although he showed consistent strong improvement in noun and light verb processing in the subtests as mentioned earlier, this did not extend to include concrete verb processing at the sentence level (fill-in verbs in a sentence task) or the word level (e.g., VAST verb comprehension, and VAST verb production). This interesting performance pattern in which light verb processing gains were associated exclusively with noun processing gains but not concrete verbs would require additional research to determine the underlying cause. It would be useful to replicate the study to include more participants with the same aphasia profile as MH (conduction aphasia) and outcome measures with verbs and noun items matched in their psycholinguistic features.

The statistical analysis of scores across participants on those two subtests, revealed statistically significant changes on the VAST fill-in verbs in a sentence subtest only and did not include LVET. The increase in scores post therapy on VAST fill-in verbs was mainly driven by participants with fluent aphasia. However, further research is required to determine the underlying influencing factors, as some participants with fluent aphasia showed the reverse pattern.

The "division of labour" theory (Gordon and Dell, 2003), explained earlier, propose that light verb deficits and over reliance on full verbs stem from a syntactic deficit. The notion was supported by findings from Thorne and Faroqi-Shah (2016). Accordingly, we anticipated that improvements in sentence production skills within- and across- subjects will be associated with improvements in light verb production skills; Nevertheless, our findings were contradictory. The only participant (MH) who showed statistically significant therapy-induced gains on LVET was also one of the only two in the sample (n=12) who did not demonstrate statistically significant gains in sentence production skills. Likewise, no association was found between therapy gains on sentence production skills and LVET skills across the group.

Therapy gains in sentence production skills

The therapy induced improvements in sentence production skills were evident in 83.3% of the sample (10 out of 12 participants), demonstrated by the statistically significant gains on the VAST sentence production subtest. The score on this task reflects meeting the minimum requirement for the verbal production of complete, grammatically correct, and informative sentences in a structured picture description task, scored with a novel scoring protocol (see Appendix 1). We examined the data for the characteristics of the participants who achieved the most therapy gains. It revealed that the subgroup with superior therapy gains performed at or below average at baseline, 67% of that subgroup was composed of participants with non-fluent aphasia. On the other hand, the subgroup that showed more modest gains performed above the average score at baseline, 67% of that subgroup was composed of participants with fluent aphasia. The factors that restricted therapy gains for the subgroup with milder sentence production deficit at baseline is likely related to the ceiling effect; it limited the maximum potential for improvement compared to the other subgroup.

In contrast, two participants did not show gains in sentence production skills which composed 16.7% of the sample. A closer look at the language profiles of those two participants, MH and WE, revealed that they both performed above the group's average on the VAST sentence production task at baseline with an accuracy score of 80% and 90%, respectively. A shared characteristic between the two participants is that they both have a classification of fluent aphasia. Their lexical therapy gains (at the word level) reflect improvements in language skills that are known to be of a relative weakness in their aphasia type. Accordingly, we can infer that participants with well-formed sentence production skills (i.e., determined by performing above 80% accuracy on VAST sentence construction task) can still benefit from the current therapy approach in advancing their lexical processing skills. They also showed noticeable therapy gains in discourse production on two measures (MLU and the total number of words), which will be discussed shortly.

Likewise, statistical analysis of changes in performance across participants on the VAST sentence production subtest showed statistically significant gains post therapy. It replicates the outcomes reported in the literature in which 86% of the sample of participants (n=84) demonstrated generalised improvements in sentence production skills to untrained items.

Through clinical observation, it was noted that following the intervention, most participants' performance on the sentence production task (in a picture description task) exceeded the minimum requirements accounted for by the novel scoring protocol (see Appendix 1). They demonstrated advanced skills in sentence expansion and in producing longer sentences with more adjectives and prepositional phrases. Those secondary sentence production gains were not reflected in their scores, with the current scoring criteria. To determine the nature and consistency of these observations, a more elaborated scoring rubric that covers all possible complex sentence elements will be needed, along with a well-controlled selection of picture stimuli designed to elicit complex sentence production at baseline and post-therapy. Since the current sentence production stimuli (20 items of the VAST sentence construction task) were designed to elicit simple sentence production, it is not clear how participants would respond at baseline if the stimuli prompted complex sentence production. Accordingly, the stimuli used in this study do not enable us to draw further conclusions about the clinically observed advancement in sentence production skills beyond the simple primary target.

Sentence processing

Therapy gains in sentence comprehension skills were tested in 4 subtests: VAST sentence comprehension, VAST grammaticality judgment task, WAB answering Yes/No questions, and WAB sequential commands. A single participant from the sample (n=12) showed significant improvement in one of the above subtests. Participant JP's increase in scores on the VAST sentence comprehension subtest was statistically significant. It was also associated with statistically significant gains in performance on the VAST sentence production task. A possible factor contributing to this indirect therapy gain could be attributed to the mapping therapy component in our method. It was designed to restore the mapping between the thematic and grammatical roles and increase metalinguistic awareness (see Chapter 2). A replication of the current study to include a larger sample size would help examine cross-modality generalisation of sentence therapy gains associated with the current novel therapy.

Across participants, statistical analysis revealed no significant therapy-induced changes on any of the above listed four subtests. This outcome is contradictory to findings reported in the literature , as inter-modality generalization from oral sentence production to oral sentence comprehension was found in 64% of a sample of 25 participants (Poirier et al., 2021).

Generalization to discourse

Three outcome measures from the SALT software default-set of variables have been selected to analyse therapy-induced changes in discourse production: 'total number of words', 'analysis set utterances' and MLU in words variables. The current therapy was designed to improve verb production within a sentence context including access to its verb argument structure, which targets both semantic and phonological processes of word retrieval (see Chapter 3 for more information on semantic and phonological processes). It also aims to improve the grammaticality of produced sentences and encourages sentence expansion. Accordingly, the intervention was expected to increase: 1. successful word retrieval attempts, indicated by 'total number of words'. 2. number of produced sentences that meet the “complete, informative, grammatically correct and relevant” criteria, as measured by ‘analysis set utterances’ variable. 3. frequency and amount of sentence expansion, estimated by MLU.

Our findings showed that 92% of the sample of participants showed noticeable increase in both number of total words and MLU, on either discourse (cookie theft or dinner party) at any post-therapy testing time-points (interim 3 or maintenance). However, only 58% of the sample also showed improvement in ‘analysis set utterances’ that indicates increase in production of complete, verbal, informative, relevant and grammatically correct sentences. The majority of participants who did not show noticeable changes in their sentence production skills were classified with fluent aphasia. Nevertheless, the fluent subgroup demonstrated strong performance on sentence production tasks at baseline, which could explain the lack of noticeable therapy-induced changes as the window for improvement was narrow on that measure.

In the current study, the observed skill generalization to discourse is in alignment with outcomes reported in the aphasia sentence production therapy literature. Within (Poirier et al., 2021) systematic review, 73% of a sample of 59 participants showed therapy-induced improvements on the following measures: mean length of utterances, number of grammatically correct sentences, number of correct information units, complexity of sentences, and number of open- and closed-class words, elicited through a story retell task.

Interpretation of the findings in light of the reported outcomes of individual methods

To the best of our knowledge, a comparable study to the multilevel therapy we presented does not exist in the literature to date. Therefore, in this section we will assess our findings against those reported of individual implementation of each of the three approaches that composed our hybrid method.

<u>VNeST</u>

A literature review presented by Edmonds (2016) of the pooled results of all VNeST studies (i.e., therapy implemented as a single approach) showed evidence of improvements in lexical retrieval at the word, sentence, and discourse levels in English (Edmonds and Babb, 2011; Edmonds et al., 2014; Edmonds et al., 2009; Edmonds et al., 2015; Furnas and Edmonds, 2014) and Korean (Kwag et al., 2014). The sample of participants’ performance (n=19) on standardized tests revealed that 86% of the sample improved in noun naming (measured by either OANB or BNT and verb naming subtest from the NAVS), and 58% of the sample also improved in verb naming (Edmonds, 2016, p.127). The extent of improvement in lexical retrieval at the word level was not replicated in our study as only 17% (2/12) of our sample showed improvement in action naming and 8% (1/12) in object naming..

A constrained sentence production task (i.e., using pictures from the NAVS for sentence elicitation and implements a scoring approach that does not account for grammatical errors) showed that 75% of the sample improved on sentence production skills on untrained items. In the current study, the participants' scores on the VAST picture description task (i.e., a constrained task that implements a scoring rubric that accounts for grammaticality, explained in Appendix 1) revealed that 83% (10/12) of the sample of participants showed significant therapy induced gains. However, a limitation to this comparison is the discrepancy in the samples' size (n=12 in our study vs n=19 in VNeST literature).

In VNeST literature, sentence production in discourse was measured by complete utterances CUs variable. These analyses revealed that 59% of the sample of participants demonstrated significant gains (Edmonds, 2016). In our study, we used the 'analysis set of utterance' variable generated by the SALT software to counted for all complete, informative, grammatically correct and relevant sentences in a given discourse. Although statistical analysis has not been implemented to determine the significance of changes, we established a criterion that identified changes above a 20% increase in scores from baseline as a robust improvement. Accordingly, our findings showed that 58% showed improvement in sentence production skills in discourse (i.e., either cookie theft of the dinner party). Also, on the total number of words variable, 92% of the sample showed therapy-induced increase.

In summary, our findings did not match the VNeST studies. The percentage of participants that showed therapy-induced gains in lexical retrieval skills in our study was much less than the number reported in VNeST studies. On the other hand, the percentage of participants who demonstrated improved sentence production skills at both sentence level and discourse level was markedly higher in our study.

<u>Mapping Therapy</u>

Rochon and colleagues (2005), tested the efficacy of the mapping therapy approach in treating sentence production in 3 participants with chronic Broca's aphasia. The training included canonical sentences (active and subject cleft) and non canonical sentences (passive and object cleft). To assess therapy-induced changes in sentence production, two constrained sentence production tests were used, the Caplan and Hanna's Sentence Production Test (Caplan and Hanna, 1998) and the Picture Description with Structure Modeling Test (Fink et al., 1995). The results showed improvements that were limited to the production of trained sentences structures and did not include untrained structures. Therapy gains were maintained at one month post therapy. However, no cross-modality therapy gains were noted, as the participants did not show improvement in their sentence comprehension skills.

On the VAST sentence production subtest outcome measure in our study, 83% of the sample of participants demonstrated statistically significant therapy gains. However, our therapy tasks included only canonical sentences and the outcome measure did not distinguish the structure of the produced sentences.

In Rochon et al. (2005) study, the generalization of therapy gains to narrative construction was assessed using the QPA discourse analysis method and story retell task.. The performance of the 3 participants showed an increase in MLU of 0.0, 0.5, and 1.0. Their scores at baseline ranged between 2-4.5 which is significantly lower than the controls' score (12 neuro-typical subjects reported in (Rochon et al., 2000)) of 8.17 (SD=1.39). However, the only common variable we used in our study in discourse analysis was the Mean Length of utterance MLU (i.e., generated by the SALT software). The narrative samples were elicited by the cookie theft and dinner party picture scene stimuli. The MLU increased between 1.65 and 3.9 post therapy (in 11 out of 12 participants), which is higher than the scores reported in Rochon et al. (2005). Nevertheless, due to the differences in sample size (3 vs 12 participants) and variation in baseline scores (2-4.5 vs 2.4-8.3 MLU scores) the results should be interpreted with caution.

<u>SentenceShaper</u>

Several studies investigated the outcomes of using SentenceShaper as a therapy tool (Linebarger et al., 2004; Linebarger et al., 2007; Linebarger et al., 2001; McCall et al., 2009). The assessment materials were similar to those used in training; however, specific items were reserved for assessment purposes only and were not included in therapy. The participants' spontaneous unaided verbal narratives were analysed using the Quantitative Production Analysis system QPA (Saffran et al., 1989) to determine intervention-induced changes. Only one case study by McCall et al. (2009) examined the outcomes of practising syntactically complex sentences (subordinate clauses) following a period of implementing the general approach described in previous studies. Their findings indicated that this approach produced further noticeable improvements, as words per sentence increased from 3.6 to 8.12 following the general therapy approach, then increased further to 11.56 words per sentence when syntactic structures were targeted in practice. In our study, the discourse elicitation method and the analysis measures we used were different from those used in SentenceShaper studies. However, the Mean Length of Utterance MLU was a common analysis variable (i.e., generated by the SALT software in our study and by the QPA method in SentenceShaper studies). In our study, we found that 67% of the participants showed at least 2.2 increase in MLU from baseline with a range of 2.2-3.9, either immediately post-therapy or at maintenance testing (on either the cookie theft elicitation task or the dinner party). The remaining 25% of the sample showed less increase with a range of 1.65- 1.8 MLU. Nevertheless, caution should be taken when comparing the results from both studies, as SentenceShaper studies included only participants with non-fluent aphasia while the sample of participants in our study encompassed a range of aphasia types, including fluent aphasia. Accordingly, performance at baseline on the MLU variable in our sample was diverse and ranged between 2.4 – 8.3 on either the cookie theft or the dinner party discourse. In comparison, the participants’ scores in the SentenceShaper studies were at the lower end of that range, possibly related to their non-fluent aphasia classification. This distinction could explain the limited increase noticed in our sample compared to participants' performance in SentenceShaper studies. Another reason could be related to the type of intervention in our study that involved structure-specific sentence therapy rather than the narrative construction training reported in SentenceShaper studies. Although our approach included sentence expansion and combining sentences, it is likely that these tasks did not engage message-level processing (therapy at the level of the event) (Marshall, 2009; Marshall, 2017) to the same extent that a narrative construction task would.

In conclusion, we consider the combined approach as feasible as individual implementation of each method in isolation. It generated statistically significant improvements in a considerably reduced intensity plan, compared to the original intervention protocol reported in the literature. Also, since it incorporated user-friendly technology and fostered independence in home-practice, we assume that the accessibility, affordability and practicality was an additionally positive feature of our study.

Maintenance of therapy gains

The participants' scores immediately after the discontinuation of therapy and after 6 weeks period with no direct therapy were compared for stability. The results on all outcome measures except the grammaticality judgment task showed no change, indicating maintenance of therapy gains. However, at the maintenance testing time-point, the participants' scores on the grammaticality judgment task showed an increase.

The current intervention method incorporated the principles of mapping therapy that increases metalinguistic awareness. It also involved constant prompting for self-assessment of verbal productions and self-correction as needed. Those prompts closely resemble the task of sentence grammaticality judgment, which likely influenced the participants' performance over time. The fact that it continued to progress, even after the discontinuation of direct therapy, may indicate internalization of the cueing strategies and generalized implementation in everyday communication, which strengthened the skill over time due to practice.

Another possible factor is likely related to the distinctive design of the task relative to the rest of the VAST subtests. It requires the participant to produce or choose one of two answers: good or bad. It means that there is a 50% of getting the correct answer by chance. The verb and sentence comprehension tasks require the participants to point to the target picture in a field of 4 which lowers the chance of getting the correct answer randomly to 25%. Furthermore, the production tasks require the participant to verbally produce the target word or sentence, eliminating the possibility of getting the right answer by chance.

Overall, performance was maintained at 6 weeks post discontinuation of direct therapy, as no change was seen across measures (i.e., no signs of deterioration were detected). The metalinguistic awareness achieved through intervention may be an integral factor in ensuring this maintenance of therapy gains as well as the increased gains in grammaticality judgement.

Studies that have examined sentence therapy in aphasia reported similar maintenance of the acquired therapy gains in majority of participants (86% of a sample on 49 participants) (Poirier et al., 2021).

The minimum dose of therapy

The multiple-testing time-points design of the current study enabled us to identify when a positive response to intervention can first be noted. The analysis revealed that for 75% of the sample of participants (9 out of 12), 8 weeks of therapy was sufficient to show signs of statistically significant gains in sentence production skills. Therefore, the minimum amount of therapy was 8 weeks of the therapy protocol outlined in this study, composed of Level 1 and Level 2. After that, the sample's average score showed stability between interim-2 and interim-3 (12^{th} – 13^{th} week of therapy) and maintenance testing time-points.

A possible solution to overcome the plateaued performance and stimulate further advancement would be to replace the current third level of the program with a task that would target increased complexity of sentence and discourse constructions or incorporate tasks that promote generalization of language skills to daily life activities.

The influence of background language and cognitive skills on therapy outcomes

The differential response to the novel sentence therapy method across participants with varying baseline language and cognitive skills was explored in the study. The statistical analysis of the participants' therapy gains in sentence production and their background scores on naming, cognitive, semantic and phonology tests (i.e., BNT, Raven's, 96 synonym judgment task, and PALPA 8, respectively) revealed an interesting correlation. It identified that the participants' baseline scores on naming and semantic tasks (BNT and 96 synonym judgment task) were strong predictors of their therapy gains. These findings are in agreement with Lambon Ralph et al. (2010) report on a related naming therapy, with some variation in the selected elicitation methods; instead of the Pyramid and Palm tree PPT (Howard and Patterson, 1992) they have used, we implemented the 96 synonym judgment task (Jefferies et al., 2009) in our study to determine the status of semantics skills. Lambon Ralph and colleagues concluded that language ability and cognitive status were both independent and important predictors of naming therapy outcomes.

The language deficit in aphasia is commonly associated with impairments in attention and executive functions (e.g., Murray, 2002; Purdy, 2002). The influence of these impairments on language production remains unclear; however, a correlation has been identified in a number of studies (e.g., Baldo, Dronkers, Wilkins, Ludy, Raskin, & Kim, 2005; Edwards, Ellams, & Thompson, 1976; Martin & Allen, 2008). A strong correlation was found between cognitive skills and language treatment gains; similarly, between cognitive skills and the maintenance of therapy gains (Yeung & Law, 2010). It was proposed in the literature that efficient cognitive skills, such as problem-solving, influences the internalization of cueing strategies and, therefore, the carryover of practising them in daily life interactions (Yeung & Law, 2010). The current study included a cognitive measure (Raven's test) to explore cognitive skills' predictability of therapy gains across aphasia subtypes and severities. The statistical analysis results showed no correlation between scores on Raven's test and therapy gains on the VAST sentence construction task. Accordingly, in our sample, cognitive skills measured by Raven's test were not strong predictors of therapy outcomes. The contrast between our findings and what has been reported in the literature could be attributed to differences in sample size or participants' characteristics. As low scores are usually observed in Global and Wernicke's aphasia (Gonzalez et al., 2020) and our sample did not include the two aphasia subtypes.

Correlation between gains in sentence production skills and changes in psychosocial measures

The Communication Outcome after Stroke (COAST) scale was used to explore any changes in the participants' perception of their communication effectiveness associated with therapy gains. Although 83% of the sample demonstrated statistically significant gains on the sentence production task, statistical analysis revealed no association between those gains and changes in the participants' self-perceived communication effectiveness post-therapy.

One possible explanation of why improved performance, detected by standardised tests, did not influence self-perceived communication effectiveness is the lack of variety in practising settings. Since language practice was limited to either a clinical setting with a clinician or at home with a personal device, the participants were not exposed to situations designed to promote generalisation of the developed skills to everyday communication.

4.6 Conclusion

In summary, we have developed the current therapy protocol to establish a single approach that targets several therapeutic goals simultaneously and delivers a cost-effective high dose of training to remediate sentence deficits in PWA with various language profiles. It was also designed to promote self-correction and self-monitoring skills in language production and foster autonomy in home-practice.

The outcomes suggest that we have, to a significant extent, achieved this goal. As 83% of the sample (10 out of 12 participants) presented statistically significant therapy-induced changes in sentence production skills, those gains were maintained after the discontinuation of direct clinician-guided therapy. Moreover, in some cases, indirect therapy gains were noted within the targeted modality (i.e., verb, noun, sentence, and discourse production) and across modalities (i.e., verb and sentence comprehension and grammaticality judgment).

Additionally, we have noted indications that supported the feasibility of the current approach. First, of all participants who started the therapy program, the retention rate was 100% with a self-report of good adherence to home practice. The initial session of demonstration of the computer program operation and the clinician-direct therapy session was sufficient for the participant to carry over home-practice independently. Only a small number of participants required minor support from their caregivers. Those participants had dexterity and mobility limitations and required assistance in positioning the computer and the computer mouse and adjusting the mouse cursor speed (i.e., slow it down) to improve its accuracy. Overall, the participants' easy adaptation to the computer-based therapy program and their adherence to the 3 months course indicated their acceptance of this approach. We may also infer that it scores highly in terms of usability and learnability.

Moreover, this approach was successful in increasing the amount of language training in a cost-effective way. Each participant was able to practice for an average of 5 hours a week; only one hour of this total amount involved a clinician-directed session. The other four hours of training were still reviewed by the clinician (i.e., by replaying the audio recordings) on a weekly basis to check the progress and give feedback. Therefore, the cost of a Speech-Language Therapist/Pathologist SLT/P -delivered therapy sessions was reduced from 5 hours to one hour per week, in addition to the one-off cost of the program ($19.99, equals £13.80 based on currency conversion on 31 August 2021, for iPad or Windows versions).

Accordingly, this approach is useful in delivering therapy to participants in rural areas with poor accessibility to regular SLT/P sessions or in places with a shortage of SLT/P s.

Although the self-reported measure of communication effectiveness and quality of life (i.e., COAST) did not capture any significant changes related to intervention, the participants and their caregivers had a positive impression. They shared stories of small successes and improvements in their language production as they have noticed it in their daily life activities. Nevertheless, the current program can be refined to promote a better transfer of therapy gains in language production to everyday communication and maximize functional communication outcomes. The modification could involve a gradual fading of the structured tasks and increasing less structured ones such as answering open-ended questions or take part in an unstructured conversation. Also, it could involve practising language production in different settings such as ordering coffee or having a conversation on the phone, first with a familiar person, then with an unfamiliar listener such as customer service agents.

As explained earlier, the therapy principles that composed our novel sentence therapy method were selected for their applicability to languages other than English, specifically Arabic. The outcomes of the current study in Chapter 4 demonstrated the feasibility of our novel method delivered in English, the same language in which the therapy principles were originally developed. With this evidence, we will endeavour to test the feasibility of a translated and adapted version of this novel method in Arabic. Also, to compare the findings in both languages. The next Chapter, 5, will present the process of translating and adapting the assessment and therapy materials before their implementation in the Arabic therapy study (an equivalent to the current study), which will be presented in Chapter 6.

CHAPTER 5 Development of Arabic assessment and treatment materials

5.1 Introduction

Stroke and neuro-disability are increasingly exerting substantial demands on health services globally, limiting functional activity and quality of life in millions of people worldwide (Johnson et al., 2016). Rehabilitation scientific knowledge and practice has emerged from Western economies such as the United States, United Kingdom and European Union countries which poses a practical problem for health practitioners in the vast number of other countries around the world, in terms of the need for linguistically and culturally appropriate translations and technologies. Within aphasiology, it is established that a valid and sensitive language assessment tool is fundamental to the development of effective and valid language therapy programs (Carter et al., 2005). Nevertheless, professionals in Arabic-speaking countries needing to implement clinical assessment of individuals with language disorders encounter significant challenges with the current lack of assessment resources and normative data (Khoja, 2019). A study by Khoja (2019) clarified some of the most common practices of Speech-Language Therapists/Pathologists (SLT/P s) in Saudi Arabia in language assessment. The questionnaire's outcome showed that 85% of the tests used in language assessment were non-standardized translations and adaptations of English tests performed independently by the treating SLT/P. In some cases such as Military Hospitals, SLT/P s followed an assessment protocol developed by other SLT/P s in the same department, in other words ad hoc assessment tools which are fine when comparing data from people with aphasia within the same institution, but very limiting for the purposes of formal audit or research. It contained a battery of tests that investigates several communication skills, including language. Likewise, the tests within the protocol did not undergo the process of standardization. Of the 122 SLT/P s participating in the questionnaires, 56 SLT/P s (less than 50%) reported formally assessing aphasia. Instead, they reported relying on informal clinical assessment on its own or with the addition of informally adapted versions, with no standardization, of the BDAE(Goodglass and Kaplan, 1983) and WAB (Kertesz, 2007).

One of the well-recognized barriers to developing language assessment tools for Arabic speakers is the complexity of the language. For example, having two variations of the language, a situation labelled “diglossia” (Ferguson, 1959) (see Chapter 2), makes it difficult to select only one variation in testing. Colloquial Arabic is the native language learned at home and used in everyday interactions, while the modern standard Arabic MSA is limited to written texts and formal oral communication (e.g., news, academic discourse, and official speeches). Also, “Arabic dialects differ from Standard Arabic at phonological, lexical, morphological and syntactic levels” (Harrat et al., 2019, p.262). Ideally, the native language of everyday interaction is the choice for testing language production skills such as sentence and discourse. However, colloquial Arabic does not have a formal written form. On the other hand, using the MSA in testing is not feasible as fluency in colloquial Arabic does not equal proficiency in MSA. Accordingly, using MSA to test language skills could be disadvantageous. Although colloquial Arabic does not have a rule-governed formal written form, an informal written form does exist. It is mainly written as it is produced, and comprehension depends mainly on the reader’s knowledge of the dialect. The closest example in English would be reading the word “read” in the present tense and the word “read” in the past tense. They are both written in the exact same form, but the reader’s perception will depend on their knowledge of the language and the context in which the word appeared. Nevertheless, even with knowledge in colloquial Arabic, frequent errors are very common even in neuro-typical adults speaking the same dialect. Accordingly, we eliminated the option of using the informal written form of colloquial Arabic in translation. However, using it as a supplement to clarify a text written in MSA will be explored in translating a questionnaire-based material (system usability scale SUS) in the current study.

Also, there is a range of distinct dialects spoken in Arabic-speaking countries such as Levantine, Gulf, Egyptian, etc. Moreover, within Saudi Arabia, several regional dialects exist, including Hejazi, Najdi, Gulf, Southern, and Bedouin (Alasmari, 2015). Several features differentiate between those Saudi dialects, which challenges the applicability of a test that implements one dialect in testing a speaker of a different dialect. Also, these variations between dialects limit the options for computerized assessments and therapy approaches in Arabic. At present, automated speech in electronic devices usually utilizes MSA, which may be appropriate and comprehendible in specific contexts such as global positioning system GPS map instructions. However, speech and language therapy aims to improve functional communication in everyday interactions, which in this case involves improving language production in the individual's native dialect (i.e., colloquial Arabic). At present, implementing computer and electronic devices that utilise automated speech in language assessment and therapy will be limited to MSA unless a new approach of customizing automated speech into specific Arabic dialect emerges, which does not exist to the best of our knowledge yet. Assessing PWA proficiency in MSA could lead to inconclusive findings regarding the participants' language profile and functional communication. Also, in therapy, training the participants, through computer-based therapy, to produce MSA to verbally express their needs and thoughts may result in forms of language change which will be of limited function communicative benefit. Alternatively, computer-based interventions can still be useful in cases where verbal instructions and the presentation test items can be recorded by a native speaker of a specific dialect. Nevertheless, this option is time-consuming and less flexible.

When planning the Arabic therapy study in 2018, presented in Chapter 6, a limited number of language assessment tools were available to test Arabic speaking adults with acquired language disorders. One of the first published tools was the Jordanian Arabic Bilingual Aphasia Test BAT (Paradis, 1987). Unfortunately, at least two factors hindered its usability in our study. The test was created for speakers of the Jordanian Arabic dialect and did not provide an in-depth investigation of sentence and discourse production. Similarly, the Arabic Diagnostic Aphasia Battery A-DAB-1 (Al-Thalaya et al., 2018), Kasr El Aini Arabic Aphasia test KAAT (Hassanein et al., 2002), and Arabic version of the Comprehensive Aphasia Test CAT (Abou El-Ella et al., 2013) all were developed in dialects that are distinct, Lebanese and Egyptian, from the dialects spoken in Saudi Arabia. Moreover, it presented a limited focus on sentence and discourse production skills. Nevertheless, one available test at that time had been adapted to the Saudi Arabic dialect. The Object and Action Naming Battery OANB (Alyahya and Druks, 2016), designed to test lexical retrieval ability in PWA, was translated from its original English into Saudi-Arabic. Yet, it did not meet the current study's requirements, which involves more focus on sentence and discourse production. In 2016, it was reported that a project to create a comprehensive assessment battery for Arabic speaking adults with acquired language disorders, with its normative database, was in development (Khwaileh et al., 2016). The dialect used in the test is Qatari/Gulf Arabic dialect, which is also spoken and well understood by the eastern province population in Saudi Arabia.

However, the project remains a work in progress. More recently, in 2020, a pilot study of a short aphasia test (‘SATG’) that was created for Saudi Arabic speakers with language disorders has been published (Altaib et al., 2020). The SATG was derived from “the aphasia diagnostic informal assessment” developed by Alzahrani (2003). According to Altaib et al. (2020), the test was based on the English version of the BDAE (Goodglass and Kaplan, 1983) and WAB (Kertesz, 2007). Thus, SATG provides an overview of language skills in several domains that could supplement our current study's aims. To establish a baseline of performance on various language tasks and compare these baselines to results post-intervention to measure therapy gains. However, the test had not yet been published at the time of conducting the therapy study reported in Chapter 6. Moreover, a primary assessment tool was still needed, in order to perform an in-depth assessment of sentence and discourse production skills in PWA. For these reasons, we have chosen to translate and adapt an assessment tool that was designed to test sentence construction extensively, which includes verb and sentence production and comprehension, from its original English language into Arabic. We further aimed to adapt it to the Saudi culture and the central Saudi dialect. However, due to the time limitation for completing this project, developing a normative database and performing an extensive psychometric evaluation of the test was not considered. Instead, the test was intended to serve as a tool to measure outcomes, to track changes in sentence and discourse production skills at several time points in response to intervention. Nonetheless, test retest reliability was carried out due to evaluate its feasibility as a post-therapy outcome measure.

Two of the most well-recognised language tests that investigate sentence construction and production are the Northwestern Assessment of Verbs and Sentences NAVS (Thompson, 2012) and the Verb and Sentence Test VAST (Bastiaanse et al., 2002). Upon comparison, we have found that the VAST exceeded the number of items in the NAVS by at least 10 items in the following subtests: verb comprehension, sentence comprehension, and verb production. Also, the two tests implemented distinct approaches in eliciting sentence production. For example, the VAST required the participants to verbally produce a sentence in a picture description task following two practice items. On the other hand, the NAVS offered two tasks to prompt sentence production: the Argument Structure Production Test (ASPT) and the Sentence Production Priming Test (SPPT). In both tasks, the aim is to investigate the participants’ ability to construct specific sentence structures rather than test their ability to retrieve words, verbs and nouns, in a sentence context alongside their sentence construction skills. Therefore, the task includes providing the participant with cues (i.e., visual, written, semantic) to offset word retrieval difficulty and focus on examining the participants' skills in constructing specific sentence structures. For example, in the Argument Structure Production Test (ASPT), words are written on the pictures to indicate the names of the people/objects and actions. Also, in the Sentence Production Priming Test (SPPT), the participant is presented with a semantically revered counterpart of the target sentence. The VAST was selected as a primary outcome measure for the current study. As the therapy protocol that will be implemented in the study included both VNEST and the mapping therapy approach, there was an expectation that both lexical retrieval in sentences and discourse would improve. Accordingly, it is within our aims to investigate the PWA ability to retrieve words in a sentence context without external cues' assistance.

For the current therapy study, presented in Chapter 6, we have selected two primary outcome measures and three secondary measures to supplement it. Four out of the five tools are the same tools used in Chapter 4 with the English group. The primary tools are composed of the VAST and discourse analysis, using the cookie theft and dinner party stimuli for elicitation. The secondary tools are the WAB-R, COAST, and the System Usability Scale SUS (Brooke, 1996). The purpose was to create consistency in the outcome measures and therapy protocol implemented across studies when possible, to conduct a meaningful comparison between findings in both groups.

The translated and adapted version of the WAB-R (Kertesz, 2007), which we have selected as a secondary assessment tool, has become an established assessment tool and is popular among SLT/P s in Saudi Arabia to assess PWA language function (Khoja, 2019). Another factor that supported our choice is that the WAB has been identified as an essential outcome measure for aphasia measurement by the Research Outcome Measurement in Aphasia (ROMA) group, which produced a consensus statement that "provides recommendations for a core outcome set (COS) for use in aphasia treatment studies". A COS is "a minimum set of outcomes that should be measured and reported in research trials of a specific health condition or population" (Wallace et al., 2019, p.181). On the other hand, the implementation of COAST (see Chapter 4) will provide an insight into the participants' perception of their communication effectiveness in this population (i.e., Saudi individuals with aphasia) through self-report. Moreover, by comparing their scores before and after therapy, we can assess any changes attributed to the intervention.

The above described primary and secondary assessment tools were selected to establish a baseline of language performance in Saudi-Arabic PWA and as an outcome measure to capture therapy gains of a specific intervention program. The current study in Chapter 6, which involves Saudi-Arabic PWA, is an extension of the therapy protocol used in Chapter 4 with English PWA. In addition to the translation and adaptation of the therapy items to suit the Saudi-Arabic language and culture, we applied some modifications to the therapy protocol to improve its time-efficiency; however, the therapy program's main framework remained the same.

The fifth assessment tool, the System Usability Scale (SUS) (Brooke, 1996), was utilized to establish a preliminary profile of the usability and learnability of the current computer-based approach to individuals with aphasia within the Saudi population. Several factors could hinder the success of a computer-based approach in speech and language therapy, which can be attributed to: background knowledge in computer/iPad/ touch screen use, language, age/lifestyle/culture, and personal traits such as confidence and adaptability. To the best of our knowledge, the literature lacks a study investigating computer-based therapy outcomes in treating language production deficits specific to Arabic speaking individuals with aphasia, which includes Saudi-Arabic. Therefore, we anticipated that the outcomes of such a measure would supplement the literature with valuable data that would encourage further exploration and refinement of computer-based speech and language therapy in Arabic.

As outlined in the Introductory Chapter, Arabic is the fifth most spoken language globally and the native language of over 313 million people in the Middle East and North Africa (Simons and Fennig, 2018). To the best of our knowledge, a language test for adults with acquired language disorder that closely examines sentence and discourse production is yet to be published. Similarly, a sentence therapy outcome in Saudi Speaking PWA has not been reported in the literature. Therefore, this study contributes to the emerging literature relating to Arabic speaking individuals with aphasia and, specifically, the Saudi dialect. The study aims were to:

1. Report the methods adopted to produce a preliminary version of Arabic assessment tools WAB, VAST, and COAST that are linguistically and culturally adapted to Saudi-Arabic (i.e., specifically the central dialect).
2. Describe the methods used to translate and adapt the therapy materials to deliver therapy in the Saudi Arabic dialect (central dialect).
3. Adapt the therapy protocol and items used in Chapter 4 to deliver therapy in Arabic in a time-efficient manner (i.e., the minimum amount of therapy needed to achieve significant improvements compared to baseline). The purpose is to create a therapy protocol that is culturally and linguistically suitable for Saudi-Arabic individuals with aphasia.

5.2 Methods

5.2.1 Translating and adapting the outcome measures

a. Verb and Sentence Test VAST (Bastiaanse, 2003)

To produce a translated, culturally-adapted and computer-based (for ease of use) version of the VAST, we applied the same MAPI approach (Acquadro et al., 2004) used to create the Saudi-Arabic version of the WAB in 2017 (unpublished). The following is a summary of the steps we followed to achieve this goal, which implemented the linguistic validation process recommended by MAPI:

The forward translation for health interventions required bilingual health professionals who are knowledgeable of the English-speaking culture but whose native language should be the primary language of the target culture. In our study, two undergraduate students majoring in speech-language therapy, who are native Arabic speakers and fluent in English, independently carried out the first step of conducting a forward translation of the test from English to Arabic. The decision to collaborate with students was to benefit from their knowledge about the nature of language deficits in aphasia. For example, a person with a background in speech-language therapy would know that, for a repetition task, if the translated word has a different number of syllables than the original word, then the task difficulty is altered, and the produced test will no longer be equivalent to the original. Next, the principal investigator produced an agreed, reconciled version derived from the two forward translated versions of the test and reported justification for any undertaken translation decisions. Then, an independent bilingual SLT/P translator with no prior exposure to the original test, the forward translation, or the reconciliation process performed the backward translation. The translator produced an English version from the single reconciled Arabic version of the forward translation. Finally, the principal investigator checked the agreement between the backward translation and the original test. This step involved examining the differences and modifying the reconciled Arabic version of the forward translation accordingly, including the justification for any changes.

Then, we pilot tested the final Arabic version of the VAST, which included two steps that were carried out simultaneously:

- We collected experts' review from 5 qualified native-Arabic speaking SLT/P s with experience working with PWA. The SLT/P s were asked to review the VAST test's Arabic version and provide their suggestions for improvements.
- We collected performance data from 6 native Arabic healthy controls (speakers of the Saudi dialect) to investigate the Arabic VAST test's clarity, intelligibility, and appropriateness of wording and pictures to the Saudi culture. Also, we administered a short questionnaire to gather their demographical and medical history information (such as age, gender, years of education, presence of medical conditions affecting neurological/cognitive status, profession, and city/country of residence, spoken languages, spoken dialect, and their parents' dialect if different). In order to classify their Saudi subdialects, speech samples were collected from each volunteer using the cookie theft and dinner party picture stimuli.

Lastly, according to the pilot testing's outcomes, we presented a final edited version of the Arabic test to use as an outcome measure in the therapy study. Ideally, a translated and adapted version of a language test, designed to provide a numerical index of language skills, should undergo psychometric analysis to determine the efficacy of the test items and the assessment quality. However, our fundamental claim regarding the translation and adaptation of a language test in the context of this study is much more modest, since the applicability of psychometric analysis is currently a challenge in light of the lack of normative data and data collected from Arabic-speaking PWA. We aimed to create a preliminary version of the test to serve a specific purpose for the current study, which was limited to establishing a baseline of performance and measuring any performance changes compared to baseline. It is not intended for diagnostic purposes or identifying the linguistic level at which the language breakdown occurs. Hopefully, this preliminary work will set the stage for future work to develop a fully controlled and standardized Arabic version of the VAST.

Carter et al. (2005) listed several factors that need to be included in creating a culturally valid language assessment tool. It includes the involvement of native speakers, who grew up in the local area and familiar with its culture, in the design of the tool. There is also a need to pilot test assessment tools on a representative sample of the target population. It involves all aspects of the tool, such as pictures, instructions, response format, cues, and setting. Moreover, it is advisable to involve a trained native speaker of the assessment language to administer the test or assist the examiner. To avoid errors due to misunderstanding of the task requirements, it is best practice to present practice items first and provide support with cues when introducing a new task. In the current study, we aimed to incorporate all of the above suggestions into the design.

Also, we aimed to produce an Arabic version of a language test that can be used effectively to determine the language skills of a large number of PWA in Saudi Arabia. Therefore, the option of using the MSA was eliminated as it would not be the native language of any participant. Instead, the central dialect was selected for the translation since the recruitment of participants was planned to take place in Riyadh, a city in the central province of Saudi Arabia where most people speak the central dialect. However, to ensure the clarity and acceptability of the translation to speakers of the central dialect and possibly other dialects, pilot testing was administered to a sample of 6 healthy controls. It included two different Saudi dialects Najdi/Central and Hijazi/Western (i.e., 3 participants of each). The aim was to reach near ceiling scores for all items to consider the translation neutral to differences in at least two Saudi dialects. Otherwise, problematic items were excluded or replaced.

The pilot testing of the Arabic version of the VAST was administered on 6 neuro-typical controls. The sample of participants consisted of 2 males and 4 females with an age range of 21 to 64 years old. The participants' education level included 1 participant with 6th grade, 1 with a high school diploma, 2 with a bachelor's degree, 1 with a Master's degree, and 1 with a PhD. The sample can be divided into 3 participants speaking the central Saudi dialect (Najdi) and 3 participants speaking the western Saudi dialect (Hijazi). Similarly, 3 of the participants were bilinguals in Arabic and English, while 3 were primarily mono-linguals (i.e., 2 of the participants showed the ability to read, write, and speak in English; however, it was limited to working proficiency). Lastly, 5 of the participants were right-handed, and one was left-handed.

First, the pictures/items that were not relatable to the Saudi culture were eliminated (e.g., mowing, raking, skiing vs skating). For example, in task-8, an item was excluded because it represented a picture based on the princess and the frog's classical tale, which is not widely known to the Saudi population. Similarly, in task-2 the picture of a magician juggling the balls can be translated into Arabic using general terms (e.g., يلعب بالكور) but a specific equivalent of the verb juggle does not exist or not widely used. Next, the group's performance on each item in all 9 tasks was calculated in percentage. All items that did not score 100% accuracy were eliminated. In naming tasks, when the participants produced a synonym of the target word, it was considered correct(e.g., task-3: يحمل- يشيل, يطعم-يأكّل , يشعل-يولع).

Moreover, we conducted reliability testing that involved collecting test re-test data from 7 participants with aphasia. The period between the test and retest was at least 7 days. We followed a pragmatic approach to recruiting participants. Any adult participant with aphasia post-stroke (at the chronic stage) and the absence of any other neurological disorder present at the hosting hospital during the 2-week recruitment period and willing to participate were recruited. As a result, we examined the stability of performance in 7 participants in addition to the test's sensitivity in capturing changes related to treatment in 4 participants.

b. Discourse analysis

The same picture scenes used in Chapter 4 to elicit discourse samples were used in the current study. The cookie theft picture (Goodglass et al., 2001) was presented without modification as the scene it contained was clear and relatable to the Saudi culture. However, the dinner party picture sequence (Mark et al., 1983) required some modification. The sequence had to be rearranged to start from right to left in each of the 4 rows. Also, the numbers that marked each scene were replaced with Arabic numbers. Nevertheless, the English written text within the picture scenes was not modified (e.g., 'cookie jar', 'the dinner party', and 'fish n chips'). It was hypothesized that the details within the picture scenes provided enough information to elicit a comprehensive description of the event. Also, the Arabic text could distract the participants from paying attention to the visual cues.

Figure 30 Dinner Party picture sequence (Mark et al., 1983)

c. Western Aphasia Battery WAB-R (Kertesz, 2007)

One of the many efforts in translating and adapting a language test for aphasia into Arabic was produced by a team of qualified SLT/P s at King Saud University in Riyadh, Saudi Arabia. The principal investigator of the current study took part in that project; therefore, accessibility was one reason for selecting this version of the unofficial Arabic WAB. The second reason was related to the fact that it is the only project that applied a systematic approach, the MAPI method, in translation and adaptation to the best of our knowledge. Since the production of the informal Arabic version of the WAB in 2017, four qualified SLT/P s have tested the assessment with around 13-15 native Arabic speaking patients with aphasia at King Saud University Medical City KSUMC (unpublished). Based on verbal feedback from SLT/P s who tested the Arabic version, the tool proved to be effective in establishing a baseline of performance, identifying areas of weakness that required intervention, and tracking the progress of therapy gains. A noteworthy observation reported by the SLT/P s is that the aphasia classification produced by the test closely aligned with their clinical prediction of the aphasia subtype based on the informal language assessment and interview. Nevertheless, this project's work has not been published yet, and therefore data is currently unavailable.

d. COAST

The same MAPI approach used to translate and adapt both the WAB and the VAST was used to translate the COAST (Communication Outcomes after Stroke) (Long et al., 2008), which is a PROM (patient-reported outcome measure) providing an index of people’s self-rating of various aspects of their communication skills recovery after a stroke. The examiner will present the questionnaire in our study, an approach that we anticipate will prevent or alleviate possible breakdowns in comprehension by the implemented external support and explanation.

In our translation and adaptation of the COAST, we aimed to produce a dialectally neutral Arabic test to expand its applicability and include a large number of participants. In order to do so, the process involved providing four Arabic versions of the forward translation. The two translators who performed the task were asked to provide two Arabic versions, one of their native dialects and one of MSA. The pilot testing and the therapy research trials were planned to take place in Riyadh, the capital of SA, in which the majority of the population speak the central Saudi dialect. For that reason, we recruited a translator that is a native speaker of the central dialect. The other translator is a native speaker of the Hijazi Saudi dialect. As a result, the four versions of the forward translation contained two versions in MSA, one in Hijazi dialect and one in the central dialect. The consensus phase of agreeing on reconciled versions was administered by the principal investigator, who is bilingual in English and Arabic with an extensive background in three major Saudi dialects (i.e., Central/ Najdi, Eastern/ Gulfian, and Western/Hijazi). In cases where words in colloquial Arabic were biased to one region and unlikely to be understood by most of the population of another region, the words were contrasted with their equivalent in MSA to assess for clarity. The final version of the consensus was then translated back to English by an independent translator who is bilingual in Arabic and English. The agreement between the backward translation and the original test was administered by the principal investigator. The last step included collecting experts' review from 5 neuro-typical volunteers who are native Arabic speakers living in the capital of SA, Riyadh.

Table 5.1 COAST (Communication Outcomes after Stroke) items in their original English form and the version

Item No.	Original English COAST	Arabic COAST	Item No.
Practice item	In the past week or so, how well could use the phone?	خلال الأسبوع الماضي أو نحو ذلك، إلى أي مدى كنت جيد في إستخدام الهاتف؟	Practice item
1	In the past week, how well could you show that you mean Yes or No?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في إظهار أنك تقصد نعم أو لا	1
2	Nowadays, how well can you use other ways to help you communicate (e.g. pointing or writing)?	حالياً، إلى أي مدى يمكنك استخدام وسائل أخرى لتساعدك علي التواصل (مثال: الإشارة، الكتابة،...إلخ)	2
3	In the past week or so, how well could you have a chat with someone you know well?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في إجراء محادثة مع شخص تعرفه جيدًا ؟	3
4	In the past week or so, how well could you have a short conversation with an unfamiliar person?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في إجراء محادثة قصيرة مع شخص غريب؟	4
5	In the past week or so, how well could you join in a conversation with a group of people?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في إجراء محادثة مع مجموعة من الأشخاص؟	5
6	Nowadays, how well can you make yourself understood in longer sentences?	حالياً، إلى أي مدى أنت جيد في توضيح مقصدك بإستخدام جمل طويلة؟	6
7	In the past week or so, how well could you understand simple spoken information?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في فهم المعلومات الشفهية البسيطة؟	7
8	Nowadays, how well can you show that you don't understand?	حالياً، إلى أي مدى أنت جيد في اظهار انك لم تفهم الحديث؟	8
9	In the past week or so, how well could you follow a change of subject in a conversation?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في متابعة فهم الحديث مع تغير الموضوع في المحادثة؟	9
10	In the past week or so, how well could you read?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في القراءة؟	10
11	In the past week or so, how well could you write?	خلال الأسبوع الماضي أو نحو ذلك، إالى أي مدى كنت جيد في الكتابة؟	11
12	Nowadays, how well can you deal with money?	حالياً، إلى أي مدى أنت جيد في التعامل مع المال؟	12
13	How much has your communication changed since just after your stroke?	إلى أي مدى تغيرت قدرتك على التواصل منذ إصابتك بالسكتة الدماغية ؟	13
14	What do you think about your communication now?	ما رأيك بقدرتك على التواصل مع الآخرين في الفترة الحالية؟	14
15	How often does confidence about communicating affect what you do?	إلى أي مدى ثقتك بقدرتك على التواصل تؤثر على ما تريد فعله؟	15

16	Nowadays, what effect do your speech or language problems have on your family life?	مدى تأثير مشاكل النطق و اللغة لديك على حالياً، ما حياتك العائلية؟	16
17	Nowadays, what effect do your speech or language problems have on your social life?	مدى تأثير مشاكل النطق و اللغة لديك على حالياً، ما حياتك الإجتماعية؟	17
18	Nowadays, what effect do your speech or language problems have on your interests or hobbies?	مدى تأثير مشاكل النطق واللغة لديك على حالياً، ما إهتماماتك أو هواياتك؟	18
19	How often do difficulties communicating make you worried or unhappy?	الى اي مدى تجعلك صعوبات التواصل قلق أو غير سعيد؟	19
20	How do you rate your overall quality of life?	كيف تُقيّم جودة حياتك بصورة عامة؟	20

e. System Usability Scale SUS

As discussed above, the need for a quick and effective assessment tool has been identified amidst the research trial in order to investigate the usability and learnability of the current computer-based approach in relation to PWA with a Saudi-Arabic background. For that purpose, we have selected SUS for several reasons, including its accessibility and the fact that it has been released almost a decade ago and underwent extensive research to assess its effectiveness. An empirical study by Bangor et al. (2008) investigated around 2,300 individual surveys (i.e., 200 studies) that implemented SUS. The author concluded that " SUS has proven itself a valuable and robust tool in helping assess the quality of a broad spectrum of user interfaces" (Bangor et al., 2008, p.593) and that "SUS score can and does provide a very useful metric for overall product usability" (Bangor et al., 2008, p.591). The SUS is composed of 10 items in which the odd-numbered items (i.e., 1,3,5,7,9) presents positive statements and the even-numbered items (i.e., 2,4,6,8,10) presents negative statements. The participants were instructed to indicate their answer by selecting one from a five-point Likert-scale that ranges from least agree to most agree. The scoring guideline involves subtracting 1 point from the raw score of each even-numbered items (i.e., the raw score could range from 1-5). For the even-numbered items, the achieved score is the result of subtracting the raw score from 5. The next step is to calculate the sum of all the scores (i.e., for the 10 items) and multiply it by 2.5. Finally, the maximum achievable score is 100, which is not a percentage. A recent study by AlGhannam et al. (2018) presented an Arabic version of the SUS. However, the translation implemented MSA.

For time efficiency, the steps of translating and adapting the SUS to Saudi-Arabic has been simplified. The primary investigator, a bilingual SLT/P (i.e., Arabic and English), produced a single forward translation version. The translation involved producing two Arabic versions for each item, one in MSA and the other in the Saudi dialect (i.e., Central dialect). An expert review and feedback from two SLT/P s with experience in assessing and delivering speech and language therapy for at least 4 years were obtained. The review included assessing sentences' clarity, appropriateness of the selected vocabulary, its alignment with the original English version, and whether it tests the usability and learnability of the therapy program. The participants were asked to provide suggested alternatives to the phrasing of questions or word choice when applicable. The primary investigator also conducted the consensus.

An additional step was implemented to ensure the clarity of the questionnaire. The task instruction included a statement that specified to the participant that the questionnaire is about using SentenceShaper App on the iPad to complete their weekly home practice. Also, the scale direction was changed to right-to-left to match the direction of the Arabic text. Also, the font was emboldened, and the selected text size was set at 14 to direct the participant's attention to the main sentence in MSA. While the sentence in Saudi colloquial Arabic, below the main sentence in MSA, was presented with quotation marks with no emboldening of the font and a size of 10. The aim was to focus on the main MSA in answering the question since it is the only official written form of Arabic, and all participants were literate. Then to use the colloquial sentence for further clarification and confirmation of understanding. The neuro-typical participants were asked to fill the questionnaire independently, while PWA were assisted by the examiner, who explained the task and read the questions.

To better predict the possible causes of a negative or positive outlook about the computer-based therapy program, the participants were asked to answer 3 independent questions that preceded the SUS introduction. The first two questions were general and required the participant to self-evaluate their ability to communicate with others using language in general and using sentences specifically. The purpose was to prime the participants to use the five-point Likert-scale. The third question was specific and required the participant to rate their proficiency in using iPad Apps or smartphone Apps. The question also highlighted that this applies to their skills in general and before participating in this current study. The purpose of adding this question was to identify any possible correlation between background knowledge and confidence in using Apps on smartphones and iPads and SUS scores. Drawing a hypothesis could be useful in case the majority of participants achieved low SUS scores.

Table 5.2 System Usability Scale SUS items in their original English form and the Arabic version

Item No.	Original English SUS	Saudi Arabic SUS	Item No.
1	I think I would like to use this system frequently	**أعتقد أنني سأرغب باستخدام هذا البرنامج بكثره** " احس اني باستخدمه كثير"	1
2	I found the system unnecessarily complex	**لقد وجدت البرنامج معقد بشكل غير ضروري** "حسيته معقد أكثر من اللازم"	2
3	I thought the system was easy to use	وجدت أن البرنامج كان سهل الاستخدام	3
4	I think that I would need the support of a technical person to be able to use the system	**أعتقد أنني سأحتاج إلى دعم مختص تقني حتى أتمكن من** **استخدام البرنامج** " بحتاج احد يساعدني عشان استخدمه"	4
5	I found the various functions in this system were well integrated	**لقد وجدت أن الوظائف المختلفة في هذا البرنامج متكاملة بشكل جيد** "كل الازرار الموجوده بالبرنامج مفيده و تكمل بعضها"	5
6	I thought there was too much inconsistency in this system	**اعتقدت أن هناك الكثير من التناقض في هذا البرنامج** " كثير من الازرار بالبرنامج مالها فايده و ما تساعد"	6
7	I would imagine that most people would learn to use this system very quickly	**أتصور أن معظم الناس سيتعلمون استخدام هذا البرنامج بسرعة كبيرة** "اغلب الناس بيتعلمون يستخدمون البرنامج بسرعه"	7
8	I found the system very cumbersome to use	**لقد وجدت أن البرنامج مرهق للغاية في الاستخدام** " صعب الاستخدام"	8
9	I felt very confident using the system	**شعرت بثقة كبيرة باستخدام البرنامج** "حسيت اني واثق اني قادر استخدم البرنامج لحالي بدون مساعده"	9
10	I needed to learn a lot of things before I could get going with this system	**كنت بحاجة لتعلم الكثير من الأشياء قبل أن أتمكن من استخدام هذا البرنامج بسهوله** " احتجت اتعلم بعض المهارات قبل ما اكون جاهز اني استخدم البرنامج لحالي بسهوله وسلاسه"	10

5.2.2 Adapting and translating the therapy protocol and materials

a. SentenceShaper®

As mentioned previously, using computerized therapy approaches in speech and language therapy for Arabic speakers can be problematic. The assessment and treatment should implement the subject's native dialect. To the best of our knowledge, a system for creating automated speech in a specific Arabic dialect, including Saudi-Arabic, is yet to be invented. Also, a system that recognizes the input speech in a specific Arabic dialect produced by the participant does not exist. As the option of using programs that utilize automated speech has been eliminated, we aimed to adapt an existing computer software that allows the recording of instructions and prompts in a specific Arabic dialect.

The SentenceShaper® (Linebarger et al., 2000) software provided the most convenient platform to deliver our theory-driven impairment-based language therapy. It does not rely on reading and writing ability to interact successfully with the program. Instead, it provides audio verbal prompts as well as many symbols and picture stimuli. It also had several built-in workbooks that target many linguistic levels (e.g., simple SVO sentence production in a picture description task vs open-ended questions that prompts discourse production) and can be customized to increase or decrease the level of support (e.g., vocabulary cues). It also had an import and export feature, where the SLT/P can custom build a workbook with a specific therapy goal/ topic and transport it to the participant's copy of the program. Therefore, any treatment approach/ material can be designed and imported into the program without restrictions. The verbal instructions and therapy prompts could be translated into any language by following a set of editing processes and recording the prompt. Similarly, the text of the written cues can be translated into any language.

For all the above reasons, SentenceShaper® provided the most flexible computer-based medium to deliver a custom-designed theory-driven therapy program that we have planned for this study. It also permitted achieving the goal of delivering therapy in a specific Saudi-dialect through the prompt recording feature. Moreover, SentenceShaper® solves the issue of the current unavailability of a dialect-sensitive speech recognition feature required in therapy, which would be required if one were aiming to determine the accuracy of the participant's productions and provide feedback. We have been able to overcome this shortage but from a different angle. Instead of providing immediate feedback to the participant about their sentence/discourse production accuracy, the program gives the participants the chance to assess their own production. It is achieved through the playback feature, which allows the participant to self-monitor their productions and self-correct the errors through the multiple editing features (e.g., reorder, add, delete, expand). In this case, the participants will be relying on their sentence comprehension skills. However, in cases where the participants' skills in recognizing their errors in sentence production or their ability to arrive at an accurate correction are poor, another option has been provided. That is the option of having the target sentence modelled in a recording, in which they can use as a reference point. This target production is located in a specific button, and the participant is instructed to play it as needed and only after all independent attempts have failed (see Appendix 3).

The above-described strategies and modifications to the original program that we have invented were implemented to meet the requirements for a computer-based approach to aphasia therapy. Also, to overcome some of the limitations we faced in delivering computer-based therapy in the Saudi-Arabic dialect. Nevertheless, these changes provided many advantages. They fostered independence in language practice. Also, rather than having an external feedback system, it promoted self-monitoring and self-correction of speech production and the internalization of self-cueing strategies.

The only issue we initially faced in using SentenceShaper® to deliver language therapy in Arabic was the fundamental difference between English and Arabic in the direction of written text. It meant that in the original software, the play button which activates the playback of a sentence would play the snippets (i.e., individually recorded words that constitutes a sentence) from left to right, resembling the written text in English. When the produced sentence is in Arabic, the snippets' order in the sentence row should resemble the Arabic written text that starts from right to left. Therefore, the play button's original programming would play the Arabic sentence backwards, from left to right, and distort the meaning as a result. Also, training the participants to produce the sentence in Arabic, but following the English direction in ordering the snippets in the sentence row would be extremely confusing.

Fortunately, we were able to contact and request the collaboration of Dr Marcia Linebarger, one of the originators and designers of SentenceShaper®, who provided all the programming and support required to create an Arabic compatible version that allowed the snippets in the sentence row as well as the story row to play from right to left. Moreover, a new IOS version of SentenceShaper® that can be easily downloaded on an iPad was created. the IOS version has been selected for this study due to the transportability and accessibility of a touch screen to most of the participants,.

b. Therapy materials

While we have adopted an extensive process to translate and adapt the primary and secondary outcome measures into Saudi-Arabic dialect through the MAPI approach, the objective of translating and adapting the therapy materials was much simpler. One of the major differences between the two is that an assessment tool involves testing the participants' comprehension. Accordingly, many steps had to be taken to ensure the validity of the translated testing materials and their cultural relevance. Doing so minimizes the risk of reaching inaccurate conclusions that will underestimate the participants' skills or mask the existing language profile. On the other hand, our study's therapy materials serve as stimuli to prompt the participants to produce sentences and discourse using words from their vocabulary repertoire and scenarios from their own experiences. Therefore, any sentence they produce will be assessed according to its completeness, informativeness, grammaticality, and its relatedness to the presented stimulus. As a result, the process of translating and adapting the therapy materials included: therapy task instructions, picture stimuli, and target verb stimuli (i.e., for sentence construction in a picture description task and the VNeST tasks).

For the therapy task instructions, the primary investigator, a native speaker of the Saudi-Arabic dialect, specifically the central dialect, and fluent in English, performed the task of translation and adaptation. The translated task instructions and two practice items for each therapy task were pilot tested with a neuro-typical participant to ensure its clarity and appropriateness. Similarly, the primary investigator performed the picture stimuli' exclusion and inclusion decisions based on their cultural appropriateness. Lastly, the verbs chosen as therapy targets for the VNeST were selected based on their frequency. A list of 30 verbs common in daily life interactions was selected for the VNeST therapy tasks (see Appendix 4). However, for practicality and time-efficiency, the picture stimuli in the picture description task were selected from the built-in workbooks in SentenceShaper. Therefore, the verbs targeted in therapy were bound to the available verbs in those workbooks.

Moreover, to support the participants independent home practice, additional steps were taken to provide them with built-in vocabulary cues for the VNeST task. In the VNeSt workbook within the SentenceShaper® App, the target verb is presented with 10 obscured buttons that provide vocabulary cues of 5 subjects and 5 objects which may stimulate sentence production as needed. Also, in VNeST, when the therapy target is sentence expansion, a number of side buttons with visible text and audio recording (i.e., played upon activation) are provided. These buttons serve as vocabulary cues to aid the production of an accurate preposition for the expanded sentence. In order to generate vocabulary cues in Arabic that are compatible with the Saudi dialect and culture, two undergraduate students from King Saud University KSU in Riyadh (location of the planned research trials and participant recruitment) were involved in the process. The two students and the primary investigator-generated a list of sentences with 30 different verbs (See Appendix 4). It included 5 different sentences for each verb, with no repetition of subjects or objects. The primary investigator then selected the most convenient 5 sentences, ensuring variety, for each target verb. The subjects and objects of these sentences were then recorded as verbal prompts in the VNeST workbooks.

The verbal instructions and prompts were recorded by the primary investigator and an undergraduate student volunteer from KSU with experience in recording radio advertisements. Additionally, the translated text on the buttons was performed using Word Document and Paint program to write, copy, and paste. The text was then attached to the designated button in SentenceShaper® using the upload feature. Finally, the edited Arabic version of the workbook was then exported to google drive and imported into the participants' iPads using the "import workbook" feature.

Therapy workbooks

For the current study that involved Saudi Arabic speaking individuals with aphasia, three built-in workbooks were selected from the IOS version of SentenceShaper®. It constituted workbook-1 Simple sentences, workbook-5 Sentences with adjectives and prepositions, and Verb Networking Level 2. They all underwent the process of translation and adaptation, as mentioned previously.

For the picture description task, the program's existing pictures were used. The verbal instructions in the introductory pages, the verbal prompt that presents the target production (play prompt button), and the vocabulary cues in the side buttons (subject, verb, object) were all replaced with Arabic recordings implementing the Saudi-central dialect.

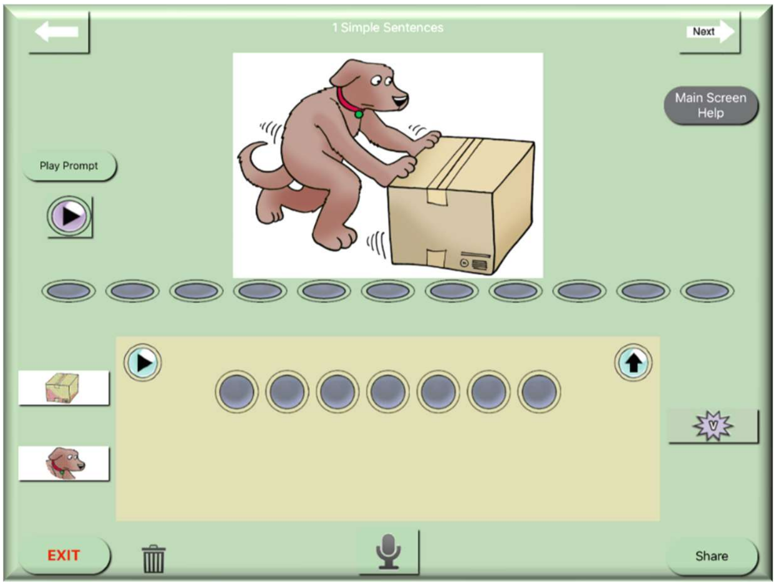


Figure 31 Example of a workbook page, Level 1

For the VNeST workbooks, the side buttons that contained Wh-question prompts, and preposition cues were replaced with Arabic text and recordings.

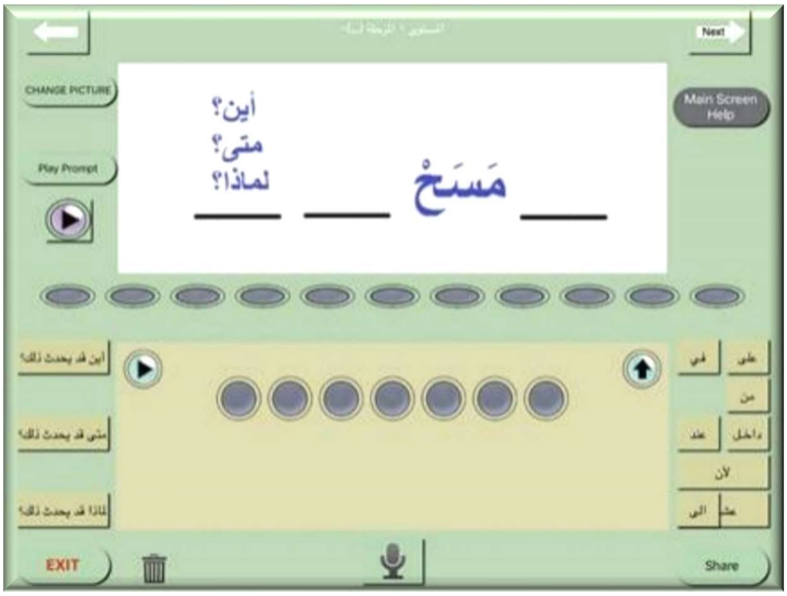
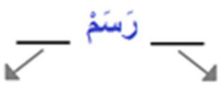


Figure 32 Example of a workbook page, Level 2 – Phase 2

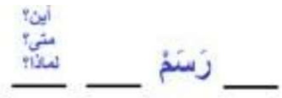
Moreover, in the VNeST workbook, an Arabic text was also added to the display picture. By using the “change picture” option, the following pictures were added to each screen:

Step one: verb رَسَمَ (past tense of the action draw)

Step two: a prompt to map the verb with a subject and an object



Step three: a prompt to expand the sentence by answering Wh-questions



Step four: a prompt to rehearse the three constructed sentences from memory to promote consolidation

For the therapy study with the Arabic group, we have chosen the workbook “Verb Networking Level-2” to implement VNeST exercises. In this workbook, written text on the vocabulary cue button was obscured (i.e., subject examples were in green and object examples in blue). On the other hand, the excluded workbook “Verb Networking Level-1” had visibly written text on the buttons, which indicated example subjects and objects. The premise behind selecting the obscured vocabulary cues on the buttons is that this layout of the page would minimize visual distraction and encourage participants to rely on their memory of events and vocabulary repertoire. It was also helpful for fading the support/cues.

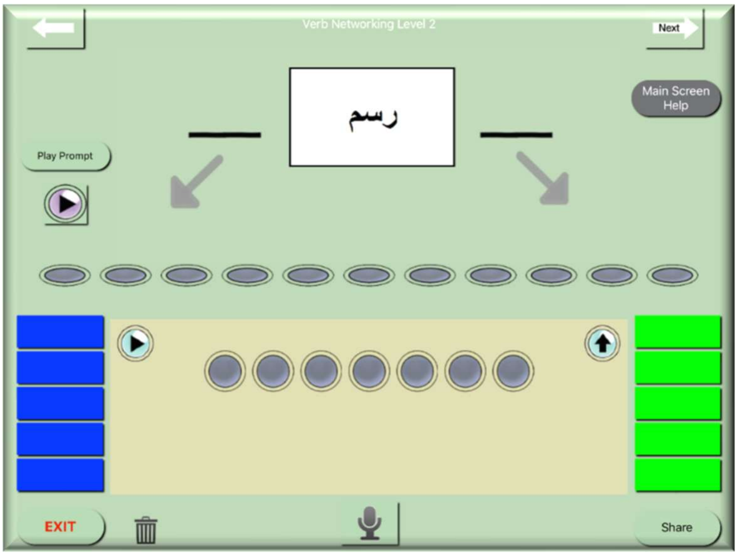


Figure 33 Example of VNeST task, Level 1-Phase 2

One of the other modifications we have made was to disable the autoplay of prompts. In the instruction prompts, one of the reasons was to minimize the frustrations that the participant might develop from listening to the same instructions repeatedly, even after mastering the task. For the "play prompt" button that presents a verbal illustration of the target production, the reason for disabling the autoplay was to minimize its influence. The participant is encouraged to practice independently and draw in self-cueing strategies. However, the button remains available for a final check to ensure their productions are comparable to the target in terms of completeness, grammaticality, and informativeness.

5.3 Results

5.3.1 VAST

After the pilot testing on neuro-typical adults and eliminating items that did not achieve 100% accuracy, further exclusion of items was implemented, the sentence production task-8 was reduced to 20 items (i.e., first 20 on the list) to keep the testing time within 2 hours. This total number of items match the equivalent test used with the English group in chapter 4. Furthermore, task-4 and task-6 (i.e., fill-in infinitive verbs in sentences and fill-in finite verbs in sentences, respectively) showed an inconsistent response in all trials in terms of the verb tense it elicited. The participants' productions alternated between verbs in the present and the past tense. Since verb morphology was not an outcome measure we focused on in this study, we eliminated one task (task-6) and focused on assessing the participants' ability to produce a verb in a sentence regardless of the selected tense. The rest of the items/tasks were kept the same.

Table 5.3 Number of items in the final Arabic version used in Chapter 6

Task No.	Subtest	The final number of items (Arabic version)
1	Verb comprehension	48
2	Semantic association	19
3	Action naming	41
4	Fill-in infinitive verbs in sentences	20
5	Object naming	45
6	Fill-in finite verbs in sentences	18
7	Sentence comprehension	45
8	Sentence production	37
9	Grammaticality judgment	30

With the Saudi speaking PWA, the test-retest scores' stability was explored by assessing the correlation between scores at baseline 1 and baseline 2 using Spearman's rank correlation coefficient. The presence of a statistically significant correlation would indicate stability of performance at two time-points and support the reliability of the test. As presented in Table 5.7 (below), the correlation results can be divided into two groups. The first group is composed of 5 out of 8 subtests that showed statistically significant P value ($p<0.05$): verb comprehension (r_s =.937, n=7, p=.002, two-tailed), action naming (r_s =.929, n=7, p=.003, two-tailed), object naming (r_s=.919, n=7, p=.003, two-tailed), fill-in verbs in sentences (r_s =.943, n=6, p=.005, two-tailed), sentence production (r_s=.900, n=5, p=.037, two-tailed). The second group is composed of 3 out of 5 subtests: sentence comprehension (was approaching significance) (r_s=.721, n=7, p=.068, two-tailed), grammaticality judgement (r_s = .721, n=6, p=.106, two-tailed) and semantic association (r_s = .516 , n=6, p=.294, two-tailed). We also considered whether the correlation r values were high, medium or low with reference to Akoglu (2018). According to this method, there were strong positive correlations in all subtests except the semantic association subtest, which was considered weak (Dancey and Reidy, 2007). A replication of the study to include a larger sample size will be required to make firm conclusion regarding the reliability of the test.

Table 5.4 Therapy participants' total comprehension and total production scores at baseline 1 and baseline 2.

Participant	No. and initials	Total comprehension		Total production	
Max. score		142		297	
		B1	B2	B1	B2
1	ND	121	117	120	112
2	NS	115	101	156	180
3	SA	132	136	254	273
4	SM	131	137	277	278
	Mean	124.75	122.75	201.75	210.75
	SD	8.18	17.17	75.65	79.78

Table 5.5 Spearman's rank correlation analysis between test and retest scores of the VAST comprehension and production subtests

Subtest	Total comprehension		Total production	
	r_s	p	r_s	p
N=4	0.80	0.20	1.00	**0.00**

Table 5.6 Participants' performance on VAST subtests at baseline 1 and baseline 2 (raw scores)

Participant No. and initials		Verb comprehension		Sentence comprehension		Grammaticality judgment*		Action naming		Object naming		Fill-in verbs in sentences*		Sentence production**		Semantic association*	
Max. score		48		45		30		41		45		19		192		19	
		B1	B2	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2	B1	B2
1	AA	35	28	16	20	16	N/A	0	0	0	0	0	0	0	0	N/A	10
2	FA	46	46	33	42	21	25	11	11	23	28	N/A	N/A	N/A	N/A	18	18
3	HZ	43	44	28	35	25	19	23	30	36	31	14	14	N/A	N/A	9	14
4	ND	44	46	41	38	19	14	7	13	26	25	2	6	85	68	17	19
5	NS	42	37	35	31	21	19	26	22	28	32	16	15	86	111	17	14
6	SA	47	48	40	42	27	28	32	31	40	43	17	19	165	180	18	18
7	SM	45	47	44	45	24	26	37	40	44	43	18	18	178	177	18	19

N/A= not available *N=6 **N=5

Table 5.7 Spearman's rank correlation analysis between test and retest scores on VAST subtests

Subtest	Verb comprehension		Sentence comprehension		Grammaticality judgment*		Action naming		Object naming		Fill-in verbs in sentences*		Sentence production **		Semantic association *	
	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p	r_s	p
N=7	.937	.002	.721	.068	.721	.106	.929	.003	.919	.003	.943	.005	.900	.037	.516	.294

*N=6 **N=5

5.3.2 WAB

The same Arabic version of the WAB produced by a team of SLT/P s at KSU in 2017 was used in the current study, with no modifications. It had been used by four SLT/P s within the organization for the past three years to test at least 15 participants with aphasia. The information was based on verbal reports by the SLT/P s as access to medical files was not available. They also reported their satisfaction with the outcomes of using this tool.

5.3.3 Speech Sample Elicitation (cookie theft and dinner party narratives)

As mentioned above in the methods section, the picture stimuli' adaptation was limited to the dinner party picture sequence. It involved switching the direction of the picture sequence to right-to-left to match it to the Arabic direction. Both the cookie theft picture and the adapted version of the dinner party picture stimuli successfully elicited speech samples from all four participants with aphasia (see Chapter 6). The pictures were clear, and the participants did not show any difficulties in comprehending the illustrated events. The relevance of the stories the participants' produced in response to the stimuli supported this conclusion.

In the current study, we analyzed the discourse produced by neuro-typical controls across English and Arabic using the cookie theft and dinner party stimuli. For the English group, we referred to data published in the literature in two studies (Alyahya and Druks, 2016; Menn et al., 1994). For the Arabic group, data were collected from 6 healthy controls with an age mean of 46.3 (SD=19.2) and with an educational background of (one participant with a high school diploma, two with a Bachelor's degree, one with a Master's degree, one with a PhD, and one with a 6^{th}-grade education). They are all native Arabic speakers (i.e., three native speakers of Najdi (central) dialect and three native speakers of Hijazi (western) dialect). Nevertheless, due to the complex nature of morphology in Arabic, achieving consistency in implementing the same guideline used in the English studies was not possible.

Table 5.8 represents the average of the total number of words produced in a discourse in response to the cookie theft and dinner party stimuli in three different healthy controls groups. For the two English studies, the guidelines listed in (Menn et al., 1994) were implemented to arrive at the score of the total number of words, which was defined as "the total number of words, word-like paraphasia, and neologisms (e.g., 'chookie, epistowdy') that the patient produced in responding to the request to provide an oral narrative description of the picture."(Menn et al., 1994, p.346).

Menn and colleagues (1994) analyzed the speech samples of healthy controls in response to the cookie theft's picture description task. The data collected from 14 healthy controls aged 51-72 showed that as a group, the average total words produced in discourse was 79.9 (SD=30.8) with a range from 46 to 150.

A more recent study by Alyahya and colleagues (2020) presented the total number of words in a discourse produced by 20 healthy controls (i.e., age mean=68.85, SD=8.47), and education mean= 14 (SD=2.8)) in response to cookie theft and dinner party stimuli. The cookie theft discourse's average score was 102.95 (SD=47.81) with a range from 53 to 243, while The dinner party average score was 253 (SD=136.77) and a range from 99 to 672 for the same group.

On the other hand, the scoring rubric for the Arabic discourse required adaptation. To estimate the average number of total words produced, we calculated the information units in each discourse. It included nouns, verbs, adjectives, adverbs, conjunctions (و), prepositions (على طاولة /للمطعم), separate and attached pronouns (هي اللي /خلته يطلع), and verb tense marker سوف (i.e., equivalent to "will" in English) which in the Saudi dialects appear as a prefix (بياخذ حيطيح /كوكيز). However, noun markers such as definite/indefinite (e.g. articles البَيْت/بَيْت) markers, gender markers (طَالِبَة/طَالِب), and singular/dual/plural markers (جَامِعَات/جَامِعَتَان/جَامِعَة) were not included. Similarly, a verb-subject agreement was not captured in the current analysis. Our reason for counting the former constituents but not the latter was that these constituents appear as independent units in English (e.g., prepositions, pronouns, modal verbs, conjunctions, etc.,) but in Arabic, they can be either separate or attached to a word. On the other hand, gender markers and dual/plural markers are more common in Arabic than in English.

The discourse samples were collected from 6 healthy controls using the cookie theft and dinner party stimuli. The findings showed that the cookie theft stimuli elicited an average of 54.2 (SD=27.1) words and a range from 22 to 88. In comparison, the dinner party stimuli elicited a mean of 102.7 (SD=45.2) words with a range from 41 to 148.

Overall, while there was no expectation that these elicitation stimuli would produce identical numbers of words across the two languages, it was interesting to notice the substantial discrepancies evident in the length of monologic discourse samples and numbers of words across Arabic and English (cookie theft: 54.2 vs 102.95, dinner party: 102.7 vs 253 respectively). The Arabic group results for cookie theft and dinner party stimuli were almost half the English group's scores in the Alyahya et al. (2020) study. However, the data also confirmed that these stimuli could be expected to elicit extended monologic samples of discourse from Arabic people with aphasia.

Table 5.8 Average of the total number of words produced in a discourse by three groups of healthy controls in response to the cookie theft and dinner party stimuli:

Healthy controls	Cookie theft Mean (SD), Range	Dinner Party Mean (SD), Range
Arabic n=6	54.2 (27.1), 22-88	102.7 (45.2), 41-148
English n=20*	102.95 (47.81), 53-243	253 (136.77), 99-672
English n=14**	79.9 (SD=30.8), 46-150	N/A

*(Alyahya et al., 2020) **(Menn et al., 1994)

5.3.4 COAST

Table 5.9 below summarizes the item-specific feedback obtained from the neuro-typical volunteers regarding the questions' clarity and their applicability to the Saudi culture. Since the questions are specific to individuals with aphasia, the pilot testing was not administered on neuro-typical volunteers. It was postponed until the research ethics approval was obtained from the hosting hospitals in Riyadh, and the principal investigator was able to recruit participants to the study. This phase took place several months following the process of translating and adapting the current study’s materials. Based on the experts’ feedback, modifications were made to the final Arabic version of the COAST. It included rephrasing the questions, being more specific with the questions, and changing some of the pictures (e.g., changing a picture that shows UK money currency and replacing it with Saudi currency).

During the study's recruitment phase, the Arabic version of the COAST was pilot tested on two participants. A participant with aphasia post-stroke and her caregiver daughter were asked to independently answer the COAST questions. The examiner, the principal investigator, administered the test and provided extra support such as rephrasing and examples to explain the question further when the participant showed signs of misunderstanding, in addition to her caregiver's help which is considered clarification from a familiar person. However, several comprehension breakdowns still occurred throughout testing. The caregiver was asked to judge the questions' clarity, and she reported that the questions were vague and could have many interpretations, or at least follow-up questions are still required to clarify some of the points. Based on this initial pilot testing findings, it was decided that the translation and adaptation of the COAST require further refinement. Unfortunately, the research trials were scheduled to start within a week which was not sufficient to modify and improve the current version. Therefore, the tool had to be excluded from the current study.

Table 5.9 The participants' feedback (n=6) on each item of the COAST.

Initials	AJ	AR	GA	SR	WM	YK
Item						
1	R					
2	R			R - SP		
3	R				Pic	
4	R				Pic	
5	R		CV	SP	Pic	
6	R		SP		SP	
7	R				Pic	
8	R				Pic	
9	R			SP	SP	
10	R				Pic - R - SP	
11	R				R - SP	
12	R - SP	R - SP				
13	SP					SP
14						R - SP
15			SP	R	Pic	
16				R	Pic	
17				R	Pic - SP	
18				R	Pic	
19		R			Pic	
20					Pic	

R= Rephrase, SP= Unclear, requires specification, CV= Unsuitable due to cultural variation, Pic= Picture choice is unsuitable. It was recommended to select pictures that are more relatable to Saudi culture

Upon reflection, we have identified several topics/items that are not very well rooted in Saudi culture and could benefit from modification and adaptation. Table 5.10 illustrates the principal investigator's comments on each item's relatability to the Saudi culture and lifestyle.

Table 5.10 The principal investigator's comments on the relatability of each topic/item to the Saudi culture and lifestyle

Item	Comments
exercise	The term " using the phone" is very vague and could benefit from some context. For example, providing a scenario such as: "can you give direction to a delivery person on the phone and answer their questions without help from others?" Start by asking the participant if they usually use phone/mobile to talk to someone on regular bases? According to their answer, we can elaborate if they can communicate as well with unfamiliar callers?
1	The term "indicate yes or No" is also vague and requires context. For example, "if someone asks for your preference such as would you like some tea? and your answer is yes, can you express that answer adequately or do you sometimes get confused as to what to say?". The examiner would then follow up, "How about in other situations such as.. etc.,?".
2	Similarly, it requires some context. For example, "in cases when you can't express yourself using language alone, what else would you do to assist you?". The examiner would allow some time for the participant to provide an answer, then prompt expansion "How about writing, gesturing, etc.,?". When the alternative ways of communication have been identified, the examiner would then ask how well the participant can use these aids.
3	The question is clear but could benefit from examples such as siblings, close friends, etc.
4	To the best of our knowledge, in Saudi Arabia, people would not usually initiate a short conversation with a stranger unless there is a purpose (e.g., asking direction, offering help etc.). Therefore, this question may seem strange. An alternative would be to give scenarios (e.g., an appointment with a health practitioner) or specify an intention of the conversations (e.g., inquire, offer help, etc.,). Also, to specify if the participant initiates the conversation.
5	Likewise, the familiarity of the "group of people" and the conversation's intention is important in determining the motivation and level of interaction. Therefore, more specificity will be beneficial.
6	Maybe a choice of two would be helpful. For example, "do you usually use long or short sentences when you talk?"
7	Similarly, examples would be beneficial, such as the pharmacist instructions about a prescription.
8	A lead-in question would be helpful. For example, "if someone said something and you didn't understand, what would you do?" then a follow-up question "How well/often do you use this strategy?".
9	A scenario would assist the participant's comprehension. For example, "when someone is talking about something, and then they suddenly changed the topic they are discussing, can you always recognize this change in the topic? Or do you sometimes get confused? How often".
10	More specificity would be helpful, such as magazines, legal documents, emails, novels, or text messages.
11	Likewise, more specificity is needed. For example, writing notes, writing instructions, writing a text message or email.
12	Clarification is needed. Does it include counting, buying from a store and calculating the returned change, investments, or savings etc.,
13	The word "communication" is broad; more specificity would be helpful. For example, "do you feel you can better express your thoughts and ideas nowadays compared to when you first had the stroke and where hospitalized? How about your ability to understand people when they talk to you?". Then follow it with the original question as a follow-up. "How much has your communication changed since just after your stroke?"
14	More details such as "Your communication in general. Your ability to express your thought and needs and understand others speech without assistance".

15	The original pictures show a man in a grocery pushing a shopping cart, a picture of a lady speaking on the phone, and a person riding a bus. It is common in Saudi Arabia that the elderly are taken care of by their grown-up children; therefore, shopping for groceries is rarely part of their lifestyle. Also, public transport, especially busses, is rarely used in Saudi Arabia.
16	Requires context and examples
17	Requires context and examples
18	Could benefit from lead-in questions such as “what are hobbies and interests?”. Also, the pictures could be replaced by games and hobbies that are more common in Saudi.
19	Lead-in questions could help clarify the question such as “ does your difficulties in communicating affect you?” then “ how often does it make you worried or unhappy?”
20	Adding context and examples to “overall quality of life” could be helpful.

Overall, the suggested adaptation requires more contexts, scenarios, examples and lead-in questions to improve the items' clarity to a Saudi-Arabic speaker. The impression was derived from experts’ review of 6 neuro-typical Saudi individuals with an age range of 21-64 years old and a range of educational backgrounds and the principal investigators’ background knowledge and experience. The pilot testing, which included PWA and her caregiver, further confirmed the need for these adjustments to the approach in presenting questions. However, a clear limitation of a new version that implements these changes would be the complexity and ‘wordiness’ of questions. The increased processing overload on the participant, especially those with language comprehension deficits, could be counter-productive. It will also increase the time required to complete the questionnaire, which is a drawback in a busy clinical setting.

5.3.5 SUS

The SUS tool has been implemented in this study to collect feedback from three different groups. The first group was composed of 5 participants who participated in at least 3 weeks of the therapy program; this included in-person sessions and home practice. The test was administered at two-time points, during the course of therapy (between week 4 and week 5 of therapy) and after the program's conclusion (i.e., between week 8 and week 9). The purpose was to examine any changes in the total score with increased familiarity and adaptation. It was observed with one participant, NS, whose score increased from 50 to 85 out of 100 within 3 weeks. He did not feel confident in practising independently and required the help of his spouse or daughters with home practice. The rest of the group’s scores showed either stability such as ND or a drop such as SM (2.5 scores) and ME (10 scores). The average score of 4 out of 5 participants’ ranged between 72.5-100, which, according to Bangor et al. (2008), is considered good, excellent, and best imaginable. Referring back to the outcomes of the 3 independent questions that preceded the SUS questionnaire (see P.186), we found that NS, who was the only participant with a poor SUS score at the first testing time-point, was also the only participant who self-rated his ability to independently use iPad and smartphone apps with a score below 4 out of 5. We may infer that the increase in scores is associated with increased familiarity and adaptability to the program.

The second group of participants was composed of the participants' caregivers, in this case, their spouses. Three participants presented an average score of 82.5, 85, and 88.75; all considered good to excellent.

The third group of participants represents SLT/P s with different background experience; nevertheless, all were bilinguals in Arabic and English. Participants ALH and NGH, whom each have experience in computer-based language therapy for Arabic speaking individuals with aphasia, scored 82.5 and 87.5, respectively. GHA and HZ, who have experience delivering paper-based assessment and therapy sessions in Arabic speaking individuals with aphasia, scored 77.5 and 85. Lastly, AN, who scored 65, is a new graduate SLT/P with no experience in delivering independent assessment and therapy sessions to individuals with aphasia. It is noteworthy to mention that the SLT/P s' input, including the SUS, was taken following their observation of a complete therapy session delivered by the primary investigator. Although they had the chance to examine the App and workbooks' content briefly, they could not download it to their own devices to access it for a longer period for a better investigation. At that time, only the pilot version of the App was available.

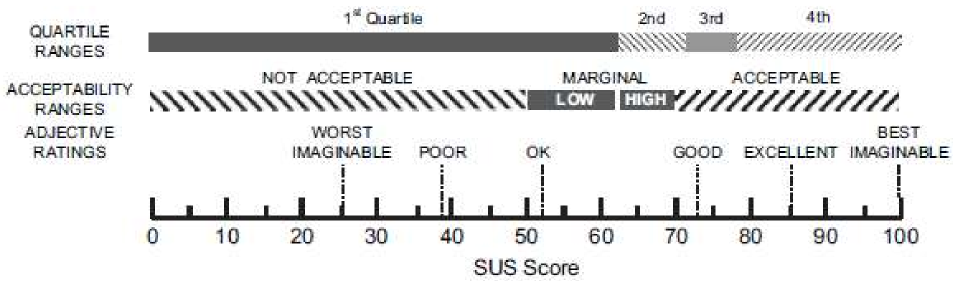


FIGURE 13 A comparison of mean System Usability Scale (SUS) scores by quartile, adjective ratings, and the acceptability of the overall SUS score.

Figure 34 System Usability Scale (SUS) (Bangor et al., 2008, p.592)

Table 5.11 Participants' SUS scores (maximum score= 100 points)

Participant's initials	Score 1	Score 2	Average score
NS	50	85	67.5
SA	100	N/A**	
SM	90	87.5	88.75
ND	72.5	72.5	72.5
ME	100	90	95

*Average duration between the two testing time-points is 21 days. **N/A= not available

Table 5.12 Care-givers' SUS scores (maximum score= 100 points)

Participant's initials	Score 1	Score 2	Average score
NS spouse	80	90	85
ND spouse	80	85	82.5
ME spouse	87.5	90	88.75

*Average duration between the two testing time-points is 21 days.

Table 5.13 **SLT/P** s SUS scores (maximum score= 100 points)

SLT/P's initials	Score
GHA	77.5
ALH	82.5
NGH	87.5
HZ	85
AN	65

5.4 Discussion

As mentioned in the Introduction, the need for language assessment tools suitable for testing Arabic speaking individuals with language disorders is clear. Some of the most prominent limitations that hinder the standardization of translated and adapted language assessment tools are the dialectal variations in Arabic and the lack of normative data. The adoption of the MAPI approach had been useful in creating translated and culturally adapted versions of widely known language tests, such as the VAST and the WAB, in a systematic way. However, the validity of some of the subtests in the Arabic versions remains questionable regardless of the fact that controls achieved ceiling scores. The repetition subtest in the WAB provides an example of this. Although a great effort had been made to match the number of syllables and words in the translated version, which sometimes required replacing the sentence with a novel example in Arabic, the hierarchy of difficulty in phoneme production could not be matched. For example, the emphatic articulation placement of consonants is more dominant in Arabic (ط /tˤ/, ظ /ðˤ/, ص /sˤ/, ض /dˤ/, ق /q/) than in English (/k/ and /g/),therefore, determining the level of difficulty of phonemes proved to be a challenge. This is especially true with the current lack of normative data on phoneme acquisition in the Saudi central dialect. A repetition task could provide us with useful information about the presence or absence of apraxic or dysarthric characteristics in Arabic speakers. In light of these difficulties, a novel assessment that is developed based on a normative data of consonant acquisition in a specific population (e.g., Saudi central dialect) could prove to be of a superior validity.

Throughout the process of translating and adapting the VAST assessment, we have arrived to the conclusion that not all tasks/items are suitable for direct translation and adaptation. For instance, the sentence comprehension task in the VAST presents only four types of sentences (active, passive, subject cleft, and object cleft); therefore, using the English version as a guideline for creating a translated and adapted Arabic version will eliminate various sentence types that are common in Arabic and occurs in everyday conversations. It includes sentences that start with a verb (i.e., VSO فتح الولد الباب), sentences that begins with verbs and the subject is implied within the verb instead of explicitly mentioned (e.g. كَتَبْتُ الرسالة/ I wrote the letter). Moreover, the inflection in Arabic sentences adds another layer of complexity, even in simple active sentences, which should be tested in aphasia language assessments. Accordingly, a more thorough process should be taken in developing some of the tasks within the assessment batteries to ensure that it covers all aspects that are fundamental to Arabic (e.g., subject verb agreement) which may not exist in the original language of the test. This will ensure that the test provides a sensitive measure of the functional language that occurs in everyday communication.

Lastly, creating an Arabic version of the grammaticality judgment task proved to be the most challenging. Although the MAPI approach has been implemented and extensive modification has been made to adapt the sentences, the final product remained unsatisfactory. One clear distinction between English and Arabic was evident in passive sentences that are semantically non-reversible. In Arabic, the subject would be deleted, which is not the case in English. For example, in English, "The boy is eaten by the dessert" (i.e., an example of a grammatically incorrect passive sentence) would be translated to " The boy is eaten" because adding the "desert" is not grammatically correct in Arabic. Similarly, "The sandwich is made by the boy" (i.e., an example of a grammatically correct passive sentence) would be translated to "the sandwich has been made". Although only items in which controls achieved above 90% accuracy of performance were included in the final version, the overall product can be further refined and improved. Moreover, an assessment developed based on normative data and Arabic linguistic expertise would be expected to be of superior validity.

We aimed to explore the stability of the test-retest scores of the Arabic version of the VAST as a preliminary investigation of its reliability. The interpretation of Spearman's correlation coefficient value indicated a strong correlation between scores on baseline 1 and baseline 2 in all subtests except the semantic association subtest. It could be attributed to the subtest content difference (i.e., semantics vs meta-linguistic processing in complex tasks). Moreover, higher scores on the second baseline could reflect better execution of the task rather than the effect of learning the test items. Also, given that the sample size was limited to 4 participants, the results remain inconclusive. A replication of the pilot testing on a larger sample and combining it with other measures could be more informative in determining the reliability of the Arabic version of the VAST. For example, an inter-rater/inter-observer reliability test (i.e., assesses the degree of agreement between the estimates of two different raters/observers). Also, parallel-forms reliability test (i.e., assess the consistency of performance across two equivalent language tools administered simultaneously), or internal consistency reliability test (i.e., assesses item-specific consistency of scores within a test).

Although our recommendation was to replace some of the VAST subtests with novel assessments based on normative data and psycholinguistic research on Arabic, we still support the translation and adaptation of some established and widely used assessment tools. For example, having a translated and adapted Arabic version of the core outcome set (COS) recommended by Research Outcome Measurement in Aphasia (ROMA) group (Wallace et al. 2019) for use in aphasia treatment studies would be a valuable contribution to the literature and will assist in producing high-quality studies. Also, the use of Arabic versions of widely used assessment tools such as WAB, VAST, BDAE, etc., could be beneficial in many ways. For example, standardising these tests and implementing them in clinical settings and research would produce quantitative data that can be understood by non-Arabic speaking researchers who are familiar with the original tests. Also, the data can be compared to its equivalent in different languages. Ultimately, this may encourage international researchers to conduct studies on Arabic speaking individuals with aphasia, a contribution that would hopefully address the paucity of studies we currently have in the literature due to the limited number of Arabic speaking researchers in the field of aphasiology. Finally, cross linguistic comparisons will be achievable when we have standardised Arabic versions of some of the language assessment tools that are popular in other languages.

The translation and adaptation of the WAB provided us with an Arabic assessment tool that tested the participants' overall language performance. On the other hand, the Arabic VAST provided us with a more in-depth assessment of verb and sentence comprehension and production. As mentioned above, developing an Arabic version of the sentence comprehension and grammaticality judgment subtests equivalent to the English version proved challenging. However, the translation and adaptation of subtests that prompted verb and sentence production were straight forward. The most noticeable changes to the Arabic version was the deletion of pictures that were not relatable to the Saudi culture. In general, the selected pictures were clear and elicited sentences that matched the events illustrated in the picture. However, it is noteworthy to mention that the original version's items determined the targeted verbs and their related nouns. Therefore, the verb and sentence production subtest in the Arabic version of the VAST was not designed to target a specific category of verbs based on frequency, age of acquisition, etc. Developing an Arabic version that controls those variations would be of great value. It also applies to the therapy stimuli used in the study. Designing a list of target verbs and nouns based on their features in Arabic (e.g., concreteness vs abstractness, high frequency vs low frequency) and choosing the picture prompts would produce a more refined and psycholinguistically controlled approach to therapy.

Finally, the difficulty we encountered in adapting the COAST, as the only PROM (patient-reported outcome measure), provided some insight into cultural challenges instead of purely linguistic adaptation. Where one is attempting to probe complex concepts such as perceptions of communication, societal issues like typical patterns of communication (e.g. whether talking to strangers is considered routine or not), issues like family structure, and practical issues such as typical self-care work carried out by elderly people, imply that the varying concepts of normality for stroke survivor means simply borrowing and adjusting from one language to another appears questionable as an approach. Instead, it seems likely that bottom-up development work may be required within given cultures, whereby the important issues from a patient perspective on, for example, what a good recovery with stroke and aphasia is, are determined locally and testing stems from this ecological basis, rather than from a relatively distant cultural space.

A study by Khwaileh and Grosvald (2019) is one of the few studies that discussed the linguistic and cultural challenges of translating and adapting language and quality of life assessment tools from English into Arabic. In addition to arriving at the same conclusion that we have presented, they provided a more in-depth discussion of the fundamental differences in linguistic properties between English and Arabic, specifically at the syntactic and morpho-syntactic levels. The article also highlighted the authors' impression of how the approach of "importing" language tests and quality of life post CVA assessments via translation into the local language "fails to address a host of linguistic and cultural issues specific to the Arabic-speaking world". Therefore, the authors called for the development of language assessments that are "based on data that originate from native speakers of the dialect of Arabic specific to the region in question" (Khwaileh and Grosvald, 2019, p.44).

The Bilingual Aphasia Test BAT (Paradis, 1987) is one of the first aphasia language tests that produced an Arabic version (Paradis et al., n.d.). The author emphasized the importance of creating equivalence in structural complexity. That "a person who adapts the BAT to another language must apply the very same rationale that led to the selection of particular stimuli in the original design" (Paradis, 1987, p.430). He also cautioned that "the cross-linguistic equivalence of the various versions of the test is threatened not only by the cultural differences between the communities in which the languages are spoken, but, especially in those tasks dealing with syntax and morphology, by the structural diversity of languages." (Paradis, 1987, p.430). In conclusion, to overcome the current challenges to producing Saudi-Arabic versions of any language test equivalent to the original tests and measures the same skills, normative data on language acquisition in Saudi-Arabic is needed. Moreover, a database of the type of errors and their frequency in Arabic-speaking individuals with aphasia would be valuable in adapting and translating theoretically informative and clinically useful assessment and therapy tools for aphasia practice.

CHAPTER 6 Piloting and evaluation of an adapted and translated multilevel sentence therapy applied to Arabic stroke survivors with aphasia in Saudi Arabia

6.1 Introduction

The kingdom of Saudi Arabia is the largest country in the Middle East, covering around 4/5 of the Arabian Peninsula (2,150,000 km^2). It is divided into 13 administrative regions and has a population of 34,218,169 people (the Saudi general authority for statistics, 2019). A study by Alqahtani and colleagues (2020) attempted to estimate the annual incidence of stroke for people residing in Saudi Arabia through a systematic review. The pooled annual incidence of stroke for people living in Saudi Arabia, from the five studies that met the inclusion criteria, estimated an average of 29 stroke cases for every 100,000 people annually. In Saudi Arabia, healthcare services are primarily government funded; however, it is currently not clear what the financial burden related to treatment and post-stroke care in Saudi Arabia is.

The primary provider of health care services in Saudi Arabia is the Ministry of Health (MOH). Its health care delivery model can be categorised into five tiers: primary health care centres, district hospitals, general hospitals, central hospitals and medical cities. A survey study by Khoja and Sheeshah (2018) provided an overview of the availability of speech-language pathology services in Saudi Arabia, which was obtained from 196 (from a total of 206) government hospitals managed by the MOH. The findings revealed that only 29 hospitals, distributed across 7 of the 13 regions, provided speech-language pathology services. However, almost 70% of the total of 183 Speech-Language Therapists/Pathologists (SLT/P s) employed by these hospitals were in the Riyadh region. In addition to the apparent unequal distribution, the overall number of professionals in the field in Saudi Arabia is low, with a ratio of 0.67 SLT/P per 100,000 people. Nevertheless, the barriers to SLT/P service delivery in Saudi Arabia are not limited to the reduced number of professionals or the uneven distribution of facilities across the regions. It extends to include the lack of assessment tools and literature specific to Arabic speaking individuals with speech and language disorders (Shaalan, 2009).

One possible solution to reducing these barriers to SLT/P service delivery in Saudi Arabia is establishing a computer-based therapy for individuals with aphasia. It is expected to facilitate the accessibility of language therapy services to a large number of PWA who do not live within reasonable proximity to hospitals that offer those services. After the initial in-person speech and language assessment session and training on using the computer program, the in-person follow up sessions can be less frequent. This technological innovation could allow the SLT/P to develop a therapy plan according to the language assessment findings and program the software with therapy tasks to target the identified therapy goals on the participant's personal or loaned device. After that, the monitoring of therapy progress can be done remotely, and in-person therapy sessions are kept to a minimum.

Additionally, a computer-based therapy could be a reliable mode of providing PWA with the much-needed increased intensity of therapy in an accessible and cost-effective way. Increasing the dose of therapy will not require additional SLT/P -directed sessions, which are currently limited in light of the shortage of SLT/P s. Also, this approach could be advanced to deliver teletherapy services.

To bridge the barriers identified in Saudi Arabia and provide a sustainable, cost-effective, and practical way of delivering SLT/P services to Arabic speaking individuals with aphasia, we aimed to produce a compatible computer-based therapy. The adopted approach was to create an Arabic equivalent of the original computer therapy program introduced in Chapter 4 through translation and cultural adaptation. We predicted that a translated and adapted version of the therapy program, which implements the same theoretical model-driven principles such as the mapping therapy and temporal window widening (SentenceShaper), will yield similarly positive outcomes as observed in Chapter 4 with the English-speaking group of PWA.

Several studies have investigated language processing in Arabic speakers with aphasia (Albustanji, 2009; Khwaileh et al., 2015; Khwaileh et al., 2017; Khwaileh et al., 2020; Mimouni and Jarema, 1997; Mimouni et al., 1998) that included dialects such as Jordanian Arabic, Algerian Arabic and Gulf-Arabic. The sample size of people with aphasia in these studies has ranged between 2-15 participants. Although the Gulf-Arabic dialect is understood and spoken by Saudi people in the eastern region of Saudi Arabia, it contains features that are distinct from other major Saudi subdialects such as the central (Najdi), western (Hijazi), southern (Jinoubi), and Bedouin. It includes morphological, phonological and lexical variations. Al-Twairesh and colleagues (2018) presented a Table (see Table 6.1 below) that highlights some of the most prominent phonological variations within Saudi subdialects (for further information on morphological and lexical variations across the Saudi dialects, refer to (Al-Twairesh et al., 2018).

Table 6.1 Phonological variations of Saudi dialect (Al-Twairesh et al., 2018, p. 75)

Phoneme feature	In Saudi dialect	Example
/q/ phoneme (MSA ق)	/g/ in almost all Saudi subdialects	قلب /qalb/ becomes /galb/
/D/ phoneme (MSA ض)	/Dˇ/	ضرس /Dirs/ becomes /Dˇirs/
MSA glottal stop phoneme	Disappeared in most cases (i.e. if not preceded or followed by a vowel)	ذئب /ðiʼb/ becomes /ðīb/
/k/ phoneme (MSA ك)	In some cases, is transformed to: • /s/ or /tš/ in Najd • /ج/ in Gulf Arabic • /ش/ in Southern dialect	لك/lak/ becomes: • /lis/ or /litš • /lij/ • /lish/
/θ/ phoneme (MSA ث)	/s/ or /t/ in Hijazi	ثاني/θani/ becomes /sani/ or /tani/
/ð/ phoneme (MSA ذ)	/d/ in Hijazi	كذب/kaðib/ becomes /Kdb/
Short vowels	SD omits many short vowels that appear in the MSA	بيوت /biyot/ becomes /byot/

Accordingly, the literature on Gulf-Arabic is not sufficient to fully understand language processing in other Saudi subdialects. Studies investigating error types in PWA speaking different major Saudi dialects are still needed. To the best of our knowledge, the current study will be the first attempt in describing the features of language disorders observed in people with aphasia speaking the Saudi central dialect.

Moreover, intervention studies that have explored the outcomes of aphasia speech and language therapy delivered in Arabic are presently limited to one study. This single case study reported Melodic Intonation Therapy MIT's application to the treatment of a native speaker of Jordanian-Arabic with aphasia (Al-Shdifat et al., 2018). In the context of this limited literature, the current study will be the first to report the outcomes of a theoretical model-driven language intervention delivered in Arabic that targets language production deficits in native Arabic speaking individuals with aphasia and assess its feasibility.

The final phase of empirical research presented in this thesis will describe a study in which the adapted and translated assessment outcome measures and therapy materials, presented in

Chapter 5, were piloted with a small number of Arabic participants with stroke aphasia, using single case study design. The specific study aims were to:

1. Describe the language profiles found in Saudi-Arabic speaking individuals with aphasia (central dialect) and the applicability of implementing the classical aphasia subtypes proposed by Kertesz (1982) to classify them.
2. Assess the feasibility of a novel computer-based sentence therapy method in remediating sentence production deficits in Arabic speakers with aphasia.

6.2 Methods

6.2.1 Study design

A single-case experimental design with multiple assessment points has been selected to evaluate changes in language production skills from baseline and during and after the intervention.

Three major hospitals in Riyadh, Saudi Arabia, were selected as study sites for participant recruitment and language testing. The ethics approval for conducting research in those sites was obtained from the institutional review board IRB in each hospital: King Fahad Medical City KFMC, King Saud University Medical City KSUMC, and Prince Sultan Bin Abdulaziz Humanitarian City PSBAHC (See appendix 5). The recruitment methods included internal referrals from the hospital departments such as speech-language therapy, rehabilitation, cardiology and neurology departments. Additionally, self-referred participants from aphasia support groups were also accepted. Then, candidates who met the inclusion/exclusion criteria for therapy were invited to attend weekly therapy sessions at the out-patient clinic in King Saud University Medical City (KSUMC).

Also, with the participants' signed consent, performance on language tests and therapy tasks was recorded on two audio recorders (to ensure backup). The recordings were saved in a password-encrypted hard drive for later analysis and score checking.

6.2.2 Participants

We applied the same inclusion criteria as in chapter 4 with one addition that specified the dialect. The participants were required to be native Arabic speakers of a Saudi dialect. Thus, the inclusion criteria encompassed presenting a history of aphasia post left hemisphere stroke at the chronic stage (6-8 months post-stroke), a noticeable language production deficit, and adequate verbal comprehension skills determined by scoring above 50% on the sentence comprehension task. Also, the absence of any neurological disorders such as brain tumour, uncontrollable seizures, traumatic brain injury, etc. was required. The presence of dysarthria or apraxia was not an indication for exclusion in this study; however, it was required that the participant produce words with adequate speech intelligibility (i.e., speech intelligibility subjectively judged by a qualified speech therapist). Moreover, adequate visual and hearing acuity was required, with or without correction. All ages over 18 were included. Additionally, participants were required to have basic literacy and computer skills and fair hand mobility/dexterity to operate and engage with the iPad's therapy program independently. The exclusion criteria included premorbid cognitive deficit determined by history taking and interview with the participant and caregivers. Finally, only participants who were able to commit to the eight weekly consecutive sessions were recruited.

Four native Saudi-speaking individuals with aphasia post-stroke were recruited to the study. They participated in 8 consecutive clinician-delivered computer-based therapy sessions and three clinician-administered language testing sessions over three months. They also completed 4 hours of weekly independent home practice over a period of 8 weeks (see Table 6.3).

Table 6.2 Basic demographic details of participants*

Participant No. and Initials		Age (years)	Gender	Education (years)	Time-post onset (months)	clinical aphasia classification**
1	NS	60	M	16	18	non-fluent
2	SA	54	F	12	36	fluent
3	SM	33	F	16	36	fluent
4	ND	21	F	6	18	non-fluent

*No medical records were presented at the time of recruitment; therefore, dates and durations were estimated based on participant and caregivers' verbal report **Based on the clinical impression of at least two qualified native Arabic speaker SLT/P s

6.2.3 Assessment tools

The VAST test used in the current study differs from the original paper-based version mentioned in Chapter 4, as the Arabic version of the VAST test was translated from a revised version of the VAST, which is computer-based. The two versions differ by the type of subtests they contain and the items' characteristics (i.e., number, type, and picture stimuli). In addition to the subtypes of the original VAST, verb comprehension, sentence comprehension, grammaticality judgement task, action naming, fill-in verbs in sentences (i.e., finite verbs and infinitives), and sentence production, two new subtests were added: object naming and semantic association. The original scoring guidelines were used in the current study, except the sentence construction task. It implemented a new scoring rubric (see Appendix 1.B). For the sentence construction subtest, the Arabic version of the scoring rubric's maximum score was determined by the average performance of 5 neuro-typical native Saudi speaking individuals. It involved quantifying the minimum number of sentence constituents required to describe a picture scene in complete, grammatically correct, relevant and informative sentences (See Chapter 5).

The WAB-R had been selected as a secondary outcome tool to measure the overall language performance on various language tasks before and after therapy. No changes were made to the original translated and adapted Arabic version produced in 2017 by a team of SLT/P s at King Saud University Medical City KSUMC. The scoring approach followed the original guidelines (Kertesz, 2007) with minor modifications to include only spontaneous answers and eliminate cueing in assessment (see Chapter 5).

The narrative samples were collected in a picture description task using two types of stimuli. First, the cookie theft scene obtained from the BDAE test (Goodglass and Kaplan, 1983) was presented in one 6x8 inch picture. The dinner party story (Mark et al., 1983) was introduced in 8 picture sequence fitted in one A4 page. The method used to analyze speech samples in chapter 4 had to be replaced, as the Systematic Analysis of Language Transcripts SALT software is not compatible with Arabic. Also, to the best of our knowledge, no equivalent software or scoring rubric exists to analyse Arabic speech samples. As a result, we have chosen to manually transcribe speech samples for sentences completeness, grammaticality, relevance and informativeness (See Appendix 1.B for scoring rubric).

Table 6.3 Testing frequency

	Test	Baseline 1	Baseline 2	Interim 1	Post-therapy
1	VAST (Sentence construction task only)	√	√	√	√
2	VAST (all subtests)	√	√	-	√
3	Narrative samples	√	√	√	√
4	WAB	√	-	-	√

(√) administered at this time point (-) not administered at this time point

Baseline 1: completed during the recruitment phase and at least one week apart from baseline 2. Baseline 2: testing was conducted at least one week before the start of the therapy program. Interim 1: the beginning of the fifth therapy session and before the start of the therapy tasks. Post-therapy: In the thirteenth session, no therapy tasks were administered.

6.2.4 Therapy protocol

The previous study's findings with the English-speaking group of participants with aphasia in Chapter 4 showed that most participants presented a peak in their sentence construction skills at the second-interim assessment time-point. Since the current study implements the same therapy protocol, we anticipated that the Arabic-speaking participants would show a comparable response at the same time-point. Accordingly, it was determined that the first two levels were sufficient to judge the participants' response to therapy. The modification of the original therapy protocol, used in Chapter 4, entailed omitting the third level of therapy and keeping the first two levels, including the interim assessments following each level (see Table 6.3 and Table 6.4). Thus, testing was conducted at four time-points only (i.e., baseline 1, baseline 2, interim 1, post-therapy). Also, due to time limitations, the maintenance assessment at four weeks post-therapy was not administered. For further details about Levels 1 and 2 of the therapy program, please refer to the methods' section 'Therapy Protocol and Procedure' in Chapter 4, the methods' section 'Adapting and translating the therapy protocol and materials' in Chapter 5, and Appendix 3.

Table 6.4 Levels of therapy delivered over the time course of the program

ACTIVITY	START	DURATION (weeks)	WEEKS							
			1	2	3	4	5	6	7	8
Level 1 - Phase A	1	2								
Level 1 - Phase B	3	2								
Level 2 - Phase A	5	2								
Level 2 - Phase B	7	2								

The recruited participants were asked to commit to 12 visits to the clinic at the research site, including 8 consecutive weekly therapy sessions and 3-4 sessions of assessment. The duration of each therapy session ranged between 75-90 minutes, and the assessment sessions were limited to 2 hours (with breaks as required) to reduce the risk of fatigue. Additionally, in the treatment phase, the participants were required to practice language therapy tasks for 4 hours/week independently at home. A detailed explanation of the homework tasks and a written reminder were provided. It mainly focused on carrying over the practice of tasks presented in the sessions.

The testing and therapy sessions were delivered by the principal investigator, a qualified speech-language pathologist. The testing materials included hard and soft copies of words, pictures, and symbols (See Chapter 5). In addition to the online scoring during the session, each session was audio-recorded for further analysis and speech sample transcription.

6.3 Results

Study aim 1: Describe the language profiles found in Saudi-Arabic speaking individuals with aphasia (central dialect) and the applicability of implementing the classical aphasia subtypes proposed by Kertesz (1982) to classify them.

6.3.1 Performance at baseline (average performance on two baselines at least one week apart)

The expected normal performance on all subtests of the Arabic version of the VAST and WAB would be at ceiling, as only items scored 100% accuracy by a sample of 6 neuro-typical native speakers were included (see Chapter 5 for further information).

6.3.1.1 Verb and Sentence Test VAST

The VAST receptive language tasks consisted of three comprehension subtests and a semantic association task. For the verb comprehension, sentence comprehension subtests, and semantic association task, 3 out of 4 participants performed towards the ceiling at baseline (92.7, 87.8, and 92.1 % accuracy respectively), except for participant NS (82.3,73.3, and 81.6 % accuracy, respectively). The VAST production section is composed of 4 subtests: action naming, object naming, fill-in verbs in sentences, and sentence production. The participants' average scores at baseline (63.4, 78.1, 73, 68.4% accuracy) are noticeably more inferior than their performance on the comprehension subtests. As can be seen from the Table 6.5, participant SA showed a high level of performance (towards the ceiling) on 7 out of 8 VAST subtests (99, 91.1, 91.7, 92.2, 94.7, 89.8 and 94.7% accuracy), except for action naming with a score of 76.8% accuracy. Similarly, participant SM showed a high level of performance on most VAST subtests (95.8, 98.9, 93.9, 96.7, 94.7,92.4, and 97.4% accuracy) with a slight relative decrease in performance on the grammaticality judgment task (83.3% accuracy). On the other hand, participants NS and ND presented discrepancy in performance between tasks. For example, participant NS's performance on the verb comprehension (82.3% accuracy), sentence comprehension (73.3% accuracy), fill-in verbs in sentences (81.6% accuracy) and semantic association (81.6% accuracy) was superior to performance on the grammaticality judgment task (66.7% accuracy), action naming (58.5% accuracy), object naming (66.7% accuracy) and sentence production (51.3% accuracy). Although participant NS did not show a clear distinction in performance on comprehension versus production tasks, this pattern was clearly detected in participant ND's performance. Her scores on verb comprehension (93.8% accuracy), sentence comprehension (87.8% accuracy), and semantic association subtests (94.7% accuracy) were noticeably superior to her scores on action naming (24.4% accuracy), object naming (56.7% accuracy) and fill-in verbs in sentences subtest (21.1% accuracy) and sentence production (39.8% accuracy). Although her performance on the grammaticality judgment task was relatively low (55% accuracy), this subtest's validity remains questionable, as mentioned in Chapter 5.

Table 6.5 Participants' performance on VAST subtests before and after therapy (percentage)

Participant No. and initials		Verb comprehension		Sentence comprehension		Grammaticality judgment		Action naming		Object naming		Fill-in verbs in sentences		Sentence production		Semantic association	
		Pre-*	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
1	NS	82.3	89.6	73.3	91.1	66.7	80	58.5	73.2	66.7	86.7	81.6	89.5	51.3	91.1	81.6	84.2
2	SA	99	100	91.1	100	91.7	96.7	76.8	87.8	92.2	97.8	94.7	94.7	89.8	97.4	94.7	94.7
3	SM	95.8	100	98.9	100	83.3	80	93.9	100	96.7	100	94.7	100	92.4	95.3	97.4	100
4	ND	93.8	95.8	87.8	95.6	55	53.3	24.4	80.5	56.7	66.7	21.1	63.2	39.8	85.4	94.7	100
	Mean	92.7	96.4	87.8	96.7	74.2	77.5	63.4	85.4	78.1	87.8	73	86.9	68.4	92.3	92.1	94.7
	SD	7.3	4.9	10.7	4.3	16.5	18	29.7	11.4	19.4	15.2	35.2	16.3	26.7	5.3	7.1	7.4
% of change		+4%		+10%		+4%		+35%		+12%		+19%		+35%		+3%	

*All pre-therapy scores represent the average baseline score (baseline 1 + baseline 2)

6.3.1.2 WAB

The participants showed a high level of performance on Yes/No questions, auditory word recognition, and responsive speech (noun retrieval in a sentence context) with an average of 90, 91.3, and 100% accuracy, respectively. Their performance on sequential commands, object-naming and sentence completion was also strong though less towards the ceiling with average scores of 72.7, 75, and 75% accuracy, respectively. Nevertheless, they exhibited relatively low performance on the repetition and word fluency subtests with an average of 58.3 and 40% accuracy, respectively.

Looking at individual participants' profiles, NS showed a high level of performance on the auditory word recognition (91.7 % accuracy), sequential commands (81.8 % accuracy), sentence completion (80% accuracy) and responsive speech (100% accuracy). There was more evidence of impairment in performance on Yes/No questions (75% accuracy) and object naming (75% accuracy); relatively poor performance on repetition (40% accuracy), and word fluency (5% accuracy). Participant ND also showed relatively poor performance on 3 out of 8 subtests: repetition, object naming, and sentence completion with an accuracy level of 40 % in each. In comparison, her performance on sequential commands and word fluency was stronger (63.6 and 65% accuracy, respectively). Nevertheless, her performance on the Yes/No questions, auditory word recognition and responsive speech subtests was superior to her performance on the rest of the tasks with accuracy levels of 90, 80, and 100%, respectively. Participant SA showed better overall performance than NS and ND with high-level scores that range between 90 and 100% accuracy on 5 out of 8 subtests: Yes/No questions, auditory word recognition, object naming, sentence completion, and responsive speech. Her score on the repetition task was slightly poorer (73.3% accuracy) and substantially poorer on the sequential commands (54.5% accuracy) and word fluency (45% accuracy) subtests. Lastly, participant SM showed the highest overall performance with a range of 80-100% accuracy on 7 out of 8 subtests, with relatively poor performance on the word fluency subtest (45% accuracy).

Table 6.6 Participants' performance on WAB subtests before and after therapy (percentage)

Participant No. and initials		Yes/No Qs		auditory word recognition		sequential commands		repetition		object naming		word fluency		sentence completion		responsive speech	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
1	NS	75	100	91.7	93.3	81.8	63.6	40.0	73.3	75	80	5	25	80	100	100	100
2	SA	95	95	93.3	93.3	54.5	54.5	73.3	73.3	90	95	45	60	100	100	100	100
3	SM	100	95	100	96.7	90.9	90.9	80.0	86.7	95	100	45	65	80	100	100	100
4	ND	90	70	80.0	88.3	63.6	63.6	40.0	73.3	40	70	65	40	40	80	100	80
Mean		90.0	90.0	91.3	92.9	72.7	68.2	58.3	76.7	75.0	86.3	40.0	47.5	75.0	95.0	100.0	95.0
SD		10.8	13.5	8.3	3.5	16.6	15.8	21.3	6.7	24.8	13.8	25.2	18.5	25.2	10.0	0.0	10.0
% of change		0%		+1.8%		-6.2%		+31.6%		+15.1%		+18.8%		+26.7%		-5%	

6.3.1.3 Discourse

To examine the participants' performance on discourse measures, we calculated the word count of the speech sample they produced in describing each picture stimuli, the Cookie Theft and Dinner Party (see Appendix 1.C Arabic discourse scoring guideline). We then examined the speech samples' content on the following measures: phrases, complete simple sentences, compound sentences, complex sentences, total number of sentences, verbs, nouns, adjectives and prepositions.

The total number of words in each participant's speech sample was compared to the average score achieved by healthy controls (see Chapter 5, Table 5.8) and the degree of variance was illustrated in percentage with positive scores (+%) refer to a value above average, and negative scores (-%) refer to a value below average (larger values indicate a wider difference from the average score). The group of participants with aphasia (n=4) scored significantly below the average performance of healthy controls on both the cookie theft (CT) and dinner party (DP) speech samples The poorest performance was noted in participant NS's performance (CT: -92.6 % ; DP: -87.3% , followed by ND (CT: -86.2 % ; DP: -80.5%) , SA (CT: -70.5 % ; DP: -57.2%) , and SM (-67.7 % ; DP: -45.5%), respectively.

Within the group of participants with aphasia, participant NS produced the least number of words in both speech samples CT and DP. A closer examination revealed that his speech samples were limited to an average of one or two complete sentences, with one incident of compound sentence production and no evidence of complex sentences. Also, he produced an average of 1.5 and 3 verbs, 2 and 7 nouns, 0 and 0.5 adjectives, and 0.5 and 1.5 prepositions on CT and DP discourse samples respectively. It was noted that successful elicitations were more frequent in response to the picture sequence stimuli (i.e., dinner party). Participant NS's performance on the discourse production task was in alignment with his overall poor performance on the VAST and WAB language production subtests. Those findings along with his high scores on the comprehension subtests and poor scores on the speech repetition task (i.e., WAB subtest) further supported the categorisation of his symptoms as Broca's aphasia (i.e., non-fluent language production, adequate comprehension, and poor repetition).

Similarly, participant ND exhibited noticeable difficulties in discourse production. In the CT speech sample, she failed to produce any complete sentences on either of the two baseline assessments (at least one week apart). However, she succeeded once on the DP speech sample in producing one complete sentence (simple active sentence only). The content of the speech samples, CT and DP, averaged 2 and 3 verbs, 5 and 11 nouns, 0 and 1 adjectives, and no prepositions, with more successful elicitations noted on the picture sequence stimuli DP. The characteristics of her speech samples along with her performance on the WAB and VAST subtests provided evidence for her classification with Broca's aphasia.

Although participant SA produced an average of 16 words on her CT speech sample and 44 words on her DP speech sample, the total sentences in each discourse were limited to 3 and 5.5 complete sentences, only one of them is classified as a compound sentence and the rest were simple active sentences. The speech sample included 2.5 and 9.5 verbs, 6.5 and 18, 1 and 0 adjectives, and 1 and 4 prepositions, on CT and DP speech samples respectively.

Likewise, participant SM produced an average of 17.5 and 56 words on each speech sample, CT and DP discourse respectively. However, the complete sentences were limited to 4.5 and 11.5 sentences, in which only 2 and 3 of them were classified as compound sentences and no evidence of complex sentences. Nevertheless, participant SM's speech sample presented a higher rate of content words compared to the other 3 participants' production. It included 5 and 14.5 verbs, 8 and 25 nouns, 0 and 0.5 adjectives, and 1.5 and 2.5 prepositions on the CT and DP speech samples respectively.

Although both participants SA and SM showed strong performance at baseline on all VAST and WAB language production tasks (i.e., scores ranged between 76.8-100% accuracy), except for the word fluency task (i.e., 45% accuracy), their language deficiency was more evident on the discourse measures. They performed significantly below the average score of healthy controls (i.e., range of deficiency 45.5% -70.5% below average). Those findings further supported the classification of their language deficit as Anomic aphasia (i.e., fluent, adequate comprehension and good repetition skills).

Table 6.7 Cookie theft and dinner party discourse's word count per participant at baseline

	Cookie theft discourse		Dinner party discourse	
Participants	Average performance at baseline* Mean (SD)	score compared to mean score of healthy controls	Average performance at baseline* Mean (SD)	score compared to mean score of healthy controls
Healthy controls (n=6)	54.2 (27.1)		102.7 (45.2)	
NS	4	- 92.6 %	13	- 87.3 %
SA	16	- 70.5 %	44	- 57.2 %
SM	17.5	- 67.7 %	56	- 45.5 %
ND	7.5	- 86.2 %	20	- 80.5 %

*word count (-) below average (+) above average

6.3.2 Examining patterns in performance in reference to classic aphasia classification suggested by Kertesz (2007)

To determine the participants' performance profile's relevance to the classical classification of aphasia subtypes, we examined their overall performance on multiple linguistic levels. The analysis covered two language assessment batteries' combined subtests, the VAST and the WAB. The subtests from both tests complement each other. For example, the WAB does not contain an action naming task or a sentence production task (in response to a picture description task) while the VAST does. On the other hand, some subtests such as the object naming task exist in both tests with some variance in the number of items (20 vs 45 items), stimulus type (objects vs pictures), and category. In this case, the stability of performance across tests will further support the reliability of the results. It is evident in the current sample, illustrated in Table 6.8, in which ND performed poorly on the object naming task on both the VAST and the WAB, while the rest of the participants showed similar strong performance on the same subtest on both VAST and WAB.

Through clinical impression, derived from the interview and unstructured conversations, NS and ND exhibited non-fluent aphasia features, while SA and SM presented fluent aphasia characteristics. Interestingly, 3 (ND, SM, and SA) out of 4 participants showed performance on VAST and WAB subtests that supported this bedside clinical impression, with only NS presenting a less defined pattern.

Participant ND's performance closely resembled the characteristics of Broca's aphasia that include non-fluent speech production, adequate comprehension, and poor repetition. It was reflected by her good performance on the verb and sentence comprehension scores: VAST verb comprehension (93.8%accuracy), VAST sentence comprehension (87.8% accuracy), WAB Yes/No questions (90% accuracy), and WAB auditory word recognition (80% accuracy); her consistently and markedly poorer performance on production tasks: VAST action naming (24.4% accuracy), VAST object naming (56.7% accuracy), VAST fill-in verbs in a sentence (21.1% accuracy), VAST sentence production (39.8% accuracy), WAB object naming (40% accuracy), WAB sentence completion (40% accuracy), and WAB word fluency (65% accuracy); and her poor performance on the WAB repetition task (40% accuracy).

Participant SA's performance represented the characteristics of Anomic aphasia which includes fluent speech, good comprehension, and the ability to repeat words and phrases with high accuracy. It was evident in her high accuracy level on most VAST and WAB subtests, except on two subtests. As shown in Table 6.8, her scores ranged from 73.3-100% accuracy level, with the majority being above 90%. Nevertheless, her performance on the sequential commands and word-fluency was poor, with 54.5 and 45 % accuracy, respectively. Her poor performance on the sequential commands could be attributed to limitations in short term verbal memory. On the other hand, her poor performance on the word fluency task can be explained by her clear clinical symptoms of anomia and apraxia.

Similarly, participant SM's skills represented the characteristics of Anomic aphasia. She showed strong performance on all subtests within the VAST and WAB assessment battery with a range of 80-100% accuracy except for one task. She scored 45% accuracy on the WAB word fluency subtest. Her poor performance could be attributed to signs of apraxia which were noted throughout the assessment and therapy sessions combined with her underlying anomic symptoms. Overall, SM's performance was also consistent with the initial clinical impression that classified her symptoms as anomia aphasia subtype.

Lastly, participant NS exhibited characteristics of Broca's aphasia. He showed low performance on VAST action naming (58.5% accuracy), VAST sentence production (51.3% accuracy), and WAB word fluency (5% accuracy). However, better performance was noted on: VAST object naming (66.7% accuracy), WAB object naming (75% accuracy), VAST fill-in verbs in sentences (81.6% accuracy), WAB sentence completion (80% accuracy), and responsive speech (100%accuracy). The above results may indicate that NS's poor performance is related to verb production in a confrontational task, as well as in a sentence context in a picture description task. In comparison, his language production deficit is less evident with nouns in confrontational naming, sentence completion, and responsive speech tasks. On the other hand, his strong performance on the fill-in verbs in sentences task could be explained by the semantic cues' availability in the sentence context that supported verb retrieval. Finally, his poor performance on the word fluency task can be explained by his speech characteristics of dysarthria and apraxia, in addition to his word retrieval deficit. His overall performance resembled a symptom classification of Broca's aphasia subtype with non-fluent speech features, poor repetition, adequate language comprehensions, and poor language production. However, the language production deficit is more evident with verbs.

Table 6.8 Participants scores at baseline (percentage)

	Participant initials	NS	SA	SM	ND
VAST	Verb comprehension	82.3	99	95.8	93.8
	Sentence comprehension	73.3	91.1	98.9	87.8
	Grammaticality judgment	66.7	91.7	83.3	55
	Action naming	58.5	76.8	93.9	24.4
	Object naming	66.7	92.2	96.7	56.7
	Fill-in verbs in sentences	81.6	94.7	94.7	21.1
	Sentence production	51.3	89.8	92.4	39.8
	Semantic association	81.6	94.7	97.4	94.7
WAB	Yes/No Qs	75	95	100	90
	Auditory word recognition	91.7	93.3	100	80
	Sequential commands	81.8	54.5	90.9	63.6
	Repetition	40	73.3	80	40
	Object naming	75	90	95	40
	Word fluency	5	45	45	65
	Sentence completion	80	100	80	40
	Responsive speech	100	100	100	100
Discourse*	Cookie theft discourse	-92.60%	-70.50%	-67.70%	-86.20%
	Dinner party discourse	-87.30%	-57.20%	-45.50%	-80.50%

* Percentage % scores refer to variance compared to the average score of healthy controls (see Table 6.7 above and Table 5.8 in Chapter 5 for more details on controls), Underlined subtests: language production measures.

Study aim 2: Assess the feasibility and of a novel computer-based sentence therapy method in remediating sentence production deficits in Arabic speakers with aphasia.

6.3.3 Treatment Results

6.3.3.1 Sentence production

To examine the statistical significance of gains in scores within individual participants, Wilcoxon matched pair tests of ordinal data were conducted. Given that in a therapy study, we predict a direction of change towards post-therapy improvement, 1-tail results have been presented. Wilcoxon matched-pairs test (Table 6.9 below) revealed that changes in scores were statistically significant for only 3 participants: NS (z= 4.19, p= 0.0), SA (z= 2.4, p= 0.016), and ND (z= 4.19, p= 0.0), while participant SM's results were not significant (z= 1.33, p= 0.18).

Table 6.9 Individual participant statistics for significance ($p<0.05$) of therapy gains on the VAST sentence construction subtest using Wilcoxon matched pairs test of ordinal data.

Participants' No. and initials		Pre-therapy mean (SD)	Post-therapy mean (SD)	Wilcoxon Z	Asymp. Sig. (1-tailed)
1	NS	4.1 (2.4)	7.3 (1.6)	4.19	<0.001
2	SA	7.2 (1.5)	7.8 (1.2)	2.40	0.016
3	SM	7.4 (1.0)	8.0 (1.0)	1.33	0.18
4	ND	3.2 (2.0)	6.8 (1.7)	4.19	<0.001

6.3.3.2 VAST and WAB results

In section “6.3.3.1” above, we have focused our analysis on therapy outcomes measured by a single subtest within the VAST battery, the sentence production subtest. In this current section “6.3.3.2”, we will examine therapy outcomes on multiple linguistic levels as measured by all subtests within the VAST and WAB, including the sentence construction task, as shown in Table 6.5, Table 6.6, Table 6.7 and Table 6.8. Also, the statistical significance ($p < 0.05$) of within-subject therapy-induced changes was examined through the implementation McNemar test and Wilcoxon matched-pairs test of ordinal data (see Table 6.9, Table 6.10 and Table 6.11). Again, for a therapy study, we have reported 1-tail results.

Participant NS showed noticeable gains in performance on three subtests: VAST sentence comprehension, VAST object naming, and VAST sentence construction. More modest increases were noted on verb comprehension, fill-in verbs in sentences, grammaticality judgment, and action naming subtests, respectively, although these scores did not reach statistical significance. Similarly, a minimal change on semantic association subtests did not reach significance. On the other hand, NS’s performance on the WAB subtests showed a noticeable increase on the word fluency, sentence completion, Yes/No questions, and repetition subtests, respectively; nevertheless, only the last two scores were statistically significant. His performance on auditory word recognition, responsive speech, and object naming showed slight changes that did not reach significance. Also, a noticeable decrease in performance on the sequential commands was noted.

Participant SA showed noticeable changes in her performance on three subtests of the VAST: sentence construction, sentence comprehension and action naming subtests; however, only the first two were statistically significant. Her performance on the other VAST subtests showed less remarkable changes. Likewise, her performance on the WAB subtests was mainly consistent before and after therapy. Nevertheless, a noticeable observed on the object naming and word fluency subtests, respectively. However, none of SA’s scores on the WAB post-therapy subtests reached statistical significance.

Participant SM showed limited performance changes on all VAST subtests. Similarly, changes in her performance on the WAB subtests were limited except for two subtests. The participant showed noticeable gains on the word fluency and sentence completion subtests. Nevertheless, none of the above-listed scores reached statistical significance.

Lastly, participant ND showed a very striking increase on the action naming subtest of the VAST post-therapy. She also showed strong gains on the VAST fill-in verbs in sentences and sentence production subtests, respectively. All of ND's score increases reached statistical significance, though her scores on the remaining subtests within the VAST did not. It includes VAST sentence comprehension and VAST object naming subtests in which changes in scores were less prominent. Also, scores on the VAST verb comprehension, grammaticality judgment task, and semantic association were mainly stable. On the other hand, her performance on the WAB subtests before and after therapy showed a robust increase on sentence completion, repetition and object naming subtests, in which only the last two reached statistical significance. However, a substantial decrease in scores was noted on three subtests: word fluency, responsive speech and yes/no questions. Nevertheless, her performance on the sequential commands was stable with no change in scores, while a slight increase was noted on the auditory word recognition subtest, which was statistically significant.

In conclusion, looking at statistically significant therapy gains, the findings revealed that on the VAST subtests, participant ND and NS showed the most noticeable gains within the group, with statistically significant scores on two subtests of the VAST. It included a statistically significant score on the action naming (p=0.000) and fill-in verbs in sentences (p=0.001) subtests for ND, and the sentence comprehension (approaching significance at p= 0.0547) and the object naming (p= 0.0037) subtests for NS. Next, SA showed statistically significant gains on at least one subtest of the VAST, the sentence comprehensions subtest (p= 0.0313). Lastly, SM had a score that approached significance (p=0.0625) on the action naming subtest.

Table 6.10 Individual participant statistics for significance ($p<0.05$, one-tailed) of therapy gains on the VAST subtests using McNemar's test.

Participants' No. and initials		Verb compreh ension	Sentence compreh ension	Gramma ticality judgmen t	Action naming	Object naming	Fill-in verbs in sentence s	Semanti c associati on
1	NS	0.5000	0.0547	0.2266	0.1938	0.0037	0.5000	0.5000
2	SA	0.5000	0.0313	0.2500	0.1094	0.1094	0.3125	0.7500
3	SM	0.1250	0.5000	0.6367	0.0625	0.5000	0.5000	0.5000
4	ND	0.3125	0.3438	0.2905	0.0000	0.1094	0.0010	0.2500

The McNemar test was implemented to test the statistical significance of therapy gains on the WAB subtests. Participant ND showed the most gains post-therapy, with statistically significant improvements on 4 out of 8 subtests of the WAB: the auditory word recognition (p=0.0313, one-tailed), repetition (p=0.0313, one-tailed), and object naming (p=0.0156, one-tailed) subtests. Surprisingly, she showed a significant decrease in her performance on the word fluency (p=0.0313, one-tailed) subtest. Next, participant NS who showed statistically significant gains on the Yes/No questions (p=0.0313, one-tailed) and the repetition (p=0.0313, one-tailed) subtests, as well as a score approaching significance on the word fluency subtest (p=0.0625, one-tailed). On the other hand, participant SM presented a score approaching significance on the word fluency subtest (p=0.0625, one-tailed). Lastly, participants SA showed no significant changes in her performance on the WAB subtests before and after therapy.

Table 6.11 Individual participant statistics for significance (p<0.05) (*p*= one-tailed) of therapy gains on the WAB subtest using McNemar's test.

No. and Initials	Yes/No Qs	auditory word recognition	sequential commands	Repetition	Object naming	word fluency	sentence completion	responsive speech
1 NS	0.0313	0.5000	0.2500	0.0313	0.5000	0.0625	0.5000	1.0000
2 SA	0.7500	1.0000	0.6875	0.7500	0.5000	0.1250	1.0000	1.0000
3 SM	0.5000	0.2500	0.7500	0.5000	0.5000	0.0625	0.5000	1.0000
4 ND	0.0625	0.0313	0.7500	0.0313	0.0156	0.0313	0.2500	0.5000

6.3.3.3 Discourse production

In order to examine generalisation of therapy outcomes to discourse production, we assessed the participants' skills at four time-points, using the data elicited by the cookie theft (Goodglass and Kaplan, 1983) and the dinner party (Kertesz, 2007) picture stimuli. The dinner party picture stimulus underwent a slight alteration to its sequence to match the Arabic text's direction starting from right to left (see Chapter 5 for details). In analysing the elicited Arabic speech samples, the systematic analysis of language transcripts (SALT) software used to analyse English speech samples (see Chapter 4) was not an available option, as it is not compatible with Arabic. Therefore, we developed a new scoring rubric for Arabic discourse and selected the following variables to examine: word count, complete sentences, phrases, compound sentences, complex sentences, verbs, nouns, adjectives, and prepositions (see Appendix 1.C), based on them being the most clinically relevant to the type of therapy delivered. The selected variables were calculated for each transcript, elicited at four time-points: baseline 1, baseline 2, Interim 1, and interim 2. (See Table 6.12 and Table 6.13). All participants (n=4) underwent two baseline assessments, and their performance before therapy intervention was determined by the average score of baseline 1 and baseline 2.

Given the high levels of variability between and within neuro-typical controls and participants with aphasia, we have presented these data in descriptive terms rather than statistical analyses. To identify substantial changes, we focused on the percentage of increase in scores compared to baseline. Accordingly, we implemented the following formula to calculate the percentage of increase: (Interim 1 / Baseline) X 100 and (Interim 2/ Baseline) X 100. The criterion for substantial improvement was set at 20% and above increase in scores compared to baseline.

On both discourse tasks, cookie theft (CT) and dinner party (DP), as can be seen in Table 6.12 and Table 6.13, participant NS showed a steady increase in word count number at interim 1 (CT:175%, DP: 23%) and interim 2 (CT: 225%, DP:107%) that roughly doubled compared to baseline. Similarly, participant ND showed a very strong increase in word count at interim1 (CT: 33.3%, DP:85%) and interim2 (CT: 126.7%, DP:145%) that roughly doubled at each time-point. Participant SA demonstrated gains in word count number, albeit more prominent on the cookie theft discourse (interim 1: 31.3%, interim2:125%) than on the dinner party discourse (interim1: 4.5%, interim2: 54.5%). However, participant SM's gain in word count was more modest, with more increase noted on the dinner party discourse (interim1:12.5%, interim2: 19.6%) than the cookie theft discourse (Interim1: -14%, Interim2: 14.3%). Nevertheless, these results need to be interpreted with caution as the percentage of increase was affected by scores at baseline especially low scores (e.g., participant NS's word count on the cookie theft discourse at baseline was 4 points and increased to 13 points at interim 2 which resulted in a significant increase in percentage 225%) .

As can be seen from the data in Table 6.8, participants NS and ND showed that most prominent therapy gains in word count on discourse measures were also the ones who scored below the sample average at baseline and significantly below the controls' average performance. On the other hand, participants SA and SM, who showed more modest gains, performed above the sample average at baseline. Although SA and SM's scores at baseline remain below the average performance of controls, the gap between their performances is less substantial (range 42-70% below controls' average) than the gap between NS and ND's performance and that of the controls (80-90% below controls' average).

The initial clinical impression of the aphasia symptoms displayed by participants NS and ND led to non-fluent aphasia classification, while SA and SM were classified with fluent aphasia. However, since a distinction in severity was also noted between the two subgroups (non-fluent and fluent), it is unclear whether significant therapy gains were more associated with severity or type of aphasia at baseline.

Discourse Analysis results (Arabic group)

Table 6.12 Cookie theft discourse details per participant (raw scores)

Participant initials	Testing time-point	Word count	Phrases	Complete sentences	Compound sentences	Complex sentences	Total sentences	Verbs	Nouns	Adjectives	prepositions
NS	Baseline 1	2	0	1	0	0	1	1	1	0	0
	Baseline 2	6	0	1	0	0	1	2	3	0	1
	Average Baseline	4	0	1	0	0	1	1.5	2	0	0.5
	Interim 1	11	0	2	0	0	2	3	6	0	0
	Interim 2	13	1	2	1	0	3	3	7	1	1
% change	B to Int1	+175%	0	+100%	0	0	+100%	+100%	+200%	0	0
	B to Int2	+225%	+100%	+100%	+100%	0	+200%	+100%	+250%	0	+100%
SA	Baseline 1	14	1	1	1	0	2	2	7	1	1
	Baseline 2	18	0	3	1	0	4	3	6	1	1
	Average Baseline	16	0.5	2	1	0	3	2.5	6.5	1	1
	Interim 1	21	0	3	1	0	4	3	10	1	3
	Interim 2	36	0	5	2	0	7	5	18	1	1
% change	B to Int1	+31%	0	+50%	0	0	+33%	+20%	+54%	0	+200%

	B to Int2	+125 %	0	+150%	+100%0	0	+133%	+100%	+177%	0	0
SM	Baseline 1	15	0	2	2	0	4	5	7	0	1
	Baseline 2	20	0	3	2	0	5	5	9	0	2
	Average Baseline	17.5	0	2.5	2	0	4.5	5	8	0	1.5
	Interim 1	15	0	3	3	0	6	6	7	0	0
	Interim 2	20	0	3	3	0	6	6	9	0	1
% change	B to Int1	-14%	0	+20%	+50%	0	+33%	+20%	+12%	0	0
	B to Int2	+14%	0	+20%	+50%	0	+33%	+20%	+13%	0	+33%
ND	Baseline 1	8	0	0	0	0	0	2	5	0	0
	Baseline 2	7	1	0	0	0	0	2	5	0	0
	Average Baseline	7.5	0.5	0	0	0	0	2	5	0	0
	Interim 1	10	1	3	0	0	3	4	6	0	0
	Interim 2	17	1	3	0	0	3	4	10	1	1
% change	B to Int1	+33%	+100%	+200%	0	0	+200%	+100%	+20%	0	0
	B to Int2	+127 %	+100%	+200%	0	0	+200%	+100%	+100%	+100%	+100%

*Abbreviations: %= percentage, B= average baseline, Int1= Interim 1, and Int2=Interim 2.

**Formula used to calculate percentage of increase in scores: (Scores post therapy/ scores at baseline)*100

Table 6.13 Dinner party discourse details per participant (raw scores)

Participant initials	Testing time-point	Word count	Phrases	Complete sentences	Compound sentences	Complex sentences	Total sentences	Verbs	Nouns	Adjectives	prepositions
NS	Baseline 1	4	1	0	0	0	0	0	4	0	0
	Baseline 2	22	6	3	1	0	4	6	10	1	3
	Average Baseline	13	3.5	1.5	0.5	0	2	3	7	0.5	1.5
	Interim 1	16	3	2	0	0	2	5	9	1	0
	Interim 2	27	6	6	1	0	7	8	14	1	1
% change	B to Int1	+23%	-14%	133	0	0	0	+67%	+29%	+100%	-150%
	B to Int2	+108%	+71%	+300%	+100%	0	+250%	+167%	+100%	+100%	-33%
SA	Baseline 1	28	4	2	0	0	2	7	11	0	4
	Baseline 2	60	2	9	0	0	9	12	25	0	4
	Average Baseline	44	3	5.5	0	0	5.5	9.5	18	0	4
	Interim 1	46	5	2	1	0	3	11	18	0	1
	Interim 2	68	1	9	6	3	18	15	22	1	4
% change	B to Int1	+5%	+67%	-64%	+100%	0	-45%	+16%	0	0	-75%
	B to Int2	+55%	-67%	+64%	+600%	+300%	+227%	+58%	+22%	100%	0

SM	Baseline 1	67	1	10	4	0	14	17	30	1	5
	Baseline 2	45	1	7	2	0	9	12	20	0	0
	Average Baseline	56	1	8.5	3	0	11.5	14.5	25	0.5	2.5
	Interim 1	49	0	9	3	0	12	13	20	0	5
	Interim 2	67	0	12	6	0	18	18	25	0	6
% change	B to Int1	-12%	0	+6%	0	0	+4%	-10%	-20%	0	+100%
	B to Int2	+20%	0	+41%	+100%	0	+57%	+24%	0	0	+140%
ND	Baseline 1	14	2	0	0	0	0	2	6	0	0
	Baseline 2	26	2	1	0	0	1	4	16	2	0
	Average Baseline	20	2	0.5	0	0	0.5	3	11	1	0
	Interim 1	37	1	3	0	0	3	8	20	0	0
	Interim 2	49	1	7	1	0	8	11	26	2	0
% change	B to Int1	+85%	-50%	+500%	0	0	+500%	+167%	+82%	0	0
	B to Int2	+145%	-50%	+1300%	+100%	0	+1500%	+267%	+136%	+100%	0

*Abbreviations: %= percentage, B= average baseline, Int1= Interim 1, and Int2=Interim 2.

**Formula used to calculate percentage of increase in scores: (Scores post therapy/ scores at baseline)*100

Table 6.14 Summary of therapy gains per participant across different language modalities and linguistic levels

		Linguistic level	Skills (outcome measure)	NS	SA	SM	ND
Production	1	Word	Action naming (VAST)				✓
			Object naming (VAST or WAB object naming)	✓			✓
	2	Sentence	Picture description (VAST)	✓	✓	✓	✓
			Fill-in verbs in sentences (VAST)				✓
	3	Discourse (cookie theft and dinner party)*	Number total Words*	✓	✓		✓
			Number of sentences*	✓	✓	✓	✓
Comprehension	4	Word	Verbs (VAST)				
			Nouns (WAB auditory word recognition				✓
	5	Sentence	Sentence picture matching (VAST)	✓	✓		
			Grammaticality judgment (VAST)				
			Following sequential commands (WAB)				

✓= statistically significant, AS= approaching significant at 0.0625

*the criterion for substantial improvement was set at 20% and above increase in scores compared to baseline

Summary of overall therapy-induced improvements in language performance

Analyses of individual participant performance across all the outcome measures (Table 6.14) showed evidence of therapy-induced improvement, widespread in measures of expressive skills and some measures of receptive skills. In this analysis, we only included statistically significant scores as determined by McNemar of paired nominal data or Wilcoxon matched-pairs statistical tests, except for the discourse measures. We distinguished substantial improvements by setting a criterion at 20%, and above increase in scores compared to baseline and implemented the following formula to calculate the percentage of increase: (Interim3 / Baseline) X 100 and (Maintenance/ Baseline) X 100.

From this sample of four participants with chronic aphasia, participant ND showed the most noticeable therapy gains with a statistically significant increase in scores on 6 subtests representing 4 linguistics level of skills (word production, sentence production, discourse production, and noun comprehension). The second participant that presented an improved performance on several subtests was participant NS. He showed statistically significant gains in 4 subtests representing 4 linguistic levels of skills (noun comprehension, sentence production, discourse production, and sentence comprehension). On the other hand, participant SA showed a statistically significant increase in 2 subtests representing 2 linguistic levels (sentence production and sentence comprehension) and robust increases in discourse production, especially word count production. Lastly, participant SM showed statistically significant improvement in performance in 1 subtest representing sentence production and some evidence of enhanced discourse production, specifically sentence production, in elicited connected speech.

From this sample of 4 participants, NS and ND showed the most noticeable therapy gains with substantial increases in scores on 4-6 subtests representing 4 linguistic levels. Interestingly, these two participants were clinically classified with non-fluent aphasia; they also had lower starting points in terms of baseline severity than SA and SM, who had been classified with fluent aphasia, and demonstrated superior performance at baseline.

6.4 Discussion

This study evaluated a linguistically and culturally adapted hybrid therapy designed to remediate sentence and discourse production deficits in a small cohort of heterogeneous Arabic participants with stroke aphasia. Doing so allowed us to describe the patterns of language impairment profiles in Arabic speaking PWA, the central Saudi dialect, and compare these to patterns typical of English speaking PWA. Specifically, we analysed the performance of 4 participants on three main outcome measures at baseline: the Verb and Sentence Test VAST (Bastiaanse et al., 2002), Western Aphasia Battery WAB (Kertesz, 2007), and measures of elicited discourse. On the VAST and WAB, the two participants with fluent aphasia, SA and SM, showed relatively superior overall performance within the cohort, while NS and ND with non-fluent aphasia showed better comprehension relative to production performance. Regardless of broader severity, all 4 participants showed poor performance on the WAB word fluency subtest. Considering SA and SM's adequate overall performance on the language production subtests, a comparable performance on the word fluency subtest might have been expected. Nevertheless, this discrepancy in performance could be explained by the signs of speech apraxia they have exhibited throughout the session, which has not been formally tested in this study. Also, word fluency as a task is uniquely dependent on both language (semantic +/ phonological) and executive/speed of processing skills, which were not directly measured in this study.

The key performance parameters of aphasia which have supported aphasia diagnosis, such as fluency, comprehension and repetition skills (Goodglass et al., 2001; Kertesz, 2007), were informative in arriving at aphasia subtypes with these Arabic PWA. Referring to classical aphasia subtypes, participants NS and ND presented Broca's Aphasia characteristics, which included non-fluent speech production, functional comprehension, and poor repetition. Meanwhile, participants SA and SM presented characteristics of Anomic Aphasia with fluent speech, good comprehension, and the ability to repeat words and phrases with high accuracy. Although both NS and ND showed signs of Broca's aphasia, a difference between participants was noted in their relative performance with nouns versus verbs. Participant ND showed comparably poor performance on subtests that required either noun or verb retrieval; her performance was consistent across multiple linguistic levels (e.g., confrontational naming actions and objects, fill-in verbs in sentences, sentence completion with noun retrieval, and sentence construction). On the other hand, participant NS's performance was markedly poorer with verbs, specifically action naming and sentence production. Nevertheless, the outcome measures selected to test action and object naming in the current study (VAST subtests: action naming and object naming) require further investigation to confirm that the psycholinguistic features of the translated items in each subtest are matched. The author of the computer-based VAST battery (unpublished), Prof. Bastiaanse, confirmed that the verbs and nouns items were matched for psycholinguistic features in the English version; however, these data are yet to be published. Accordingly, the replication of the process to produce an Arabic equivalent was not available at the time of developing the materials for the current study. This information could be relevant in understanding the discrepancy in language production impairment observed in participant NS's performance.

A study by Alyahya et al. (2018b) showed evidence that word-class related differences in performance between verbs and nouns in a confrontation naming task were neutralized once the items were matched on several psycholinguistic features (i.e., word imageability, frequency, familiarity, age-of-acquisition, length and visual complexity). It was tested in a relatively large and diverse cohort of PWA (n=48), representing a wide range of aphasia subtypes and severities. Although these findings contradicted several studies that found a discrepancy in word processing between nouns and verbs (Bastiaanse and Jonkers, 1998; Luzzatti et al., 2002; Mätzig et al., 2009), it was further supported by a replication study (Alyahya et al., 2018b). When unmatched items (a subset of the original items) were used to assess the performance of the same cohort of patients (n=48), a discrepancy was found. It aligned with the above studies' results that suggested better performance of PWA on nouns compared to verbs. Accordingly, to interpret within-subject differences in verb and noun processing (e.g., the performance of participant NS), care should be taken to ensure that psycholinguistic features of the items used in assessment are matched. We can only attempt to further analyse word class differences in performance in our study when it is confirmed that items are matched for psycholinguistic features.

Although participants SA and SM exhibited generally adequate language production skills on the VAST and WAB subtests, their impairments were more clearly evident in the discourse measures. When their production was compared to the average word count produced by 6 healthy control participants, they scored markedly below average. Interestingly, the hierarchy of severity presented by the 4 participants with aphasia on the discourse measures aligned in broad terms with the severity of their overall performance on the VAST and WAB subtests combined, including the sentence production task. On the VAST and WAB test batteries, participant ND demonstrated the most severe language production profile followed in rank order by NS, SA, then SM, while on the discourse measures, participant NS scored the lowest followed in rank order by ND, SA and SM.

Even with our best efforts to capture the characteristics of language deficits observed in Arabic speaking individuals with aphasia, the sample size remains small. Accordingly, the range of symptoms and severities presented in the sample is limited (i.e., mainly moderately to mildly severe). It could be attributed to the short window of recruitment which was limited to 2-4 weeks. The SLT/P colleagues in Saudi Arabia, who collaborated in the recruitment process for this study, referred only patients who were actively engaging in SLT and support group sessions. Due to the lack of a database that stores the information and contact details of all discharged patients with aphasia. Having a database that collates all the cases and specifies their language skills on multiple linguistic levels, which also indicates the willingness to participate in future research, could have supported the current study's expansion.

Another limitation was related to the small sample size of the normative data. Due to time constraints and limited access to Arabic speaking healthy controls, our current data was collected from 6 neuro-typical healthy volunteers only. The normative data was required for planning, including materials development, before the primary investigator’s relocation to Riyadh for the research trials' commencement. Therefore, the information was obtained remotely through student collaborators from King Saud University in Riyadh, Saudi Arabia. To ensure that both groups, the PWA and healthy controls, spoke/were familiar with the central Saudi dialect (spoken in Riyadh, Saudi Arabia) and were exposed to this region's same culture. Nevertheless, a larger sample size of controls would strengthen the results of the normative data.

Participants response to therapy

The current computer-based therapy program's main therapy target was to remediate sentence production deficits in people with aphasia. The findings showed that the goal was achieved in all 4 participants who demonstrated a statistically significant increase in performance post-therapy on the VAST sentence construction task. It is important to note, the two participants SA and SM, who were more mildly impaired at baseline, still responded to the intervention as noted by their statistically significant improvement on the task. The less structured picture scene description tasks (cookie theft and dinner party), elicited discourse production, also captured remarkable increases in word count post-therapy compared to baseline in at least 3 out of 4 participants. Although improvements were noted, both baseline and post-intervention scores of the PWA group remained well below the average healthy control performance. Nevertheless, the results need to be interpreted with caution as the measure we implemented to track improvement in discourse production depended on the percentage of increase in scores from baseline.

Indirect therapy effects were also noted, as participants showed statistically significant improvements in skills that were not directly targeted in therapy, such as object naming, action naming, fill-in verbs in sentences, auditory word recognition, and sentence comprehension.

Overall, participants who showed more severe language production impairment characteristics such as ND and NS showed more noticeable therapy-induced gains on multiple linguistic levels than SA and SM, who performed close to the ceiling at baseline on most tasks. Nevertheless, significant therapy gains were still achieved on target therapy goals, sentence construction and discourse production in all participants regardless of their severity or aphasia type (i.e., fluent versus non-fluent). Our findings contradicted those reported by the VNeST studies, which found no relationship between the degree of impairment and the amount/extent of improvement (Edmonds, 2016).

Further interpretation of the outcomes of implementing this multilevel theory-driven hybrid approach to aphasia therapy in both languages English and Arabic (presented in Chapter 4 and Chapter 6) in light of wider literature will be revisited in Chapter 7.

Participants' acceptance of a computer-based aphasia therapy

At the time of conducting the therapy trials in Saudi Arabia, computer-based therapy approaches in treating Arabic adults with language production disorders did not exist. The barriers to this type of approach were explained in the introduction of Chapter 5. Accordingly, the current study aimed to examine the acceptability and learnability of this new computer-based approach in therapy to a sample of Arabic speaking individuals with aphasia from Riyadh, Saudi Arabia. These impressions were derived from clinical observations, patient and caregiver interviews, and an objective measure System Usability Scale SUS that quantified the users' feedback.

One of the main indications of this approach's feasibility and acceptability is the 100% retention rate. It involved a commitment to 12 to 13 visits to the hospital site in a weekly fixed time slot within a period of 10 to 12 weeks. Impressively, there were no incidents of delays or 'no shows'. Moreover, all participants actively engaged in all presented tasks during the sessions and showed compliance with the home practice requirements. However, in one incident, after the first week of therapy, participant SM reported that the picture description tasks were too easy and preferred more challenging stimuli. Nevertheless, she found the self-generated sentences in the VNeST tasks more challenging and engaging. The rest of the group relayed that the tasks presented an adequate level of difficulty. We here acknowledge that the preselected generic stimuli were purposefully implemented to systematically apply the same therapy protocol across a range of PWA with different language profiles. To examine the variance in their response to the same intervention. In clinical settings, it is recommended to tailor the stimuli to match each participant's lifestyle and personal interests and to customize the length of each phase/level of the therapy program according to the participant's response. For example, if the participant showed a high level of accuracy in Level 1 phase 1 (simple SVO sentence production in a picture description task) consistently on the first session, the next phase (self-generated SVO sentence with a written verb prompt, a VNeST task) can be introduced sooner. Therefore, this current study's same therapy protocol can be modified to suit either research or clinical settings. The criteria for progressing to the next level/phase in the research implementation depends on time duration (Level 1 phase 1 is implemented in week 1 and week 2 of the therapy program). In contrast, the clinical implementation should be constantly adjusted according to the participant's performance and response to therapy. We anticipate that clinical implementation of the current therapy protocol would be more enjoyable to participants and achieve more functional gains (through custom-tailored stimuli related to each participant’s lifestyle and interests). Lastly, the findings on the SUS scale further supported the clinical impression of the acceptability of the program. The adjective rating of the acceptability indicated by the scores of 4 participants and 3 caregivers ranged between good, excellent, and best imaginable.

The intervention program we have reported fitted into broader packages of care available to these participants with aphasia. The current clinical pathways for patients post-stroke in Ministry of Health MOH hospitals can be summarized into discharge to home with outpatient rehabilitation, inpatient rehabilitation, and long-term care facility “nursing home”, which applies to patients within the acute stage (i.e., defined as 90 days post-stroke onset in the article) and after (Al-Senani et al., 2019). To the best of our knowledge, the Speech and Language rehabilitation plan, including the frequency, intensity, and duration of the therapy sessions, is determined by the treating SLT/P and is case-dependent. They are usually not restricted by policies similar to those implemented by health insurance companies that set a cap on annual benefits or the total number of sessions for one beneficiary. Nevertheless, the shortage of SLT/P s in SA and their concentration in main cities meant that many patients are discharged home without access to regular therapy sessions, especially if they lived in rural areas. Based on the verbal report during case history and interview, the four participants in our study reported access to different SLT plans/settings following the onset of aphasia. For example, participant NS received SLT assessment and therapy session during his hospital stay (inpatient) and outpatient follow-up appointments after discharge. However, due to SLT/P shortage and the large caseload, the outpatient sessions were not as frequent as he would have preferred. Therefore, he pursued additional therapy sessions in a private clinic. Similarly, participant SM received inpatient, and outpatient SLT sessions in a MOH hospital complemented with additional therapy sessions in a private clinic. Moreover, she enrolled in an intensive course of comprehensive rehabilitation (i.e., included physical therapy, occupational therapy, etc.) which required her to be admitted to a rehab hospital for around 6 weeks. Participant SA also chose the same option and received the same intensive program twice, approximately one year apart. However, only patients with more than one disability (e.g., mobility and language) are eligible to apply for this type of intensive rehab program, and the duration of the wait time for admission could extend to months depending on the caseload and waitlist at the time of application. On the other hand, due to the long waitlist, participant ND did not receive any SLT outpatient sessions until 8 months post aphasia onset.

As illustrated by those four cases, even with close geographical proximity to several hospitals that offer SLT services in Riyadh, other factors influence the access, intensity, and frequency to those services. In light of the current situation, we anticipate that a computer-based language therapy pathway will provide a much-needed resource to bridge the current gaps in SLT service delivery and support equal distribution of services to all PWA.

In the current study, we aimed to translate existing and widely used test batteries. One of the goals was to ensure that at least some of the core outcome set COS recommended by Research Outcome Measurement in Aphasia (ROMA) were met, which will "address the heterogeneous measurement of outcomes in aphasia treatment research" and "facilitate the production of transparent, meaningful, and efficient outcome data" (Wallace et al., 2019, p.180). Also, having an Arabic equivalent of various well-known standardized measures will provide means for comparing language impairment profiles across languages beyond the core outcome set. Researchers and clinicians with no background in Arabic will still interpret the participants' performance on those tests. Additionally, some of the tests, such as the WAB, offer a total quotient of performance which is useful when measuring progress over time for research or clinical purposes.

Nonetheless, we have acquired valuable insight from the process of developing the present assessment and therapy materials. It includes identifying tasks that require further refinement to improve its validity and accordingly establish a plan to improve it. Likewise, to ensure the availability of a sensitive outcome measure that captures functional language characteristics that occurs in everyday communication. A multidisciplinary approach will be necessary to accomplish this aim, such as the contribution of Arabic linguists familiar with the variations within the Saudi subdialects. Moreover, due to the scarcity of previous normative literature, establishing a normative database will also be needed to determine the psycholinguistic features (e.g., age of acquisition, frequency, etc.). On the other hand, SLT/P scientists' contribution may include approaches that increase the validity of the test, such as matching the task difficulty (i.e., linguistic processing load) with the original test, match the length/duration it requires to complete the task, avoid priming which may interfere with the accuracy of results, etc. In conclusion, we recommend a combined approach to the creation of Arabic language assessment tools. It includes translating existing well-known test batteries with more language-specific linguistic tasks to ensure valid, reliable, and ecological measures.

Compared to the process of developing Arabic assessment tools, minimal efforts were required to create the Arabic therapy materials. It was largely due to the type of skills that we targeted in therapy, sentence construction and discourse production, in which elicitation depends on the picture stimulus. Therefore, the process focused on translating the task instructions and eliminating culturally irrelevant picture scenes from the selected workbooks within the SentenceShaper. Nevertheless, one of the limitations of this approach is that we were bound to the pre-selected verbs/actions available in these workbooks. Ideally, trained verbs should be selected according to their psycholinguistic features in Arabic (dialect-specific) and their relevance to each participant's lifestyle and personal therapy goals. Nevertheless, more flexibility was permitted with the VNeST task in which the primary investigator developed a list of 30 common verbs in Arabic that were relevant to the Saudi culture to be targeted in therapy.

Another possible limitation of our selected approach in eliciting sentence construction and discourse production was the difficulty of overcoming the participants' avoidance of complex sentence structures. Therefore, the therapy may have been restricted to the sentence structures that the participant chose to produce. A possible solution for including various sentence structures in therapy could be the implementation of sentence priming. However, further research is needed to determine the frequency of specific sentence structures in Arabic to determine the functional value of targeting each sentence type in therapy.

One of the successful decisions we have made in designing the therapy program has been selecting SentenceShaper as a medium to deliver therapy tasks. We initially used the computer software version with the English group (Chapter 4) and then switched to the iOS when it was made available, which is compatible with iPads. They were both easy to use and could be easily adapted to improve accessibility to participants with limited dexterity (e.g., the mouse cursor speed was adjusted for participant AD, mentioned in Chapter 4, which significantly improved her control and accuracy of selection). The new iOS version released in 2019 offered a compatibility option for languages with a right-to-left script direction, which was a necessary feature for the Arabic therapy trials. When this option was selected, the replay of the recorded snippets in the sentence row followed the right-to-left direction, which matched the Arabic text's direction. Nevertheless, the direction of page-turning remained the same as the original English copy, from left-to-right. Arguably, this did not pose any difficulties for Arabic speakers since most electronics, home appliances, and smartphones, which people use and interact with daily, apply the same layout. An additional feature of SentenceShaper that made it language-neutral and easy to use for delivering therapy in multiple languages was the ability to erase the built-in recorded instructions and re-record it in any other language. It also applied to verbal recordings stored within the vocabulary cues (i.e., includes buttons with or without visible text). Moreover, the text on the vocabulary cues buttons could also be deleted and new ones inserted. For this task, we used the "paint app" to create new buttons with Arabic text. Additionally, the pictures in all built-in therapy workbooks were replaceable. The features mentioned above within SentenceShaper allowed easy customization of the therapy materials to implement countless therapy tasks in many languages, allowing the adaptation of stimuli for patient-centred therapy plans (Hinckley, 2016).

In conclusion, the outcomes of the study supported the feasibility and acceptability of this novel computer-based sentence therapy method in treating Arabic speaking individuals with aphasia.

CHAPTER 7 Discussion

7.1 Summary of chapters and findings

Chapter 1 included an overview of the thesis's content, structure and research aims. **Chapter 2** presented the literature review that motivated our experimental work. In **Chapter 3** we examined sentence processing skills across a sample of 29 participants, with a range of language and cognitive skills, using a novel scoring rubric that was developed to gain specificity in performance measurement across the key skills in sentence production; the pilot testing showed high inter-rater reliability. The findings indicated that most participants who scored highly on the sentence production primary outcome measure (VAST sentence production task) were composed of the fluent-aphasia subgroup. We hypothesized that their scores were driven by their ability to generate word substitution when experiencing lexical retrieval difficulty. Accordingly, a future replication of the study with two suggested modifications is anticipated to unmask sentence production deficits in participants with fluent aphasia. First, introducing response timing in the assessment (e.g., 30-60 seconds, depending on the controls average), is likely to highlight the word finding/substitution attempts and would simulate the time pressure in everyday conversations. Second, the addition of constrained sentence production tasks that examines variety of sentence structures which was not covered in the current study, will provide a more clear overview of this subgroup's skills.

The outcomes also revealed, that fluent and non-fluent subgroups presented comparable performance on the language comprehension tasks, which was below the neurotypical controls' scores. The non-fluent subgroup's scores on the receptive and expressive skills showed a noticeable gap, with markedly poorer expressive skills. On the other hand, the fluent subgroup's receptive and expressive skills were comparable. The participants' scores on the LVET test, that examine light verb production, showed markedly low accuracy level relative to their scores on the VAST production subtests. A closer examintation according to aphasia subtype revealed that the non-fluent group's scores were significantly lower than their scores on the rest of the outcome measures. On the other hand, the fluent subgroup's performance on the LVET was within the range of their performance on the other outcome measures.

In **Chapter 4**, we tested the feasibility of a novel hybrid method in treating sentence production deficits in a sample of 12 English participants with aphasia. The outcomes showed that 83% of the sample showed statistically significant therapy-induced gains in their sentence production skills. The subgroup of participants within the sample that showed the most therapy gains were composed of those who performed at or below the group's average score at baseline, 67% of that subgroup were classified with non-fluent aphasia. Also, generalisation of therapy gains to discourse production elicited by the cookie theft or the dinner party picture scene description task was noticed in 92% of the sample on the MLU and total word count discourse measures; and 58% of the sample showed gains on the analysis set utterance measure generated by the SALT software. In some cases, indirect therapy gains were noted within the targeted modality (verb, noun, sentence, and discourse production) and across modalities (verb and sentence comprehension and grammaticality judgment). Overall, the participants required at least 8 weeks of therapy, following the same protocol described in the chapter, to start noticing statistically significant gains on standardized language assessments. Therapy gains were maintained 6 week after the discontinuation of direct-therapy.

The statistical analysis of the participants' performance on naming, cognitive, semantic and phonology tests (i.e., BNT, Raven's, 96 synonym judgment task, and PALPA 8, respectively) showed that strong predictors of therapy gains in sentence production skills were limited to baseline scores on naming and semantic tasks (BNT and 96 synonym judgment task) only. In summary, the outcomes of Chapter 4 demonstrated the feasibility of our novel method in treating English speaking participants with aphasia.

Chapter 5 described the methods we used to produce a preliminary version of the Arabic assessment tools WAB, VAST, and COAST and the results of a pilot test with a sample of 6 native Arabic speaker volunteers. It also included an explanation of the process of translation and adaptation of the therapy materials to deliver therapy in the Saudi Arabic dialect. Furthermore, the challenges we faced in creating a translated and culturally adapted version of the assessment tools and our suggestions to overcome the barriers in future research were highlighted in the chapter. Also, the rationale for adjusting the time-frame for delivering the therapy protocol to Arabic participants with aphasia was explained.

The pilot testing in **Chapter 6** of the linguistically and culturally adapted hybrid sentence therapy method showed evidence of its feasibility. The sample of participants represented two participants with fluent aphasia with high performance in sentence production skills and two with non-fluent aphasia and low performance, they all benefitted from the intervention. All 4 participants showed statistically significant therapy-induced improvement in sentence production skills. They also demonstrated generalised therapy gains to discourse production on the total word-count measure. The study also captured the characteristics of language deficits observed in Arabic speaking individuals with aphasia at baseline and compared them to the classic aphasia subtypes presented in the English literature (Goodglass et al., 2001). In some cases, indirect therapy gains in object naming, action naming, fill-in verbs in sentences, auditory word recognition, and sentence comprehension tasks were noted.

The acceptability and learnability of a computer-based aphasia therapy approach to Arabic PWA and the Saudi culture was explored through the implementation of the System Usability Scale (SUS). The adjective rating collected from a sample of 4 participants and 3 caregivers indicated an acceptability and learnability rate that ranged between good, excellent, and best imaginable. The outcomes further supported the feasibility of this novel method in treating Arabic speaking individuals with aphasia.

The current **Chapter 7**, will discuss the novelty of the therapy method design presented in the thesis, interpret the outcomes (English and Arabic) in light of wider literature and its novel contributions including its impact on the Arabic aphasia literature. It will also explore the advantages of our computer-based approach in light of post-Coronavirus disease 2019 COVID-19 adjustments in SLT service delivery. Finally, it will explain the study limitation and directions for future research.

7.2 The innovation of a multilevel computer-based hybrid sentence therapy program

The therapy program we presented in this thesis matches the care bundle's definition that implements the integrative model as described by Hinckley (2017) (see Chapter 2 within the thesis for further details). We integrated the delivery of three different impairment-focused and evidence-based methods and delivered the therapy in a hierarchical, multileveled format. This combined the principles of the VNeST (Edmonds et al., 2009) and mapping therapy (Schwartz et al., 1994), delivered through SentenceShaper that provides processing support (i.e., widening of the temporal window) (Linebarger et al., 2007). A fundamental aspect of the integrative model is that “the ultimate goal of the activity or the task is preserved across all selected therapies” (Hinckley, 2017, p. 349). Hence, our study's consistent aim across the therapy tasks was to support independence in producing complete, informative, and grammatically correct sentences verbally in a picture description task (simple sentences). Beyond this, the program aimed to prompt sentence expansion or combine two sentences or more in a discourse, in a picture scene description task. Nevertheless, a clear distinction from the traditional integrative model is that the desired activity (therapy goal) in our study was pre-determined by the research team and delivered systematically across participants with a wide range of language impairment profiles. One of the main purposes of doing so was to investigate the different responses to the same therapy method across participants.

7.3 Interpretation of therapy study outcomes (English and Arabic) in light of wider literature

The thesis presented the outcomes of implementing this multilevel theory-driven hybrid approach to aphasia therapy. Our findings showed evidence of this method's feasibility in remediating a range of language deficits in PWA with different language profiles in both languages, English and Arabic. It yielded statistically significant improvements in sentence construction skills in the vast majority of participants in each group (83% of the English group, 100% of the Arabic group, and 87.5% across both groups). In addition to achieving gains on untrained items of targeted skills (i.e., sentence and discourse production), it also generated indirect gains on multiple linguistic skills.

The VNeST has been classified in the literature as a therapy that aims to improve verb retrieval, while the mapping therapy as an approach that aims to improve sentence processing (Marshall, 2015). Accordingly, our hybrid approach that combines the two in addition to the processing support (i.e., Sentence Shaper) could be categorized as a multilevel treatment as described by Webster et al. (2015) since it works across multiple linguistic levels (i.e., word level and sentence level).

Comparing our findings to multilevel therapy studies reported in the literature proved to be challenging due to the wide variations in the elicitation methods and outcome measures across studies. Furthermore, a comparable study to the one we presented does not exist in the literature to this date. Therefore, we resolved to compare our findings to studies that reported individual implementation of each of the three approaches that compose our hybrid method.

Table 7.1 Summary of significant therapy gains per participant across different language modalities and linguistic levels.

Modality		Linguistic level	Skills	English	Arabic	Total	% of the sample
				n=12	n=4	n=16	n=16
Production	1	Word	Action naming	2	1	3	19%
			Object naming	1	2	3	19%
	2	Sentence	Picture description	10	4	14	87.5%
			Fill-in verbs in sentences	0	1	1	6.25%
			LVET	1	N/A	N/A	-
	3	Discourse (either cookie theft or dinner party)*	Analysis set utterance	7	4	11	69%
			MLU	11	N/A	N/A	-
			Number total Words	11	3	13	81%
Comprehension	4	Word	Verbs	3	0	3	19%
			Nouns	1	1	2	12.5%
	5	Sentence	Sentence picture matching	1	2	3	19%
			Grammaticality judgment	0	0	0	0
			Following sequential commands	0	0	0	0

*Discourse analysis methods were different for each language

N/A= not applicable

7.3.1 VNeST (Edmonds et al., 2009)

A review of the outcomes of studies that implemented VNeST as a single therapy approach showed preliminary evidence of its efficacy in improving lexical retrieval at the word, sentence, and discourse levels (Edmonds, 2016). The review examined the findings of a pool of 19 English speaking participants with a range of aphasia types and severities reported in 5 studies (Edmonds and Babb, 2011; Edmonds et al., 2014; Edmonds et al., 2009; Edmonds et al., 2015; Furnas and Edmonds, 2014). Another study by Kwag et al. (2014) investigated the outcomes of a modified version of the original VNeST therapy protocol in treating 3 monolingual Korean speakers with aphasia. Findings were similar to the ones reported by Edmonds and colleagues with the English speakers; however, generalization to discourse was not tested in this study. Accordingly, the findings demonstrated the success of implementing VNeST in languages other than English.

Overall, the finding reported in this literature indicated generalized improvements in lexical retrieval at the word level. The sample of participants' performance (n=19) on standardized tests revealed that 86% of the sample improved in noun naming (measured by either OANB or BNT and verb naming subtest from the NAVS), and 58% of the sample also improved in verb naming (Edmonds, 2016, p.127). The authors hypothesized that this was attributed to the semantic network activation induced by VNeST. However, this extent of improvement in lexical retrieval at the word level was not replicated in our study since only 19% (n=3) of the total sample of participants (n=16 both English and Arabic) achieved a statistically significant increase in scores post-therapy. Nevertheless, we argue that high accuracy level of performance on confrontational naming tasks is not a strong indicator of an equally strong performance on sentence and discourse production which were the aims of our intervention. Also, a number of studies found no association between improvements in lexical retrieval and gains in functional communication, which is the ultimate goal of speech and language therapy (Carragher et al., 2012; Webster and Whitworth, 2012; Wilshire and McCarthy, 2002).

A constrained sentence production task (i.e., using pictures from the NAVS for sentence elicitation) showed that 75% of the sample improved on untrained sentence production skills. It is important to note that the sentence scoring approach in the constrained task in VNeST studies did not account for grammatical errors (Edmonds et al., 2009), which was a fundamental variable in our scoring rubric that determined accuracy. Therefore, caution should be taken when comparing the VNeST studies' findings on the constrained sentence production task and the ones we have presented. Nevertheless, even with the advantage given in scoring the participants' performance in the VNeST studies on the sentence production task, the numbers in our study exceeded it. 87.5% of the participants in our sample showed statistically significant improvement on the VAST sentence construction task (i.e., a constrained task, although different than the NAVS used in the VNeST studies). However, a limitation to this comparison is the discrepancy in the samples' size (n=16 in our study vs n=19 in VNeST literature).

Lastly, sentence production in discourse, as measured by complete utterances CUs (this measure is not standardized, but high interrater reliability has been reported), was only tested with English participants in the VNest literature and revealed that 59% of the sample showed significant gains (Edmonds, 2016). Likewise, the percentage of participants in our sample that showed statistically significant therapy-induced improvements exceeded that number. We implemented a different methodology than the one used in the VNeST studies, and our method also differed by language. For example, with the English discourse samples, we used the analysis set of utterance generated by the SALT software to counted for all complete, informative, and relevant sentences in a given discourse. However, with the Arabic discourse samples, we implemented a manual speech sample analysis that looked for the same criteria (see Appendix 1.C for details). Additionally, we looked into the changes in the total number of words produced in a sample before and after therapy. Although statistical analysis has not been implemented to determine the significance of changes, we established a criterion that identified changes above a 20% increase in scores from baseline as a robust improvement. Accordingly, our findings showed that of the total number of participants in both language groups (n=16), 69% improved on the amount of complete, relevant, informative and grammatically correct sentences produced in discourse (i.e., either cookie theft of the dinner party), and 81% of the sample showed a noticeable increase in the total number of words produced in a given discourse. Similarly, acknowledging the limitation of using percentages across different participant sample sizes, our results exceeded those reported by studies that implemented VNeST as a single therapy approach.

In summary, our findings did not match the VNeST studies. The number of participants that showed therapy-induced gains in lexical retrieval in our study was much less than the one reported in VNeST studies. On the contrary, the number of participants who improved on sentence production skills and sentence construction in discourse was markedly higher in our study.

All the above mentioned VNeST studies implemented a high dose of therapy with at least two sessions per week, totalling 3-3.5 hours per week (i.e., duration ranged between 4-15 weeks with the majority receiving 10 weeks of therapy). A more recent single case study by Parkes (2017) investigated the outcomes of delivering a lower dose, 1.5 hours per week of clinician-direct therapy sessions (total number of hours was 4.5 hours), of VNeST therapy to treat a participant with Broca's aphasia. However, the study's outcomes did not replicate the findings reported by Edmonds (2016). On the other hand, the therapy dosage and intensity of clinician-directed sessions have been much lower in our study as participants received only one 90-minute clinician-directed session per week, along with independent home practice that averaged 3 hours per week (i.e., therapy period was 12 weeks for English group and 8 weeks for Arabic group). Also, the VNeST approach was embedded within a multilevel approach, which meant that the time spent practising VNeST tasks in the clinician-guided sessions was less than the total duration of the session.

In conclusion, our findings indicated that as a component of a specific hybrid multilevel therapy approach, the implementation of VNeST in a lower dose and intensity was effective in remediating sentence and discourse production (as measured by structured tasks) across English and Arabic speaking individuals with aphasia.

7.3.2 Mapping therapy (Schwartz et al., 1994)

The mapping therapy studies encompass a group of highly diverse techniques, strategies, and outcome measures (Berndt and Mitchum, 1997; Byng, 1988; Byng et al., 1994; Dorze et al., 1991; Haendiges et al., 1996; Jones, 1986; Marshall et al., 1997; Mitchum et al., 1997b; Mitchum et al., 1995; Nickels et al., 1991; Rochon et al., 2005; Schwartz et al., 1994). Nevertheless, a common finding shared by these studies has indicated that mapping therapy was effective in remediating sentence processing in trained structures (e.g., active, passive, object cleft, etc.) and trained modalities (comprehension, production), except for the comprehension of non-canonical sentences in which mixed results were found (Beveridge and Crerar, 2002; Schwartz et al., 1994). Also, a study by Rochon et al. (2005) found improvement in the trained production of canonical and non-canonical sentences and generalization of therapy gains to narrative production.

Rochon and colleagues (2005), tested the efficacy of the mapping therapy approach in treating sentence production in 3 participants with chronic Broca's aphasia, delivered in an average of 19 sessions of 1-hour biweekly sessions (over the course of 6 months which included interim testing sessions). The training included canonical sentences (active and subject cleft) and non canonical sentences (passive and object cleft). To assess therapy-induced changes in sentence production, two constrained sentence production tests were used, the Caplan and Hanna's Sentence Production Test (Caplan and Hanna, 1998) and the Picture Description with Structure Modeling Test (Fink et al., 1995). The results showed improvements that were limited to the production of trained sentences structures and did not include untrained structures. Therapy gains were maintained at one month post therapy. However, no cross-modality therapy gains were noted, as the participants did not show improvement in their sentence comprehension skills.

In our study we have used the sentence construction subtest in the Verb and Sentence Test VAST (Bastiaanse et al., 2003), a different sentence elicitation task than the two implemented in Rochon et al. (2005) study. In the Caplan and Hanna's sentence production test, the root form of the verb is provided and visual cues to prompt the inclusion of all items in a picture stimuli were included. On the other hand, the PDSM test restricted the participant's production to a target (pre-selected by the examiner) sentence structure. Those features were not present in the VAST, as it involved a line drawing picture stimuli with no additional verbal or visual cues. Also, the elicitation was not restricted to a specific sentence structure. The produced sentences that were grammatically correct, complete, informative and relevant sentence produced, in any sentence structure, received a full score. On the VAST sentence production subtest, our findings across the two groups (English and Arabic) showed similar results, as 87.5% of the sample of participants demonstrated statistically significant therapy gains. However, our therapy tasks included only canonical sentences and the outcome measure did not distinguish the structure of the produced sentences.

The generalization of therapy gains to narrative construction was examined through a Cinderella story retelling task and the QPA discourse analysis method in Rochon et al. (2005). The performance of the 3 participants showed an increase in MLU of 0.0, 0.5, and 1.0. Their scores at baseline ranged between 2-4.5 which is significantly lower than the controls' score (12 neuro-typical subjects reported in (Rochon et al., 2000)) of 8.17 (SD=1.39)

The only common variable in our discourse analysis and the one reported by Rochon et al. (2005) is the Mean Length of utterance MLU which was generated by the SALT software for the English group only (n=12). The narrative samples were elicited by the cookie theft and dinner party picture scene stimuli. The MLU increased between 1.65 and 3.9 post therapy (in 11 out of 12 participants), which is higher than the scores reported in Rochon et al. (2005). Nevertheless, due to the differences in sample size (3 vs 12 participants) and variation in baseline scores (2-4.5 vs 2.4-8.3 MLU scores) the results should be interpreted with caution.

7.3.3 SentenceShaper (Linebarger et al., 2001)

Several studies investigated the outcomes of using SentenceShaper as a therapy tool (Linebarger et al., 2004; Linebarger et al., 2007; Linebarger et al., 2001; McCall et al., 2009). The intervention included practising story retelling from wordless picture books or silent videos and producing narratives of personally relevant topics, movies, or television shows. After the initial training on using the program, the practice was carried out independently by the participants at home with weekly follow up visits to the research site. The assessment materials were similar to those used in training; however, specific items were reserved for assessment purposes only and were not included in therapy. The participants' spontaneous unaided verbal narratives were analyzed using the Quantitative Production Analysis system QPA (Saffran et al., 1989) to determine intervention-induced changes. After at least 15 hours of home use, the following characteristics were noted: increased MLU, more structured utterances, improved grammatical well-formedness, and an increased proportion of words in sentences. Also, narratives scored higher on informativeness measures. Only one case study by McCall et al. (2009) examined the outcomes of practising syntactically complex sentences (subordinate clauses) following a period of implementing the general approach described in previous studies. Their findings indicated that this approach produced further noticeable improvements (i.e., following 27 hours of practice over 5 months, words per sentence increased from 3.6 to 8.12 following the general therapy approach, then increased further to 11.56 words per sentence when syntactic structures were targeted in practice). The authors concluded that "practising production in narrative-level contexts with processing support is sufficient to induce at least some structural improvements without specific sentence-level training." (McCall et al., 2009, p.455). In our study, the discourse measures and the analysis methods we used were different from those used in SentenceShaper studies and distinct for each language. However, the Mean Length of Utterance MLU was a common analysis variable in both studies, although it was generated by the SALT software in our English study and by the QPA method in SentenceShaper studies. In our study with the English participants, we found that 67% of the participants (8 out of 12 participants) showed at least 2.2 increase in MLU from baseline with a range of 2.2-3.9, either immediately post-therapy or at maintenance testing (on either the cookie theft elicitation task or the dinner party). The remaining 25% of the sample showed less increase with a range of 1.65- 1.8 MLU. Nevertheless, caution should be taken when comparing the results from both studies, as

SentenceShaper studies included only participants with non-fluent aphasia while the sample of participants in our study encompassed a range of aphasia types, including fluent aphasia. Accordingly, performance at baseline on the MLU variable in our sample was diverse and ranged between 2.4 – 8.3 on either the cookie theft or the dinner party discourse. In comparison, the participants' scores in the SentenceShaper studies were at the lower end of that range, possibly related to their non-fluent aphasia classification. This distinction could explain the limited increase noticed in our sample compared to participants' performance in SentenceShaper studies. Another reason could be related to the type of intervention in our study that involved structure-specific sentence therapy rather than the narrative construction training reported in SentenceShaper studies. Although our approach included sentence expansion and combining sentences, it is likely that these tasks did not engage message-level processing (therapy at the level of the event) (Marshall, 2009; Marshall, 2017) to the same extent that a narrative construction task would.

Overall, the studies showed evidence that SentenceShaper was useful in inducing at least some structural improvements in narrative production in people with fluent aphasia. Similar findings were observed in our study across languages, although our approach was composed of impairment-based explicit training of sentence production, expansion, and discourse construction. Our study also extended the research sample to include participants with fluent aphasia, mainly anomia or conduction aphasia, who similarly demonstrated a positive response to intervention as the non-fluent participants. Nevertheless, it is currently unclear if the impairment-based approaches we have incorporated added further value to the amount and quality of improvements noticed in the participants compared to the approach reported by the SentenceShaper studies.

In conclusion, the mapping therapy on its own may repair and strengthen the mapping procedures between the thematic roles and grammatical roles, but is unlikely to induce semantic network strengthening that the VNeST provide. This added benefit is anticipated to improve word retrieval in sentences that extends to untreated verbs and nouns, and thus improves time-efficiency. On the other hand, implementing the VNeST on its own is unlikely to improve the grammaticality of the produced sentences since it is not included in training nor monitored in the outcome measures. Finally, SentenceShaper on its own is unlikely to achieve the same level of metalinguistic awareness that a clinician-directed and impairment-based approaches would achieve. It is anticipated that improving metalinguistic awareness will increase the success of self-correction attempts which SentenceShaper supports.

7.4 Summary of the novel contributions

In the context of the current scarcity of studies investigating the outcomes of combining different impairment-focused therapy approaches into a care bundle (see Chapter 2), our findings represent a novel contribution to the aphasiology treatment literature. First, we explained our reasoning and the evidence that supports our method of integrating this novel bundle of impairment-focused approaches: the mapping therapy, VNeST, and SentenceShaper as a processing support to widen the temporal window. We also provided a detailed description of the innovative approach we adopted in modifying SentenceShaper software to create a language-neutral medium for delivering impairment-based language therapy across languages. Then, we tested our method's efficacy with a sample of participants that encompassed a range of language impairment profiles (n= 16) in two distinct languages (English and Arabic). The outcome analysis we presented in the thesis was not limited to targeted and untargeted language skills. It extended to include an investigation of response variations to therapy across participants following a systematic application of our pre-designed therapy program. Furthermore, we attempted to search for patterns that could link therapy outcomes to aphasia type and severity at baseline. Lastly, the analysis was supplemented with an investigation of the computer program's usability and learnability (i.e., as measured by the System Usability Scale SUS) and the self-reported communication effectiveness before and after the intervention (COAST).

7.5 Contributions specific to the Arabic aphasia literature

The study we presented is the first to examine aphasia language profiles in native Arabic speakers of the Saudi central dialect in a sample of 4 participants with a range of age groups, educational level, and time of onset to the best of our knowledge. To date, this is the only study that has investigated the outcomes of applying an impairment-focused therapy approach in treating language production deficits in Arabic speaking individuals with aphasia (Saudi dialect). Before this, the only other study that investigated an impairment-focused approach in remediating aphasia in Arabic was a single case study that reported the outcomes of implementing the Melodic Intonation Therapy MIT in treating a person with aphasia speaking the Jordanian dialect (Al-Shdifat et al., 2018). The investigation we delivered was the first of its kind in Saudi Arabic speakers with aphasia. Although the sample size of four participants is considered limited for making firm observations of the trends identified in performance, either at baseline or post-therapy, the findings were indicative of this therapy method's feasibility. Nonetheless, we aim to further refine the assessment tools and carry on the testing process to create a large database of Arabic aphasia profiles. It will capture the language skills of a wide range of aphasia subtypes and severities identified in native Arabic speakers of the Saudi dialect (all subdialects). It will help define the error types that are most common in this population and the most common language impairment profiles. Moreover, this database will also serve as an effective and time-efficient source for identifying research candidates for recruitment.

7.6 The advantages of our computer-based approach in light of post-Coronavirus disease 2019 COVID-19 adjustments in SLT service delivery

In 2017, the time of starting the current project, computer-based language therapy was limited to specialized technology centres that offered Augmentative and Alternative Communication AAC fittings and training and some school-based SLT services that provided support to students with AAC devices. In out-patient SLT clinical settings, there were a number of known computer programs for children with voice disorders or stuttering which were mainly focused on providing the participants with either visual reinforcement for performance or biofeedback in the context of computer games (King et al., 2012; Umanski et al., 2008). On the other hand, due to infection control guidelines and hygiene purposes, in-patient clinical settings focused mainly on using low-technology AAC devices to enable the in-patient to communicate basic needs and thoughts if and when needed.

Moreover, prerequisite academic coursework for the qualification/ licensure of Speech-Language Therapists/Pathologists did not include courses focused on AAC or SLT computer-based training and service delivery. If a given undergraduate/graduate program offered the specialized courses, they were classified as elective. Therefore, many qualified SLT/Ps did not receive enough training either through undergraduate/graduate curriculum coursework or clinical internships/placements to reach confidence in delivering computer-based SLT services (Alateeq, 2014).

Therefore, at the time of designing the project presented in this thesis, there was no pressing need for computer-based and tele-therapy approaches to SLT services. Many SLT/P s expressed their strong preference for in-person sessions, in which the motivation of the participant can be easily examined and level of support/ reinforcement can be adjusted accordingly. Also, many relied on paper and pencil cueing strategies, face-to-face speech articulation modeling (e.g., for dysarthria errors), and non-verbal/total communication methods which can be difficult to demonstrate in other than in-person settings. Nonetheless, we predicted that the need for computer-based therapy will arise in the future and tele-therapy will become a mainstream at some point. Accordingly, we implemented it in the design of our program. We reasoned that this approach presented an obvious solution for the shortage of qualified SLT/P s in some countries/cities which commonly prolongs the waitlist of patients and wait time to receive the much needed SLT services post stroke onset. Also, this method provides means for increasing the intensity of therapy, in which its benefits are well established in the literature (Cherney et al., 2008b), without increasing the financial burden. Nevertheless, the acceptability of this approach to the consumers (i.e., SLT/P s, clients, caregivers) had to be addressed first.

Today, the COVID-19 pandemic mandated essential adjustments to lifestyle and service delivery which included resolving to tele-therapy as a single mode of delivery for speech and language therapy services. It was necessary to protect vulnerable clients/patients who were shielding and those at high risk of developing life-threatening health complications upon contracting the infection, which likely applied to a large proportion of PWA. Additionally, when national lockdowns were implemented, and clinicians were required to work from home for several months, the only available option was to deliver therapy remotely via tele-therapy. These circumstances forced everyone (i.e., SLT/Ps, clients, and caregivers) to learn and adapt to this new way of delivering SLT services. As a result, most SLT/Ps worldwide have now gained the training and experience to accept computer-based approaches in therapy and the requirements to excel in delivering it. Now, more than ever, our current therapy program's contribution will be tested and utilized thoroughly.

7.7 Study limitations

Throughout developing and testing the research projects presented in this thesis, we have identified a number of limitations and possible ways of addressing them in future research. First, the sample size was limited, especially the Arabic group. A replication study with a large sample size will support our findings and enable us to make firm conclusions. Next, increasing the validity of the translated and adapted language assessment tools (i.e., from English to Arabic) proved challenging in light of the current limited resources in the literature. Therefore, it would be useful in future research to adopt a more thorough approach and collaborate with a multidisciplinary team, including a linguist familiar with the Saudi subdialects. In order to match the psycholinguistic features of the Arabic version with the original test as well as to ensure strong the validity of the tasks. Also, some language skills, such as sentence comprehension and grammaticality judgment tasks, would require the task development to be derived directly from the principles of Arabic linguistics. As the translation and adaption from pre-existing English test batteries overlooked many fundamental elements of Arabic and proved to be insufficient. Adopting this suggested alternative will likely ensure better comprehensiveness of the test, especially in capturing functional verbal communication of everyday life in Arabic.

Another limitation is related to our limited understanding of the factors that might have influenced PWA's performance on baseline outcome measures (e.g., SM’s apraxia could have interfered with her performance on the word fluency task). A possible solution to incorporate in future research is the addition of supplemental tests for the assessment of apraxia, dysarthria, verbal memory and executive function skills, and functional communication skills.

Finally, the self-monitoring and self-correction skills that SentenceShaper engaged in therapy were novel additions to the structured impairment-based method. The current study's design did not examine performance on those two skills at baseline or post-intervention. It would be beneficial to include it in the baseline performance assessment and the interval testing throughout the course of therapy. Doing so will enable us to assess the correlation between gains in self-monitoring and self-correction skills (i.e., likely induced by SentenceShaper practise) and gains in sentence and discourse production.

7.8 Directions for future research

The lack of standardized assessment tools, normative data, or studies investigating therapy outcomes in Arabic speaking individuals with aphasia is a clear barrier to Evidence-Based Practice EBP. It prevents the advancement of speech and language therapy services in Arabic speaking countries. Nevertheless, researchers’ interest in this field is usually challenged by the current scarcity of resources.

Our experience in conducting aphasia research in Saudi Arabia (SA) helped us reach a strategic plan to address the issues. It consists of creating a database of participants with aphasia, which specifies all medical background information and performance on language and cognitive assessment batteries. We expect implementing this proposed solution to support researchers in allocating resources within SA effectively for either in-person research projects or remotely conducted projects (e.g., retrospective studies). This facilitation will encourage the participation of an increased number of researchers within SA and abroad. Ultimately, it will enrich the aphasia literature with studies on Arabic speakers and possibly help establish international research collaborations.

Essentially, the validity testing and standardization of the assessment tools need to proceed. Therefore, one of the near future research projects would involve expanding the network of student and clinician collaborators who took part in the current research to carry over collecting normative data, translating and adapting assessment tools, and pilot testing the materials. It will serve as a mutually beneficial collaboration that will create research opportunities for students and clinicians and support researchers with volunteer research assistants. As a result, more goals can be accomplished in less time. Moreover, the inclusion of Arabic linguists, who are familiar with the Saudi subdialects, in the team will be vital to accomplishing the following aims: 1. categorize and describe the psycholinguistic features (e.g., the frequency of specific sentence structure, age of acquisition, etc.) found in the language samples of native Arabic speaking neuro-typical controls representing a range of Saudi subdialects. This step is the foundation that will inform our selection of testing/therapy items in future research. 2. contribute to the analysis of aphasia error types identified in the performance of Arabic speaking individuals with aphasia from a linguistic point of view. 3. support the process of translating and adapting existing language assessment tools into Arabic by establishing equivalence of psycholinguistic features in both versions.

Following the establishment of valid and reliable language assessment tools and improved access to candidates for recruitment through a database of participants, more therapy studies can be conducted. For example, we can replicate the therapy study presented in this thesis to include a large sample of Arabic participants with aphasia. We can also create an updated version of our therapy program to incorporate newly identified interesting concepts such as speed- and accuracy-focused intervention (Conroy et al., 2018), which showed evidence of significantly increased generalization of therapy gains to connected speech. Likewise, an interesting topic for future studies would be to examine the effect of combining our present therapy program with a participation-focused therapy (Hinckley, 2017) to facilitate the generalization of skills to different contexts, including spontaneous speech production. Outcome measures may include tests that measure functional communication, such as the scenario test (van der Meulen et al., 2010) and speech sample analysis of procedural narratives and unstructured conversations.

REFERENCES

Abou El-Ella, M., Alloush, T., El-Shobary, A., El-Dien Hafez, N., Abd EL-Halim, A. & El-Rouby, I. (2013). 'Modification and standardisation of Arabic version of the Comprehensive Aphasia Test', *Aphasiology,* 27(5), pp. 599-614.

Acquadro, C., Conway, K., Girourdet, C. & Mear, I. (2004). 'Linguistic validation manual for patient-reported outcomes (PRO) instruments. 2004', *Lyon: Mapi Research Institute*.

Akoglu, H. (2018). 'User's guide to correlation coefficients', *Turkish journal of emergency medicine,* 18(3), pp. 91-93.

Al-Senani, F., Al-Johani, M., Salawati, M., ElSheikh, S., AlQahtani, M., Muthana, J., AlZahrani, S., Shore, J., Taylor, M. & Ravest, V. S. (2019). 'A national economic and clinical model for ischemic stroke care development in Saudi Arabia: A call for change', *International Journal of Stroke,* 14(8), pp. 835-842.

Al-Shdifat, K. G., Sarsak, J. & Ghareeb, F. A. (2018). 'Exploring the efficacy of melodic intonation therapy with Broca's aphasia in Arabic', *South African Journal of Communication Disorders,* 65(1), pp. 1-8.

Al-Thalaya, Z., Nilipour, R., Sadat Ghoreyshi, Z., Pourshahbaz, A., Nassar, Z. & Younes, M. (2018). 'Reliability and validity of bedside version of Arabic Diagnostic Aphasia Battery (A-DAB-1) for Lebanese individuals', *Aphasiology,* 32(3), pp. 323-339.

Al-Twairesh, N., Al-Matham, R., Madi, N., Almugren, N., Al-Aljmi, A.-H., Alshalan, S., Alshalan, R., Alrumayyan, N., Al-Manea, S. & Bawazeer, S. (2018). 'Suar: Towards building a corpus for the Saudi dialect', *Procedia computer science,* 142, pp. 72-82.

Alasmari, M. (2015). *Lenition in the Rijal Alhajir tribal dialect in the southern region of Saudi Arabia*: Southern Illinois University at Carbondale.

Alateeq, H. A. (2014). *Reported confidence of speech-language pathology graduates.* Texas A&M University-Kingsville

Albert, M. L., Goodglass, H., Helm, N. A., Rubens, A. B. & Alexander, M. P. (2013). *Clinical aspects of dysphasia* (Vol. 2): Springer Science & Business Media.

Albright, E. & Purves, B. (2008). 'Exploring SentenceShaperTM: Treatment and augmentative possibilities', *Aphasiology,* 22(7-8), pp. 741-752.

Albustanji, Y. M. (2009). *Agrammatism In Jordanian–Arabic Speakers.* The Ohio State University

Albustanji, Y. M., Milman, L. H., Fox, R. A. & Bourgeois, M. S. (2013). 'Agrammatism in Jordanian-Arabic speakers', *Clinical linguistics & phonetics,* 27(2), pp. 94-110.

AlGhannam, B. A., Alsuwaidi, M. & Almayyan, W. (2018). 'Perceived Usability Using Arabic System Usability Scale (A-SUS): Student Perspective of Smart PAAET App', *International Journal of Computer Science and Information Security (IJCSIS),* 16(7).

Alhawary, M. T. (2009). *Arabic second language acquisition of morphosyntax*: Yale University Press.

Alqahtani, B. A., Alenazi, A. M., Hoover, J. C., Alshehri, M. M., Alghamdi, M. S., Osailan, A. M. & Khunti, K. (2020). 'Incidence of stroke among Saudi population: a systematic review and meta-analysis', *Neurological Sciences*, pp. 1-6.

Altaib, M. K., Falouda, M. & Meteyard, L. (2020). 'From informal to formal: the preliminary psychometric evaluation of the short aphasia test for Gulf Arabic speakers (SATG)', *Aphasiology*, pp. 1-19.
Alyahya, R. S. & Druks, J. (2016). 'The adaptation of the object and action naming battery into Saudi Arabic', *Aphasiology,* 30(4), pp. 463-482.
Alyahya, R. S., Halai, A. D., Conroy, P. & Lambon Ralph, M. A. (2020). 'A unified model of post-stroke language deficits including discourse production and their neural correlates', *Brain,* 143(5), pp. 1541-1554.
Alyahya, R. S., Halai, A. D., Conroy, P. & Ralph, M. A. L. (2018a). 'The behavioural patterns and neural correlates of concrete and abstract verb processing in aphasia: A novel verb semantic battery', *NeuroImage: Clinical,* 17, pp. 811-825.
Alyahya, R. S., Halai, A. D., Conroy, P. & Ralph, M. A. L. (2018b). 'Noun and verb processing in aphasia: Behavioural profiles and neural correlates', *NeuroImage: Clinical,* 18, pp. 215-230.
Alzahrani, S. (2003). 'The aphasia diagnostic informal assessment', *Unpublished*.
Aoun, J. E., Benmamoun, E. & Choueiri, L. (2009). *The syntax of Arabic*: Cambridge University Press.
Armstrong, E. (2000). 'Aphasic discourse analysis: The story so far', *Aphasiology,* 14(9), pp. 875-892.
Ballard, K. J. & Thompson, C. K. (1999). 'Treatment and generalization of complex sentence production in agrammatism', *Journal of Speech, Language, and Hearing Research,* 42(3), pp. 690-707.
Bangor, A., Kortum, P. & Miller, J. (2008). 'The system usability scale (SUS): An empirical evaluation', *International Journal of Human-Computer Interaction,* 24(6), pp. 574-594.
Barde, L. H., Schwartz, M. F. & Boronat, C. B. (2006). 'Semantic weight and verb retrieval in aphasia', *Brain and Language,* 97(3), pp. 266-278.
Basso, A. (2005). 'How intensive/prolonged should an intensive/prolonged treatment be?', *Aphasiology,* 19(10-11), pp. 975-984.
Bastiaanse, R., Edwards, S., Mass, E. & Rispens, J. (2003). 'Assessing comprehension and production of verbs and sentences: The Verb and Sentence Test (VAST)', *Aphasiology,* 17(1), pp. 49-73.
Bastiaanse, R., Edwards, S. & Rispens, J. (2002). *Verb and sentence test (VAST)*: Thames Valley Test Company.
Bastiaanse, R. & Jonkers, R. (1998). 'Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia', *Aphasiology,* 12(11), pp. 951-969.
Bastiaanse, R., Maas, E., Rispens, J., Schaap, L. & Mellema, A. (2000). *Werkwoorden-en zinnentest (WEZT): handleiding*: Swets Test Publishers.
Bates, E., Saygın, A. P., Moineau, S., Marangolo, P. & Pizzamiglio, L. (2005). 'Analyzing aphasia data in a multidimensional symptom space', *Brain and Language,* 92(2), pp. 106-116.
Berman, R. A. & Bolozky, S. (1978). *Modern Hebrew structure*: University Pub. Projects.
Berndt, R. & Mitchum, C. (1997). 'An experimental treatment of sentence comprehension', *Approaches to the treatment of aphasia*, pp. 91-111.
Berndt, R. S. (2001). 'Sentence production', *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*, pp. 375-396.

Berndt, R. S., Haendiges, A. N., Mitchum, C. C. & Sandson, J. (1997). 'Verb retrieval in aphasia. 2. Relationship to sentence processing', *Brain and language,* 56(1), pp. 107-137.

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), pp. 2345-2357.

Beveridge, M. A. & Crerar, M. A. (2002). 'Remediation of asyntactic sentence comprehension using a multimedia microworld', *Brain and language,* 82(3), pp. 243-295.

Bhogal, S. K., Teasell, R. & Speechley, M. (2003). 'Intensity of aphasia therapy, impact on recovery', *Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews [Internet]*: Centre for Reviews and Dissemination (UK).

Biran, M. & Fisher, S. (2015). 'Structured treatment can improve predicate argument structure impairment', *Aphasiology,* 29(1), pp. 29-56.

Bird, H. & Franklin, S. (1996). 'Cinderella revisited: A comparison of fluent and non-fluent aphasic speech', *Journal of neurolinguistics,* 9(3), pp. 187-206.

Bock, J. K. (1986). 'Meaning, sound, and syntax: Lexical priming in sentence production', *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 12(4), p. 575.

Bock, K. & Levelt, W. (1994a). 'Grammatical encoding'.

Bock, K. & Levelt, W. (1994b). *Language production: Grammatical encoding*: Academic Press.

Boser, K. I., Weinrich, M. & McCall, D. (2000). 'Maintenance of oral production in agrammatic aphasia: Verb tense morphology training', *Neurorehabilitation and neural repair,* 14(2), pp. 105-118.

Boyle, M. (2017). Semantic treatments for word and sentence production deficits in aphasia. *In:* Seminars in speech and language, 2017. Thieme Medical Publishers. pp. 052-061.

Brooke, J. (1996). 'SUS: a “quick and dirty'usability', *Usability evaluation in industry*, p. 189.

Bruns, C., Varley, R., Zimmerer, V. C., Carragher, M., Brekelmans, G. & Beeke, S. (2019). '“I don’t know”: a usage-based approach to familiar collocations in non-fluent aphasia', *Aphasiology,* 33(2), pp. 140-162.

Buckingham Jr, H. W. & Kertesz, A. (1974). 'A linguistic analysis of fluent aphasia', *Brain and Language,* 1(1), pp. 43-61.

Butler, R. A., Lambon Ralph, M. A. & Woollams, A. M. (2014). 'Capturing multidimensionality in stroke aphasia: mapping principal behavioural components to neural structures', *Brain,* 137(12), pp. 3248-3266.

Butterworth, B. & Howard, D. (1987). 'Paragrammatisms', *Cognition,* 26(1), pp. 1-37.

Byng, S. (1988). 'Sentence processing deficits: Theory and therapy', *Cognitive Neuropsychology,* 5(6), pp. 629-676.

Byng, S., Nickels, L. & Black, M. (1994). 'Replicating therapy for mapping deficits in agrammatism: Remapping the deficit?', *Aphasiology,* 8(4), pp. 315-341.

Caplan, D. & Hanna, J. E. (1998). 'Sentence production by aphasic patients in a constrained task', *Brain and Language,* 63(2), pp. 184-218.

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), pp. 895-916.

Carragher, M., Sage, K. & Conroy, P. (2013). 'The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation', *Neuropsychological rehabilitation,* 23(6), pp. 846-887.

Carragher, M., Sage, K. & Conroy, P. (2015). 'Outcomes of treatment targeting syntax production in people with Broca's-type aphasia: evidence from psycholinguistic assessment tasks and everyday conversation', *International Journal of Language & Communication Disorders,* 50(3), pp. 322-336.

Carter, J. A., Lees, J. A., Murira, G. M., Gona, J., Neville, B. G. & Newton, C. R. (2005). 'Issues in the development of cross-cultural assessments of speech and language for children', *International Journal of Language & Communication Disorders,* 40(4), pp. 385-401.

Chang, F., Dell, G. S. & Bock, K. (2006). 'Becoming syntactic', *Psychological review,* 113(2), p. 234.

Cherney, L. R., Halper, A. S., Holland, A. L. & Cole, R. (2008a). 'Computerized script training for aphasia: Preliminary results', *American Journal of Speech-Language Pathology*.

Cherney, L. R., Kaye, R. C. & van Vuuren, S. (2014). 'Acquisition and maintenance of scripts in aphasia: A comparison of two cuing conditions', *American Journal of Speech-Language Pathology,* 23(2), pp. S343-S360.

Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T. & Schooling, T. (2008b). 'Evidence-based systematic review: Effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia'.

Cherney, L. R., Patterson, J. P. & Raymer, A. M. (2011). 'Intensity of aphasia therapy: Evidence and efficacy', *Current neurology and neuroscience reports,* 11(6), p. 560.

Cho-Reyes, S. & Thompson, C. K. (2012). 'Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS)', *Aphasiology,* 26(10), pp. 1250-1277.

Christiansen, J. A. (1995). 'Coherence violations and propositional usage in the narratives of fluent aphasics', *Brain and Language,* 51(2), pp. 291-317.

Clough, S. & Gordon, J. K. (2020). 'Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia', *Aphasiology,* 34(5), pp. 515-539.

Conroy, P., Sage, K. & Lambon Ralph, M. A. (2009). 'A comparison of word versus sentence cues as therapy for verb naming in aphasia', *Aphasiology,* 23(4), pp. 462-482.

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), pp. 1815-1827.

Crary, M. A., Wertz, R. T. & Deal, J. L. (1992). 'Classifying aphasias: Cluster analysis of western aphasia battery and boston diagnostic aphasia examination results', *Aphasiology,* 6(1), pp. 29-36.

Cruice, M., Worrall, L., Hickson, L. & Murison, R. (2003). 'Finding a focus for quality of life with aphasia: Social and emotional health, and psychological well-being', *Aphasiology,* 17(4), pp. 333-353.

Dancey, C. P. & Reidy, J. (2007). *Statistics without maths for psychology*: Pearson education.

Davidson, B., Worrall, L. & Hickson, L. (2003). 'Identifying the communication activities of older people with aphasia: Evidence from naturalistic observation', *Aphasiology,* 17(3), pp. 243-264.
De Bleser, R. & Papathanasiou, I. (2003). *The Sciences of Aphasia: From Therapy to Theory*: Pergamon.
Difrancesco, S., Pulvermüller, F. & Mohr, B. (2012). 'Intensive language-action therapy (ILAT): The methods', *Aphasiology,* 26(11), pp. 1317-1351.
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), pp. 254-267.
Dorze, G. L., Jacob, A. & Coderre, L. (1991). 'Aphasia rehabilitation with a case of agrammatism: A partial replication', *Aphasiology,* 5(1), pp. 63-85.
Doyle, P. J., McNeil, M. R., Bost, J. E., Ross, K. B., Wambaugh, J. L., Hula, W. D. & Mikolic, J. M. (2007). 'The Burden of Stroke Scale (BOSS) provided valid, reliable, and responsive score estimates of functioning and well-being during the first year of recovery from stroke', *Quality of Life Research,* 16(8), pp. 1389-1398.
Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S. & Laster, L. J. (1999). 'The stroke impact scale version 2.0: evaluation of reliability, validity, and sensitivity to change', *Stroke,* 30(10), pp. 2131-2140.
Edmonds, L. A. (2014). 'Tutorial for Verb Network Strengthening Treatment (VNeST): Detailed description of the treatment protocol with corresponding theoretical rationale', *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders,* 24(3), pp. 78-88.
Edmonds, L. A. (2016). 'A review of verb network strengthening treatment', *Topics in Language Disorders,* 36(2), pp. 123-135.
Edmonds, L. A. & Babb, M. (2011). 'Effect of verb network strengthening treatment in moderate-to-severe aphasia', *American Journal of Speech-Language Pathology*.
Edmonds, L. A., Mammino, K. & Ojeda, J. (2014). 'Effect of verb network strengthening treatment (VNeST) in persons with aphasia: Extension and replication of previous findings', *American Journal of Speech-Language Pathology,* 23(2), pp. S312-S329.
Edmonds, L. A. & Mizrahi, S. (2011). 'Online priming of agent and patient thematic roles and related verbs in younger and older adults', *Aphasiology,* 25(12), pp. 1488-1506.
Edmonds, L. A., Nadeau, S. E. & Kiran, S. (2009). 'Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia', *Aphasiology,* 23(3), pp. 402-424.
Edmonds, L. A., Obermeyer, J. & Kernan, B. (2015). 'Investigation of pretreatment sentence production impairments in individuals with aphasia: Towards understanding the linguistic variables that impact generalisation in Verb Network Strengthening Treatment', *Aphasiology,* 29(11), pp. 1312-1344.
Edwards, S. (1995). 'Profiling fluent aphasic spontaneous speech: A comparison of two methodologies', *International Journal of Language & Communication Disorders,* 30(3), pp. 333-345.
Edwards, S. (2005). *Fluent aphasia* (Vol. 107): Cambridge University Press.
Edwards, S. & Bastiaanse, R. (1998). 'Diversity in the lexical and syntactic abilities of fluent aphasic speakers', *Aphasiology,* 12(2), pp. 99-117.

Edwards, S. & Tucker, K. (2006). 'Verb retrieval in fluent aphasia: A clinical study', *Aphasiology,* 20(7), pp. 644-675.
El-Hajj, M., Salameh, P., Rachidi, S. & Hosseini, H. (2016). 'The epidemiology of stroke in the Middle East', *European Stroke Journal,* 1(3), pp. 180-198.
Enderby, P., John, A. & Petheram, B. (2013). *Therapy outcome measures for rehabilitation professionals: speech and language therapy, physiotherapy, occupational therapy*: John Wiley & Sons.
Faroqi-Shah, Y. (2008). 'A comparison of two theoretically driven treatments for verb inflection deficits in aphasia', *Neuropsychologia,* 46(13), pp. 3088-3100.
Faroqi-Shah, Y. & Baker, A. L. (2017). 'Agrammatic Aphasia', *Aphasia Rehabilitation: Clinical Challenges*, p. 101.
Ferguson, C. A. (1959). 'Diglossia', *word,* 15(2), pp. 325-340.
Ferretti, T. R., McRae, K. & Hatherell, A. (2001). 'Integrating verbs, situation schemas, and thematic role concepts', *Journal of Memory and Language,* 44(4), pp. 516-547.
Fink, R., Schwartz, M. & Myers, J. (1998). Investigations of the sentence-query approach to mapping therapy. *In:* Brain and language, 1998. ACADEMIC PRESS INC 525 B ST, STE 1900, SAN DIEGO, CA 92101-4495 USA. pp. 203-207.
Fink, R. B., Martin, N., Schwartz, M. F., Saffron, E. M. & Myers, J. L. (1993). 'Facilitation of verb retrieval skills in aphasia: A comparison of two approaches'.
Fink, R. B., Schwartz, M. F., Rochon, E., Myers, J. L., Socolof, G. S. & Bluestone, R. (1995). 'Syntax stimulation revisited: An analysis of generalization of treatment effects', *American Journal of Speech-Language Pathology,* 4(4), pp. 99-104.
Frattali, C. M., Thompson, C. M., Holland, A. L., Wohl, C. B. & Ferketic, M. M. (1995). 'The FACS of life ASHA facs--a functional outcome measure for adults', *Asha,* 37(4), pp. 40-46.
Frederiksen, J. R. & White, B. Y. (1989). 'An approach to training based upon principled task decomposition', *Acta psychologica,* 71(1-3), pp. 89-146.
Friedmann, N. (2006). 'Speech production in Broca’s agrammatic aphasia: Syntactic tree pruning', *Broca’s region*, pp. 63-82.
Friedmann, N. a. & Grodzinsky, Y. (1997). 'Tense and agreement in agrammatic production: Pruning the syntactic tree', *Brain and language,* 56(3), pp. 397-425.
Furnas, D. W. & Edmonds, L. A. (2014). 'The effect of computerised Verb Network Strengthening Treatment on lexical retrieval in aphasia', *Aphasiology,* 28(4), pp. 401-420.
Gamal-Eldin, S. M. (1967). 'A Syntactic Study of Egyptian Colloquial Arabic'.
Garrett, M. (1980). 'Levels of processing in sentence production', *Language production Vol. 1: Speech and talk*: Academic Presspp. 177-220.
Garrett, M. F. (1975). 'The analysis of sentence production', *Psychology of learning and motivation*: Elsevierpp. 133-177.
Garrett, M. F. (1976). 'Syntactic processes in sentence production', *New approaches to language mechanisms,* 30, pp. 231-256.
Garrett, M. F. (1982). 'Observations from normal and pathological language use', *Normality and pathology in cognitive functions*.
Garrett, M. F. (1988). Processes in language production. Linguistics: The Cambridge survey 3: Language: Psychological and biological aspects, ed. by Frederick J. Newmeyer, 69-96. Cambridge: Cambridge University Press.

Gleason, J. B., Goodglass, H., Obler, L., Green, E., Hyde, M. R. & Weintraub, S. (1980). 'Narrative strategies of aphasic and normal-speaking subjects', *Journal of Speech, Language, and Hearing Research,* 23(2), pp. 370-382.

Gonzalez, R., Rojas, M. & Ardila, A. (2020). 'Non-linguistic abilities in aphasia', *Journal of Neurolinguistics,* 56, p. 100916.

Goodglass, H. (1993). *Understanding aphasia*: Academic Press.

Goodglass, H. & Kaplan, E. (1972). *The assessment of aphasia and related disorders*: Lea & Febiger.

Goodglass, H. & Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*: Lea & Febiger.

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

Goral, M. & Kempler, D. (2009). 'Training verb production in communicative context: Evidence from a person with chronic non-fluent aphasia', *Aphasiology,* 23(12), pp. 1383-1397.

Gordon, J. K. (2007). 'A contextual approach to facilitating word retrieval in agrammatic aphasia', *Aphasiology,* 21(6-8), pp. 643-657.

Gordon, J. K. (2008). 'Measuring the lexical semantics of picture description in aphasia', *Aphasiology,* 22(7-8), pp. 839-852.

Gordon, J. K. & Dell, G. S. (2003). 'Learning to divide the labor: An account of deficits in light and heavy verb production', *Cognitive Science,* 27(1), pp. 1-40.

Grimshaw, J. (1990). *Argument structure*: the MIT Press.

Haarmann, H. J. & Kolk, H. H. (1991). 'Syntactic priming in Broca's aphasics: Evidence for slow activation', *Aphasiology,* 5(3), pp. 247-263.

Habash, N., Soudi, A. & Buckwalter, T. (2007). 'On arabic transliteration', *Arabic computational morphology*: Springerpp. 15-22.

Haendiges, A. N., Berndt, R. S. & Mitchum, C. C. (1996). 'Assessing the elements contributing to a" mapping" deficit: a targeted treatment study', *Brain and Language,* 52(1), pp. 276-302.

Harrat, S., Meftouh, K. & Smaili, K. (2019). 'Machine translation for Arabic dialects (survey)', *Information Processing & Management,* 56(2), pp. 262-273.

Harris, L., Olson, A. & Humphreys, G. (2012). 'Rehabilitation of past tense verb production and non-canonical sentence production in left inferior frontal non-fluent aphasia', *Aphasiology,* 26(2), pp. 143-161.

Hassanein, A., El-Tamawy, M., Sallam, T., Hosny, H., Abdel Naseer, M. & El-Fayoumy, N. (2002). 'Kasr El Aini Arabic Aphasia test (KAAT) simple, standardized, valid, reliable test for Egyptian patients, literate and illiterate', *Egypt J Neurol Psychiatry Neurosurg,* 39, pp. 381-95.

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*: J. Wiley; Chapman & Hall.

Heeschen, C. (1985). 'Agrammatism versus paragrammatism: A fictitious opposition', *Agrammatism*: Elsevierpp. 207-248.

Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A. & Andriacchi, K. D. (2008). 'Narrative transcription accuracy and reliability in two languages', *Topics in Language Disorders,* 28(2), pp. 178-188.

Helm-Estabrooks, N. & Ramsberger, G. (1986). 'Treatment of agrammatism in long-term Broca's aphasia', *British Journal of Disorders of Communication,* 21(1), pp. 39-45.
Herbert, R., Best, W., Hickin, J., Howard, D. & Osborne, F. (2003). 'Combining lexical and interactional approaches to therapy for word finding deficits in aphasia', *Aphasiology,* 17(12), pp. 1163-1186.
Hesketh, A. & Bishop, D. (1996). 'Agrammatism and adaptation theory', *Aphasiology,* 10(1), pp. 49-80.
Hilari, K., Byng, S., Lamping, D. L. & Smith, S. C. (2003). 'Stroke and aphasia quality of life scale-39 (SAQOL-39) evaluation of acceptability, reliability, and validity', *Stroke,* 34(8), pp. 1944-1950.
Hinckley, J. (2017). 'Selecting, Combining, and Bundling Different Therapy Approaches', *Aphasia Rehabilitation: Clinical Challenges: Clinical Challenges*, p. 331.
Hinckley, J. J. (2016). Aphasia practice in the year 2026. *In:* Seminars in speech and language, 2016. Thieme Medical Publishers. pp. 166-172.
Holland, A. L., Frattali, C. & Fromm, D. (1999). *Communication activities of daily living: CADL-2*: Pro-Ed.
Holland, A. L., Miller, J., Reinmuth, O., Bartlett, C., Fromm, D., Pashek, G., Stein, D. & Swindell, C. (1985). 'Rapid recovery from aphasia: A detailed language analysis', *Brain and Language,* 24(1), pp. 156-173.
Howard, D. & Patterson, K. (1992). 'Pyramid and palm trees: a test of semantic accessfrom pictures and words', *Bury Saint Edmunds: Thames Valley Test Company*.
Hussmann, K., Grande, M., Meffert, E., Christoph, S., Piefke, M., Willmes, K. & Huber, W. (2012). 'Computer-assisted analysis of spontaneous speech: quantification of basic parameters in aphasic and unimpaired language', *Clinical linguistics & phonetics,* 26(8), pp. 661-680.
Ibrahim, R., Eviatar, Z. & Aharon-Peretz, J. (2002). 'The characteristics of Arabic orthography slow its processing', *Neuropsychology,* 16(3), p. 322.
Ingham, B. (1994). 'Najdi Arabic', *Central Arabian, Amsterdam/Philadelphia, Benjamins, coll.“London Oriental and African Language Library*.
Jacobs, B. J. & Thompson, C. K. (2000). 'Cross-modal generalization effects of training noncanonical sentence comprehension and production in agrammatic aphasia', *Journal of Speech, Language, and Hearing Research,* 43(1), pp. 5-20.
Jefferies, E., Patterson, K., Jones, R. W. & Lambon Ralph, M. A. (2009). 'Comprehension of concrete and abstract words in semantic dementia', *Neuropsychology,* 23(4), p. 492.
Jespersen, O. (1992). *The philosophy of grammar*: University of Chicago Press.
Johnson, W., Onuma, O., Owolabi, M. & Sachdev, S. (2016). 'Stroke: a global response is needed', *Bulletin of the World Health Organization,* 94(9), p. 634.
Jones, E. V. (1986). 'Building the foundations for sentence production in a non-fluent aphasic', *British journal of disorders of communication,* 21(1), pp. 63-82.
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 and language,* 164, pp. 63-67.
Kay, J., Lesser, R. & Coltheart, M. (1996). 'Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction', *Aphasiology,* 10(2), pp. 159-180.

Kertesz, A. (1982). *Western aphasia battery test manual*: Psychological Corp.
Kertesz, A. (2007). *WAB-R: Western aphasia battery-revised*: PsychCorp.
Kertesz, A. & McCabe, P. (1975). 'Intelligence and aphasia: performance of aphasics on Raven's coloured progressive matrices (RCPM)', *Brain and Language,* 2, pp. 387-395.
Khoja, M. A. (2019). 'A survey of formal and informal assessment procedures used by speech-language pathologists in Saudi Arabia', *Speech, Language and Hearing,* 22(2), pp. 91-99.
Khoja, M. A. & Sheeshah, H. (2018). 'The human right to communicate: A survey of available services in Saudi Arabia', *International journal of speech-language pathology,* 20(1), pp. 102-107.
Khwaileh, T., Body, R. & Herbert, R. (2015). 'Morpho-syntactic processing of Arabic plurals after aphasia: dissecting lexical meaning from morpho-syntax within word boundaries', *Cognitive neuropsychology,* 32(6), pp. 340-367.
Khwaileh, T., Body, R. & Herbert, R. (2017). 'Lexical retrieval after Arabic aphasia: Syntactic access and predictors of spoken naming', *Journal of Neurolinguistics,* 42, pp. 140-155.
Khwaileh, T. & Grosvald, M. (2019). 'The Use of Translated Materials in Assessing CVA Patients in Qatar: Problems and Solutions', *Arab Journal of Applied Linguistics,* 4(1), p. 44.
Khwaileh, T., Mustafawi, E. & Albustanji, Y. (2020). 'A linguistically-driven response categorisation protocol for Arabic nouns and verbs: clinical and research applications', *Clinical Linguistics & Phonetics*, pp. 1-17.
Khwaileh, T. A., Mustafawi, E., Howard, D. & Herbert, R. (2016). An aphasia battery for Qatari/Gulf Arabic. *In:* Frontiers in Psychology 54th Annual Academy of Aphasia Meeting Conference Abstracts, 2016.
Kim, M. & Thompson, C. K. (2000). 'Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization', *Brain and language,* 74(1), pp. 1-25.
King, S. N., Davis, L., Lehman, J. J. & Ruddy, B. H. (2012). 'A model for treating voice disorders in school-age children within a video gaming environment', *Journal of Voice,* 26(5), pp. 656-663.
Kolk, H. (1995). 'A time-based approach to agrammatic production', *Brain and language,* 50(3), pp. 282-303.
Kolk, H. & Heeschen, C. (1990). 'Adaptation symptoms and impairment symptoms in Broca's aphasia', *Aphasiology,* 4(3), pp. 221-231.
Kolk, H. H., van Grunsven, M. J. & Keyser, A. (1985). 'On parallelism between production and comprehension in agrammatism', *Agrammatism*: Elsevierpp. 165-206.
Kolk, H. H. & Van Grunsven, M. M. (1985). 'Agrammatism as a variable phenomenon', *Cognitive neuropsychology,* 2(4), pp. 347-384.
Kwag, E. J., Sung, J. E., Kim, Y.-H., Cheond, H.-J., Kwag, E. J., Sung, J. E., Kim, Y.-H. & Cheond, H.-J. (2014). 'Effects of verb network strengthening treatment on retrieval of verbs and nouns in persons with aphasia', *Communication Sciences & Disorders,* 19(1), pp. 89-98.
L Tate, R., McDonald, S., Perdices, M., Togher, L., Schultz, R. & Savage, S. (2008). 'Rating the methodological quality of single-subject designs and n-of-1 trials: Introducing

the Single-Case Experimental Design (SCED) Scale', *Neuropsychological rehabilitation,* 18(4), pp. 385-401.
Laks, L. (2013). 'Passive formation in Palestinian and Standard Arabic: Lexical vs. syntactic operations', *Word Structure,* 6(2), pp. 156-180.
Lam, J. M. & 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', *Medical care*, pp. 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), pp. 289-305.
Lee, J. B., Kaye, R. C. & Cherney, L. R. (2009). 'Conversational script performance in adults with non-fluent aphasia: Treatment intensity and aphasia severity', *Aphasiology,* 23(7-8), pp. 885-897.
Lesser, R. (1989). 'Some issues in the neuropsychological rehabilitation of anomia'.
Levelt, W. (1989). 'From intention to articulation', *Cambridge, MA: The MITPress*.
Levelt, W. J. (1999). 'Models of word production', *Trends in cognitive sciences,* 3(6), pp. 223-232.
Linebarger, M. (1990). 'Neuropsychology of sentence parsing'.
Linebarger, M., McCall, D. & Berndt, R. (2004). 'The role of processing support in the remediation of aphasic language production disorders', *Cognitive Neuropsychology,* 21(2-3), pp. 267-282.
Linebarger, M., McCall, D., Virata, T. & Berndt, R. S. (2007). 'Widening the temporal window: Processing support in the treatment of aphasic language production', *Brain and Language,* 100(1), pp. 53-68.
Linebarger, M. & Romania, J. (2000). 'SentenceShaper®[Computer software, Unisys Corporation]', *Jenkintown, PA: Psycholinguistic Technologies*.
Linebarger, M., Schwartz, M. & Kohn, S. (2001). 'Computer-based training of language production: An exploratory study', *Neuropsychological Rehabilitation,* 11(1), pp. 57-96.
Linebarger, M. C. (1987). 'Negative polarity and grammatical representation', *Linguistics and philosophy*, pp. 325-387.
Linebarger, M. C., Romania, J. F., Fink, R. B., Bartlett, M. R. & Schwartz, M. F. (2008). 'Building on residual speech: A portable processing prosthesis for aphasia', *J Rehabil Res Dev,* 45(9), pp. 1401-1414.
Linebarger, M. C., Schwartz, M. F., Romania, J. R., Kohn, S. E. & Stephens, D. L. (2000). 'Grammatical encoding in aphasia: Evidence from a "processing prosthesis"', *Brain and Language,* 75(3), pp. 416-427.
Linebarger, M. C., Schwartz, M. F. & Saffran, E. M. (1983). 'Sensitivity to grammatical structure in so-called agrammatic aphasics', *Cognition,* 13(3), pp. 361-392.
Links, P., Hurkmans, J. & Bastiaanse, R. (2010). 'Training verb and sentence production in agrammatic Broca's aphasia', *Aphasiology,* 24(11), pp. 1303-1325.
Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A. & Zoghaib, C. (1989). 'The communicative effectiveness index: Development and psychometric evaluation of a functional communication measure for adult aphasia', *Journal of speech and hearing disorders,* 54(1), pp. 113-124.

Long, A., Hesketh, A., Paszek, G., Booth, M. & Bowen, A. (2008). 'Development of a reliable self-report outcome measure for pragmatic trials of communication therapy following stroke: the Communication Outcome after Stroke (COAST) scale', *Clinical Rehabilitation,* 22(12), pp. 1083-1094.

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A. & Pinna, G.-D. (2002). 'Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability', *Brain and language,* 81(1-3), pp. 432-444.

Mark, H., Fletcher, W. & Birt, D. (1983). *Storylines: Picture sequences for language practice*: Longman.

Marshall, J. (1995). 'The mapping hypothesis and aphasia therapy', *Aphasiology,* 9(6), pp. 517-539.

Marshall, J. (2009). 'Framing ideas in aphasia: The need for thinking therapy', *International journal of language & communication disorders,* 44(1), pp. 1-14.

Marshall, J. (2015). 'Assessment and treatment of sentence processing disorders', *The Handbook of Adult Language Disorders*: Psychology Presspp. 391-418.

Marshall, J. (2017). 'Disorders of sentence processing in Aphasia', *Aphasia and related neurogenic communication disorders*.

Marshall, J., Booth, T., Devane, N., Galliers, J., Greenwood, H., Hilari, K., Talbot, R., Wilson, S. & Woolf, C. (2016). 'Evaluating the benefits of aphasia intervention delivered in virtual reality: results of a quasi-randomised study', *PloS one,* 11(8), p. e0160381.

Marshall, J., Chiat, S. & Pring, T. (1997). 'An impairment in processing verbs' thematic roles: A therapy study', *Aphasiology,* 11(9), pp. 855-876.

Marshall, J., Devane, N., Edmonds, L., Talbot, R., Wilson, S., Woolf, C. & Zwart, N. (2018). 'Delivering word retrieval therapies for people with aphasia in a virtual communication environment', *Aphasiology,* 32(9), pp. 1054-1074.

Marshall, J., Pring, T. & Chiat, S. (1998). 'Verb retrieval and sentence production in aphasia', *Brain and Language,* 63(2), pp. 159-183.

Mätzig, S., Druks, J., Masterson, J. & Vigliocco, G. (2009). 'Noun and verb differences in picture naming: Past studies and new evidence', *Cortex,* 45(6), pp. 738-758.

McCall, D., Virata, T., Linebarger, M. C. & Berndt, R. S. (2009). 'Integrating technology and targeted treatment to improve narrative production in aphasia: A case study', *Aphasiology,* 23(4), pp. 438-461.

McCann, C. & Doleman, J. (2011). 'Verb retrieval in nonfluent aphasia: A replication of Edwards & Tucker, 2006', *Journal of Neurolinguistics,* 24(2), pp. 237-248.

McNeil, M. R., Robin, D. A. & Schmidt, R. A. (2009). 'Apraxia of speech: Definition and differential diagnosis', *Clinical management of sensorimotor speech disorders,* 2, pp. 249-267.

McRae, K., Hare, M., Elman, J. L. & Ferretti, T. (2005). 'A basis for generating expectancies for verbs from nouns', *Memory & cognition,* 33(7), pp. 1174-1184.

Menn, L., Ramsberger, G. & Estabrooks, N. H. (1994). 'A linguistic communication measure for aphasic narratives', *Aphasiology,* 8(4), pp. 343-359.

Miller, J. & Chapman, R. (1983). SALT: System analysis of language transcriptions. Baltimore: University Park Press.

Miller, N., Willmes, K. & De Bleser, R. (2000). 'The psychometric properties of the English language version of the Aachen Aphasia Test (EAAT)', *Aphasiology,* 14(7), pp. 683-722.

Milman, L., Vega-Mendoza, M. & Clendenen, D. (2014). 'Integrated training for aphasia: An application of part–whole learning to treat lexical retrieval, sentence production, and discourse-level communications in three cases of nonfluent aphasia', *American Journal of Speech-Language Pathology,* 23(2), pp. 105-119.
Mimouni, Z. & Jarema, G. (1997). 'Agrammatic aphasia in Arabic', *Aphasiology,* 11(2), pp. 125-144.
Mimouni, Z., Kehayia, E. & Jarema, G. (1998). 'The mental representation of singular and plural nouns in Algerian Arabic as revealed through auditory priming in agrammatic aphasic patients', *Brain and language,* 61(1), pp. 63-87.
Mitchum, C. & Berndt, R. (1994). 'Verb retrieval and sentence construction: Effects of targeted intervention', *Cognitive neuropsychology and cognitive rehabilitation*, pp. 317-348.
Mitchum, C., Greenwald, M. & Berndt, R. (1997a). Production-specific thematic mapping impairment: A treatment study. *In:* Brain and Language, 1997a. ACADEMIC PRESS INC ELSEVIER SCIENCE 525 B ST, STE 1900, SAN DIEGO, CA 92101 …. pp. 121-123.
Mitchum, C., Greenwald, M. L. & Berndt, R. S. (1997b). Production-specific thematic mapping impairment: A treatment study. *In:* Brain and Language, 1997b. ACADEMIC PRESS INC ELSEVIER SCIENCE 525 B ST, STE 1900, SAN DIEGO, CA 92101 …. pp. 121-123.
Mitchum, C. C., Greenwald, M. L. & Berndt, R. S. (2000). 'Cognitive treatments of sentence processing disorders: What have we learned?', *Neuropsychological Rehabilitation,* 10(3), pp. 311-336.
Mitchum, C. C., Haendiges, A. N. & Berndt, R. S. (1995). 'Treatment of thematic mapping in sentence comprehension: Implications for normal processing', *Cognitive Neuropsychology,* 12(5), pp. 503-547.
Murray, L. & Coppens, P. (2013). 'Formal and informal assessment of aphasia', *Aphasia and related neurogenic communication disorders*, pp. 66-91.
National Aphasia Association (2021). *Aphasia Statistics*. Available at: https://www.aphasia.org/aphasia-resources/aphasia-statistics/.
Nicholas, L. E. & Brookshire, R. H. (1993). 'A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia', *Journal of Speech, Language, and Hearing Research,* 36(2), pp. 338-350.
Nicholas, M., Obler, L. K., Albert, M. L. & Helm-Estabrooks, N. (1985). 'Empty speech in Alzheimer's disease and fluent aphasia', *Journal of Speech, Language, and Hearing Research,* 28(3), pp. 405-410.
Nickels, L., Byng, S. & Black, M. (1991). 'Sentence processing deficits: A replication of therapy', *International Journal of Language & Communication Disorders,* 26(2), pp. 175-199.
Othman, E., Shaalan, K. & Rafea, A. (2003). A chart parser for analyzing modern standard Arabic sentence. *In:* Proceedings of the MT summit IX workshop on machine translation for semitic languages: issues and approaches, 2003. pp. 37-44.
Paivio, A., Yuille, J. C. & Madigan, S. A. (1968). 'Concreteness, imagery, and meaningfulness values for 925 nouns', *Journal of experimental psychology,* 76(1p2), p. 1.
Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., Latimer, N., Julious, S., Cross, E. & Alshreef, A. (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), pp. 821-833.
Paradis, M. (1987). *Bilingual aphasia test*: Lawrence Erlbaum Associates Hillsdale, NJ, USA:.
Paradis, M., Hallis, Y., Taha, K. & Amayreh, M. (n.d.). Bilingual aphasia test (Jordanian Arabic version).
Parisi, D. (2013). 'Grammatical disturbances of speech production', *The cognitive neuropsychology of language*, pp. 201-219.
Parkes, B. P. (2017). *Is Verb Network Strengthening Treatment (VNeST) Effective at a Lower Dosage?* , Temple University. Libraries
Poirier, S.-È., Fossard, M. & Monetta, L. (2021). 'The efficacy of treatments for sentence production deficits in aphasia: a systematic review', *Aphasiology*, pp. 1-21.
Procházka, S. (2006). 'Does geographical periphery imply linguistic periphery? The examples of the Arabic dialects of Cilicia and Urfa in Southern Turkey', *Peripheral Arabic Dialects*.
Psycholinguistic Technologies (2015). 'User's Manual SentenceShaper® 2, Version 1.0.10'.
Raven, J. C., Raven, J. C. & Court, J. H. (1962). *Advanced progressive matrices*: HK Lewis London.
Raymer, A. & Ellsworth, T. (2002). 'Contrasting treatments for verb retrieval impairment in aphasia: a case study', *Aphasiology,* 16, pp. 1031-1045.
Robey, R. R. (1998). 'A meta-analysis of clinical outcomes in the treatment of aphasia', *Journal of Speech, Language, and Hearing Research,* 41(1), pp. 172-187.
Robson, J., MARSHALL, J., Pring, T. & CHIAT, S. (1998). 'Phonological naming therapy in jargon aphasia: Positive but paradoxical effects', *Journal of the International Neuropsychological Society,* 4(6), pp. 675-686.
Rochon, E., Laird, L., Bose, A. & Scofield, J. (2005). 'Mapping therapy for sentence production impairments in nonfluent aphasia', *Neuropsychological Rehabilitation,* 15(1), pp. 1-36.
Rochon, E., Saffran, E. M., Berndt, R. S. & Schwartz, M. F. (2000). 'Quantitative analysis of aphasic sentence production: Further development and new data', *Brain and language,* 72(3), pp. 193-218.
Ryding, K. C. (2014). *Arabic: A linguistic introduction*: Cambridge University Press.
Saffran, E. M. (1982). 'Neuropsychological approaches to the study of language', *British Journal of Psychology,* 73(3), pp. 317-337.
Saffran, E. M., Berndt, R. S. & Schwartz, M. F. (1989). 'The quantitative analysis of agrammatic production: Procedure and data', *Brain and language,* 37(3), pp. 440-479.
Saffran, E. M. & Schwartz, M. F. (1988). ''Agrammatic'comprehension it's not: Alternatives and implications', *Aphasiology,* 2(3-4), pp. 389-393.
Saffran, E. M., Schwartz, M. F. & Marin, O. S. (1980). 'Evidence from aphasia: Isolating the components of a production model', *Language production,* 1, pp. 221-241.
Salt Software LLC. (2015a). *Analysis set vs total utterances*. Available at: https://www.saltsoftware.com/faq/analysis-set-vs-total-utterances.
Salt Software LLC. (2015b). *Analysis set vs. total utterances*. Available at: https://www.saltsoftware.com/faq/analysis-set-vs-total-utterances/#comment-50.

Salt Software LLC. (2020). *Transcription aids*. Available at: https://www.saltsoftware.com/resources/tranaids

Schwartz, M. F., Linebarger, M. C. & Saffran, E. M. (1985). 'The status of the syntactic deficit theory of agrammatism', *Agrammatism*: Elsevierpp. 83-124.

Schwartz, M. F., Linebarger, M. C., Saffran, E. M. & Pate, D. S. (1987). 'Syntactic transparency and sentence interpretation in aphasia', *Language and Cognitive Processes,* 2(2), pp. 85-113.

Schwartz, M. F., Saffran, E. M., Fink, R. B., Myers, J. L. & Martin, N. (1994). 'Mapping therapy: A treatment programme for agrammatism', *Aphasiology,* 8(1), pp. 19-54.

Shaalan, S. (2009). 'Considerations for developing and adapting language and literacy assessments in Arabic-speaking countries', *Multicultural psychoeducational assessment*, pp. 287-314.

Shatz, C. J. (1992). 'The developing brain', *Scientific American,* 267(3), pp. 60-67.

Shewan, C. M. & Kertesz, A. (1980). 'Reliability and validity characteristics of the Western Aphasia Battery (WAB)', *Journal of Speech and Hearing Disorders,* 45(3), pp. 308-324.

Simons, G. F. & Fennig, C. D. (2018). 'Ethnologue: Languages of the world, twenty', *Dallas, Texas: SIL International. Online version: http://www. ethnologue. com Accessed,* 26(12), p. 2018.

Spreen, O. & Benton, A. (1969). 'Neurosensory center comprehensive examination for Aphasia: manual of instructions (NCCEA)', *Victoria, BC: University of Victoria*.

Springer, L., Glindemann, R., Huber, W. & Willmes, K. (1991). 'How efficacious is pace-therapy when 'language systematic training'is incorporated?', *Aphasiology,* (4-5), pp. 391-399.

Stroke Association (2021). *Stroke Statistics*. Available at: https://www.stroke.org.uk/what-is-stroke/stroke-statistics.

Swinburn, K., Porter, G. & Howard, D. (2004). *Comprehensive aphasia test*: Taylor & Francis.

Tactus Therapy Solutions Ltd. (2021). *Advanced Naming Therapy,*. Available at: https://apps.apple.com/app/advanced-naming-therapy/id1078772848?mt=8.

the Saudi general authority for statistics (2019). *annual report, population by age groups and gender* Available at: https://www.stats.gov.sa/sites/default/files/population_by_age_groups_and_gender_en.pdf (Accessed: 08/02/2021).

Thompson, C., Ballard, K., Shapiro, L. & Tait, M. (1997a). Training complex Wh-movement structures in agrammatic aphasia: Optimal order for promoting generalization. *In:* BRAIN AND LANGUAGE, 1997a. ACADEMIC PRESS INC JNL-COMP SUBSCRIPTIONS 525 B ST, STE 1900, SAN DIEGO, CA …. pp. 124-127.

Thompson, C., Ballard, K. J., Shapiro, L. P. & Tait, M. E. (1997b). Training complex Wh-movement structures in agrammatic aphasia: Optimal order for promoting generalization. *In:* BRAIN AND LANGUAGE, 1997b. ACADEMIC PRESS INC JNL-COMP SUBSCRIPTIONS 525 B ST, STE 1900, SAN DIEGO, CA …. pp. 124-127.

Thompson, C. & Shapiro, L. (2005). 'Treating agrammatic aphasia within a linguistic framework: Treatment of underlying forms', *Aphasiology,* 19(10-11), pp. 1021-1036.

Thompson, C. K. (2008). 'The state of impairment-and consequences-based approaches to treatment for aphasia: Final commentary: The impairment and its consequences', *Aphasia rehabilitation: The impairment and its consequences*: Plural Publishing, Incpp. 261-269.

Thompson, C. K. (2012). 'Treatment of syntactic and morphologic deficits in agrammatic aphasia: Treatment of underlying forms', *Language Intervention Strategies in Aphasia and Related Neurogenic Communication Disorders: Fifth Edition*: Wolters Kluwer Health Adis (ESP)pp. 735-755.

Thompson, C. K., Choy, J. J., Holland, A. & Cole, R. (2010). 'Sentactics®: Computer-automated treatment of underlying forms', *Aphasiology,* 24(10), pp. 1242-1266.

Thompson, C. K., Faroqi-Shah, Y. & Lee, J. (2015). 'Models of sentence production'.

Thompson, C. K. & Shapiro, L. P. (1995). 'Training sentence production in agrammatism: Implications for normal and disordered language', *Brain and Language,* 50(2), pp. 201-224.

Thompson, C. K., Shapiro, L. P., Kiran, S. & Sobecks, J. (2003). 'The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia', *Journal of Speech, Language, and Hearing Research*.

Thompson, C. K., Shapiro, L. P. & Roberts, M. M. (1993). 'Treatment of sentence production deficits in aphasia: A linguistic-specific approach to wh-interrogative training and generalization', *Aphasiology,* 7(1), pp. 111-133.

Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J. & Schneider, S. L. (1996). 'TrainingWh-Question Production in Agrammatic Aphasia: Analysis of Argument and Adjunct Movement', *Brain and Language,* 52(1), pp. 175-228.

Thorne, J. & Faroqi-Shah, Y. (2016). Verb production in aphasia: Testing the division of labor between syntax and semantics. *In:* Seminars in speech and language, 2016. Thieme Medical Publishers. pp. 023-033.

Tombaugh, T. N. & Hubiey, A. M. (1997). 'The 60-item Boston Naming Test: Norms for cognitively intact adults aged 25 to 88 years', *Journal of Clinical and Experimental Neuropsychology,* 19(6), pp. 922-932.

True, G., Bartlett, M. R., Fink, R. B., Linebarger, M. C. & Schwartz, M. (2010). 'Perspectives of persons with aphasia towards SentenceShaper To Go: A qualitative study', *Aphasiology,* 24(9), pp. 1032-1050.

Umanski, D., Kosters, W., Verbeek, F. & Schiller, N. O. (2008). Integrating computer games in speech therapy for children who stutter. *In:* First Workshop on Child, Computer and Interaction, 2008.

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), pp. 424-435.

Van Gorp, W. G., Satz, P., Kiersch, M. E. & Henry, R. (1986). 'Normative data on the boston maming test for a group of normal older adults', *Journal of Clinical and Experimental Neuropsychology,* 8(6), pp. 702-705.

Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*: Harvard university press.

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M. & Cherney, L. R. (2019). 'A core outcome set for aphasia treatment research: The ROMA consensus statement', *International Journal of Stroke,* 14(2), pp. 180-185.

Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Kirke, E. & Kolomeitz, D. (2018a). 'Report from ROMA: an update on the development of a core outcome set for aphasia research', *Aphasiology,* 32(sup1), pp. 241-242.

Wallace, S. J., Worrall, L. E., Rose, T. & Le Dorze, G. (2018b). 'Discourse measurement in aphasia research: have we reached the tipping point? A core outcome set… or greater standardisation of discourse measures?', *Aphasiology,* 32(4), pp. 479-482.

Wambaugh, J. L., Doyle, P. J., Martinez, A. L. & Kalinyak-Fliszar, M. (2002). 'Effects of two lexical retrieval cueing treatments on action naming in aphasia', *Journal of Rehabilitation Research and Development,* 39(4), pp. 455-466.

Webster, J., Morris, J. & Franklin, S. (2005). 'Effects of therapy targeted at verb retrieval and the realisation of the predicate argument structure: A case study', *Aphasiology,* 19(8), pp. 748-764.

Webster, J. & Whitworth, A. (2012). 'Treating verbs in aphasia: exploring the impact of therapy at the single word and sentence levels', *International Journal of Language & Communication Disorders,* 47(6), pp. 619-636.

Webster, J., Whitworth, A. & Morris, J. (2015). 'Is it time to stop “fishing”? A review of generalisation following aphasia intervention', *Aphasiology,* 29(11), pp. 1240-1264.

Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*: Psychological Corporation.

Weinrich, M., Shelton, J. R., Cox, D. M. & McCall, D. (1997). 'Remediating production of tense morphology improves verb retrieval in chronic aphasia', *Brain and Language,* 58(1), pp. 23-45.

Whitworth, A., Leitao, S., Cartwright, J., Webster, J., Hankey, G., Zach, J., Howard, D. & Wolz, V. (2015). 'NARNIA: A new twist to an old tale. A pilot RCT to evaluate a multilevel approach to improving discourse in aphasia', *Aphasiology,* 29(11), pp. 1345-1382.

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

Wilshire, C. E. & McCarthy, R. A. (2002). 'Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia', *Cognitive neuropsychology,* 19(2), pp. 165-186.

Zughoul, M. R. M. (2007). *Studies in Contemporary Arabic/English Sociolinguistcs*: Hamada Establishment for University Studies and Publ. and Distributing.

APPENDICES

Appendix 1.A Modifications to VAST test battery scoring guidelines

I. Fill-in verbs in sentences subtests (two subtests with total of 20 trials)

The assessment of verb retrieval in a sentence was prioritized over the accuracy of the verb form. Therefore, a modification was made to the scoring approach for the fill-in finite verbs and infinitives in a sentence task. The production of the target verb in any form was considered correct. For example, if the participant produced "diving" or "is diving" instead of the target "dives" in "The girl _____into the water" the answer was considered correct. Also, if the participant produced "reading" or "reads" instead of the target "read" in "The girl like to _____ a book" the answer was accepted. This modification allowed us to increase the number of trials from ten to twenty trials.

II. Sentence production subtest (20 trials)

A novel scoring protocol for VAST sentence production subtest was developed for the current study. The following table illustrates the guideline to scoring the three different types of sentences that appears in the sentence production subtest of the VAST.

	Sentence Type	Example	Maximum score	Instructions
a.	Noun Verb NV Target	*'The man is walking'*	7 points	
				NOUN AGENT 2 points for target Noun Agent (man) or synonym (e.g. chap/guy) 1 point for pronoun (he) or less informative noun (person, Tom, that one) 0 points for missing noun or non-informative or incorrect one (e.g. woman/she/that)

				<u>VERB</u> <u>2 points</u> for target Verb (walking) or synonym (e.g. strolling) <u>1 point</u> for less informative verb (e.g. moving) or light verb (going) <u>0 points</u> for missing verb or non-informative or incorrect one (e.g. falling, doing) <u>GRAMMAR</u> <u>3 points</u> for correct grammar: 1. Definite or indefinite marking on noun. 2. Tense morphology on verb. 3. Function words. <u>2 points</u> 2 of the above <u>1 point</u> for 1 of the above <u>0 points</u> is neither of the above
b.	Noun Verb Noun NVN Target	*"The boy is catching the ball"*	9 points	
				<u>NOUN AGENT</u> <u>2 points</u> for target Noun Agent (boy) or synonym (e.g. lad) <u>1 point</u> for pronoun (he) or less informative noun (person, Tom, that one) <u>0 points</u> for missing noun or non-informative or incorrect one (e.g. woman/she/that) <u>VERB</u> <u>2 points</u> for target Verb (catching) or synonym (e.g. grabbing) <u>1 point</u> for less informative verb (e.g. playing) or light verb (going) <u>0 points</u> for missing verb or non-informative or incorrect one (e.g. falling, doing) <u>NOUN PATIENT OR THEME</u> <u>2 points</u> for target Noun (ball) or synonym (e.g. lad) <u>1 point</u> for pronoun (it) or less informative noun (that one) <u>0 points</u> for missing noun or non-informative or incorrect one <u>GRAMMAR</u> <u>3 points</u> for correct grammar: 1. At least 1 definite <u>or</u> indefinite marking on noun. 2. Tense morphology on verb. 3. Function words. <u>2 points</u> 2 of the above <u>1 point</u> for 1 of the above <u>0 points</u> is neither of the above

c.	Noun Verb Noun Noun NVNN target	*"The man is drinking a glass of wine"*	9 points	
				NOUN AGENT 2 as above VERB 2 as above First and second NOUN PATIENT OR THEME 2 points 2 of the target nouns 1 point for 1 of the target nouns 0 points is neither of the target nouns GRAMMAR 3 as above

Appendix 1.B Scoring guideline for Arabic sentence production in VAST sentence construction task

NOUN AGENT الفاعل

2 points for target Noun Agent (boy) or synonym (e.g. lad)

1 point for pronoun (he) or less informative noun (person, Tom, that one)

0 points for missing noun or non-informative or incorrect one (e.g. woman/she/that)

VERB الفعل

2 points for target Verb (catching) or synonym (e.g. grabbing)

1 point for less informative verb (e.g. playing) or light verb (going)

0 points for missing verb or non-informative or incorrect one (e.g. falling, doing)

NOUN PATIENT OR THEME المفعول به

2 points for target Noun (ball) or synonym (e.g. lad)

1 point for pronoun (it) or less informative noun (that one)

0 points for missing noun or non-informative or incorrect one

GRAMMAR

3 points for correct grammar: 1. At least 1 definite or indefinite marking on noun. 2. Tense morphology on verb. 3. Verb gender marker.

2 points 2 of the above

1 point for 1 of the above

0 points is neither of the above

In creating the Arabic version of the scoring rubric the only modification to the original rubric was to the grammar scores. It involved replacing the requirement of producing a function word in a sentence with the requirement of producing a correct verb gender marker.

The structured sentence production task in the VAST usually elicits simple sentences rather than compound, complex sentences, or discourse. Therefore, it was determined that verb gender markers in Arabic are more common in simple sentences and errors in producing them affects its grammaticality (e.g., البنت تسقي الزرع vs البنت يسقي الزرع) .On the other hand, the article “the” which is also common in simple sentences in Arabic ال التعريف will be accounted for through the first requirement “definite or indefinite marking on noun”.

Table: example of a scored sentence (participants ND) in VAST sentence construction task:

errors	grammar	الورود	تسقي	البنت	Target sentence البنت تسقي الورود	Testing time-point
	0	2	0	2	بنت.. زرع.. وردة	Baseline 1
	0	0	0	0	زراعه	Baseline 2
verb gender marker	2	2	2	2	البنت .. ماي .. يسقي الورد	Interim1
verb gender marker	2	2	2	2	البنت يسقي الورد	Interim2

Adding several other sentence constituents to the scoring guideline would have been beneficial in capturing a wide range of sentence types and discourse. For example, conjunctions (و), prepositions (على الطاولة /للمطعم), separate and attached pronouns (هي خلته يطلع اللي),and singular/dual/plural markers (جَامِعَات/جَامِعَتَان/جَامِعَة). However, the current guideline is limited to the type of sentences typically produced by neuro-typical native Arabic speakers in response to the VAST sentence construction task. The target production for this structured sentence production task is always a simple sentence, which includes: nouns, verbs, noun markers such as definite/indefinite (e.g. articles البَيْت/بَيْت) markers, verb gender markers (ترسم/يرسم), and verb-subject agreement. Therefore, the current scoring rubric captures the participants’ success or failure in meeting the minimum required sentence constituents in a given task item as produced by controls. It doesn’t include variations that could occur due to elaboration or sentence expansion.

Appendix 1.C Arabic discourse scoring guideline

The table below presents an example of the method used to quantify the following discourse elements: Total word count, verbs, nouns, adjectives, prepositions, noun phrases (fragmented sentences), simple sentences, compound sentences, and complex sentences in a given speech sample. In addition to excluding incomplete and agrammatical sentences from the sentence count, complete and grammatically correct sentences that are not relevant to the stimulus were also excluded. However, the words were still calculated in the total word count.

Code	Indication
Strikethrough	mazes, false starts, repetitions, and filler words
Green font colour	complete informative and grammatically correct sentences
Three dots …	pauses of more than 3 seconds between productions
WC word count	verbs, nouns, adjectives, adverbs, pronouns (detached and overt only), prepositions, and conjunctions
simple sentence	A sentence that contains at least Subject + Verb للفعل اللازم Or Subject + Verb + Object للفعل المتعدي
compound sentence	two independent clauses connected with a coordindate conjunction (but, and, so)
complex sentence	main clause and a subordinate clause connected with a subordinate connective (if , when, while, because)

Table: samples of a scored discourse in Arabic

	word count	Sentences			Discourse (ND_CT_Int1)
		complex	compound	simple	
					~~البنت تشيل الكيكه .. والولد .. لا ..~~
	5			✓	البنت تمسك الكيك..، الولد .. ~~يرمي .. لا لا~~ .. يطيح .. ~~الولد~~ ..
	3			✓	الحرمه تغسل مواعين..
	2				يكب الماء ..
Total	10			2	verbs= 4, nouns= 6, and adjectives= 0

*Participant ND's cookie theft discourse sample at interim 1 time-point

word count	Sentences			Discourse (SA_DP_Int2)
	complex	compoun d	simple	
3			✓	محمد كلّم فيصل..
3				~~قال له وش.. يعني~~.. قال على العشاء..
6			✓	~~يعني~~ محمد قال : فيصل تجينا على العشاء?...
4			✓	قال إن شاء الله...
3			✓	بعدين راح البيت...
11		✓ (but)		وزوجته تطبخ ويساعدها في المطبخ... سمك ... لكنه السمكه نسيها …
6			✓	و هم يلبسون هو و زوجته ...

	10	✓ (if)			~~السفره~~...بعد ذلك ..~~اللبس~~ راحوا يشوفون السفره ، هل هي ناقصه ؟ فيها شي؟...
	2			✓	سمع الجرس...
	4				الضيوف فيصل و زوجته..
	6			✓	و هم جاهزين ..السفره.. نسوا السمكه..
	10	✓ (because)			بعد ذلك يعني .. محمد مسرعا إلى المطعم .. لكي لا يشوفون السمكه ... وخلاص يعني
Total	68	3	1	7	verbs= 15, nouns= 22, and adjectives= 1 (ناقصه)

*Participant SA's dinner party discourse sample at interim 2 time-point

Appendix 2 Light Verb Elicitation Test LVET (Carragher et al., 2013)

Instructions: the examiner reads aloud each sentence/excerpt from the accompanying stimuli document, using an obvious pause to indicate the gap. Then, examiner asks the person being tested to read the sentence/excerpt silently and suggests a word to go into the gap so that the sentence will make sense. Correct response scores 1.

Stimuli

1. When Sam heard the tickets were half-priced, he decided he would _____ to the concert after all.
2. Aisling just didn't know what to _____ about her problem.
3. You don't always ______ what you want in life.
4. Rachel was horrified when the cash-point told her she did not _____ enough money in her account
5. Let's not ______ hasty. We need to think this through.
6. Will you _____ and see me when I move to London?
7. I wouldn't ______ it a second thought.
8. Karen couldn't cope with the new litter of puppies. So she asked Tony to ______ them off her hands.
9. His Dad told Freddy that he should only _____ a promise if he knew he could keep it.
10. I didn't get the job. They don't think I _____ what it takes.
11. The neighbour was so annoyed about the ball going into his garden, that he refused to _______ it back to Jack.
12. I must try to ______ kinder to my sister.
13. Raymond is a very neat and tidy child. He will always ______ his bed before he comes down for breakfast.

14. Sheila was desperate for a cup of tea so went straight into the kitchen to ______ the kettle on.
15. I must ______ you out for Sunday Lunch sometime.
16. You need to be more independent. I can't ______ everything for you.
17. We are all off on holiday. Why don't you _____ with us.
18. It was a cold day. Dave reminded his son Michael to _______ some gloves on for the walk to school.
19. Next year, I will ______ on holiday no matter what.
20. There are easier ways to _____ money, but Frank loves his job.
21. The teachers need to keep an eye on Tom as he can often ____ advantage of younger children.
22. He doesn't seem to _____ a damn about it anymore.
23. Naz remembered she had no food at home and she needed to _____ to the supermarket.
24. The best advice in a fire is to _______ everyone out and then call the fire brigade.
25. I would love to ______ a fly on the wall when they discuss that.
26. Sophie didn't want to _____ children before her thirties.
27. Why don't we stay here and they can ______ to us
28. The best way to cope with exam pressure is to tell yourself that all you can _______ is your best.
29. Pat saw Jackie buying a drink but told her to ______ her money away, as this round was on him.
30. He'll _______ what's coming to him in the long run

Appendix 3 SentenceShaper procedure

SentenceShaper was designed to support individuals with aphasia to overcome working memory limitations that interfere with their language production. The tasks within the built-in therapy workbooks involve describing pictures, picture sequences, or answering questions by recording the user's productions and arranging the recordings, which will turn into symbols, on the screen.

Each word is recorded by pressing on the microphone icon displayed on the computer's screen (bottom); by doing so, a recording box will appear on the screen that depicts the recording's progress. After producing the intended message, the participant will press the "Done" button to stop the recording, and the box will close. A crystal ball icon will then appear on the screen, which resembles the recorded word/speech fragment, and clicking on it will replay the recording. This crystal ball icon and the recording attached to it is called a snippet. Next, the same steps will be followed to record another word, creating another snippet on the screen page. Each snippet on the sentence row will have a different symbol. They are arbitrary symbols designed to help the user visually differentiate between the recorded speech fragments (Psycholinguistic Technologies, 2015).

Then, the participant will be instructed to construct a sentence by dragging the icons to the allocated slots on the sentence row in the centre of the screen. Pressing the "play" button, located at the top left of the sentence row, will play all the snippets in sequence from left to right. For editing, the participant can rearrange the word order by dragging the icons into the target slot in the sentence row. The replay feature allows the participant to self-monitor and correct any errors at the sentence level (e.g., word order, morphology etc.). It also relieves the working memory from holding on to the produced sentence elements while planning the next. As a result, it supports planning larger structures, such as discourse and narratives. Also, it enables the revision of a sentence without having to produce all the words again.

Also, vocabulary cues can be provided on the main page (i.e., the side buttons on the screen) to assist the participant in sentence production.

Finally, the recorded sentences can be combined with the arrow button ↑ on the upper right corner of the sentence row box to form a narrative. It will save the entire sentence in a "purple bean" icon and move it to the upper part of the screen in the story row. Similarly, pressing the “play” button will play all recorded sentences on the story row, in order from left to right. Also, dragging the purple bean icon (i.e., resembles a single recorded sentence) back to the sentence row will decompose it into snippets. Therefore, narratives can still be edited at the sentence level.

<u>The therapy procedure of our novel method using SentenceShaper</u>

In the two therapy studies presented in the thesis (Chapter 4 and Chapter 6), SentenceShaper® (Linebarger and Romania, 2000) (cognitive processing support) was the computer medium for delivering the theory-driven language therapy tasks. In both forms, the computer software and the IOS application include features that allow the user to record, playback, and edit their sentences until a satisfactory production is achieved.

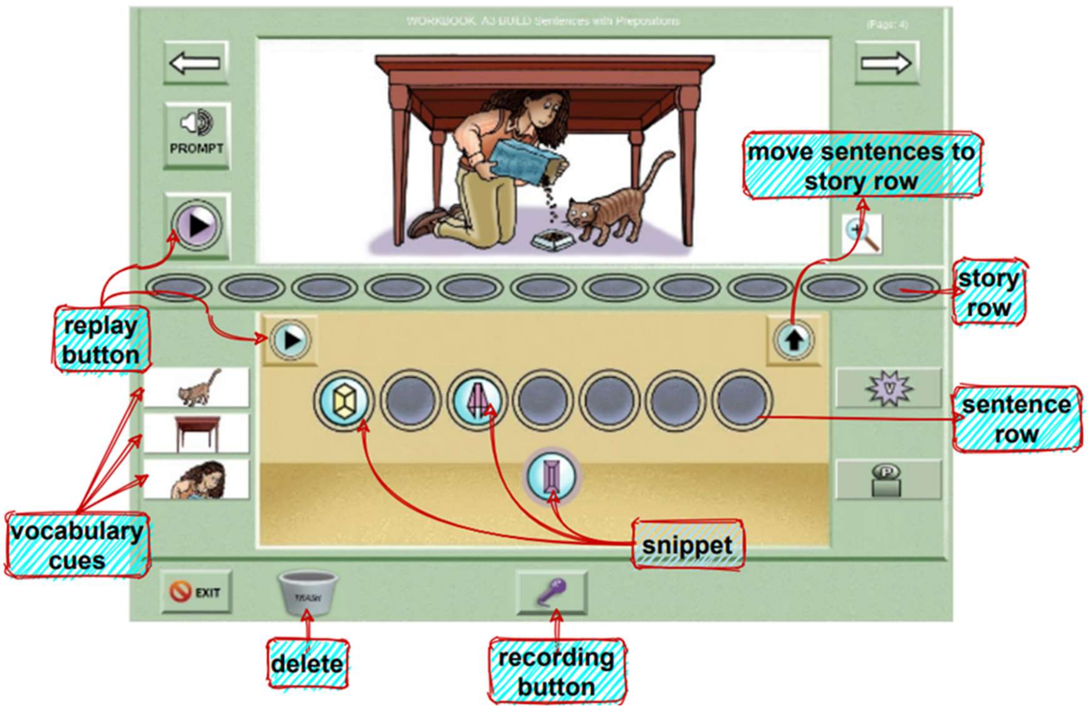
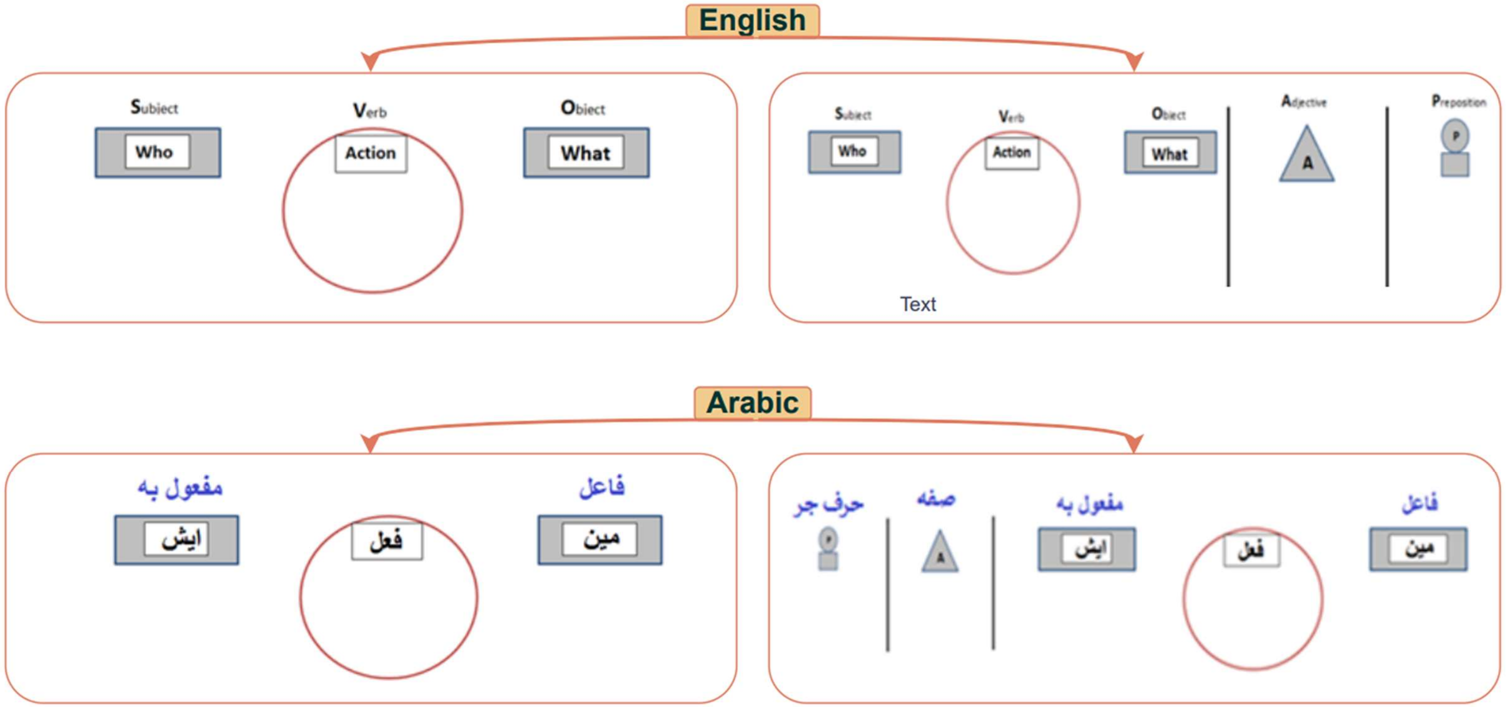


Figure 35 A screenshot of the SentenceShaper® working space layout

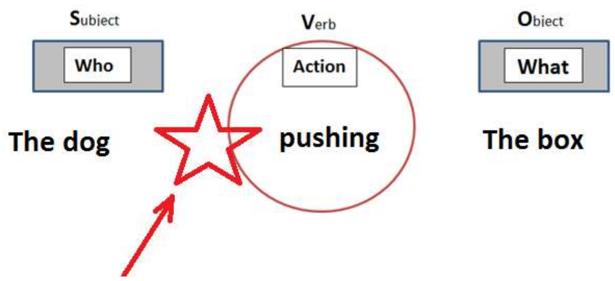
Figure 36 Example of the written scaffold used in Chapter 4 and Chapter 6



A written scaffold is used in therapy to build a mental representation of the sentence structure and stimulate self-cueing strategies. The clinician would fill the scaffold by transcribing online the participant's productions and illustrate to them the sentence they had produced

In case support is needed, the clinician will present it in the following progression:

- First step: the clinician would present a non-specific prompt, " is there anything missing in the sentence?"
- Second step: the clinician would then point to the blank in the scaffold, indicating a missing element, and ask, "what is missing here?"



- Third step: the clinician would model the missing element by saying " *is* was missing here" and writing it in the allocated slot. Then a delayed repetition will be prompted by pausing for a few seconds and then asking, "what was missing from the sentence?"

Then, the participant will be instructed to record the missing word, insert the snippet in its correct position in the sentence row, and replay it to self-assess. When a satisfactory production has been achieved, the participant is instructed to repeat the sentence production verbally from memory, as the scaffold will be taken away, and only the symbols on SentenceShaper's screen will remain. The aim is to assist in the consolidation of the practised skill.

The same above described steps will be implemented for a missing subject, verb, or object in a sentence with a few more cueing strategies. For example, if the missing word is "the dog”, if the non-specific prompt wasn’t successful, the clinician will point to the empty box in the scaffold and ask what is missing there. The next prompt will include a Wh-question, “who is pushing the box?”. If further support were needed, the following cues would be provided in the same progression and as needed:1. semantic cue such as “is it a cat?”. 2. phonemic cue of the initial sound /d/. 3. written cue of the first letter “d”. If all of the above cues did not achieve the desired outcome, the participant is then prompted to click on the button with the matching picture on the left side column of the page and to listen to the audio cue. After hearing the vocabulary cue, the participant is instructed to repeat it verbally, record it in a snippet and insert it in the correct slot in the sentence row. Then, the participant is instructed to replay the complete sentence and self-assess it. Finally, the participant is asked to verbally produce the sentence from memory three times before moving to the next page/ item.

Appendix 4 Instructions for recording the Arabic prompts on SentenceShaper for IOS.

Workbook VNest level-2

- Record the instructions in standard Arabic (الفصحى); however, when you record the target verb, pronounce it in the central dialect.
- For pages 5-9 fill in the (---) with the target verb from the list of 30 verbs.

Page		English	Arabic
1	Prompt	This workbook is based upon the VNEST approach developed by Dr. Lisa Edmonds the goal is activate verbs and their noun associates	يستند كتاب التدريبات الحالي إلى نهج VNEST الذي طورته الدكتورة ليزا إدموندز والهدف من هذا النهج هو تنشيط الأفعال و الأسماء المرتبطه بها
2	Prompt	The 15 verbs used here are from other SentenceShaper workbooks along with some light verbs like have	يحتوي كتاب التدريبات على مجموعة من الأفعال
3	Prompt	For each verb you will first record 3 transitive sentences with the verb then you will choose one sentence and expand it with information about where, why, and when and finally you will try to remember and record the 3 sentences	لكل فعل ، سوف تسجل أولاً 3 جمل مختلفة باستخدام نفس الفعل ، ثم تختار جملة واحدة منها وتسترسل بمعلومات حول أين ولماذا ومتى حدث هذا الفعل ، وأخيراً ستحاول أن تتذكر وتعيد تسجيل الجمل الثلاث.
4	Prompt	Suggested subject and object nouns are played on the side buttons in the level 1 workbook the text of these nouns are shown while it is covered in level 2 workbook, for many users level 1 will be too easy and they should start with level 2 workbook.	بضغط الأزرار الجانبية على الشاشه ، يمكنك الاستماع للاسماء المقترحة للفاعل والمفعول به
5	Prompt	We will start with the verb (---) please press forward arrow to continue	سنبدأ مع الفعل (---) يرجى الضغط على زر السهم للأمام للمتابعة
6	Prompt	Please create 3 sentences with the verb (---) you can play the blue buttons for the subject nouns or the green buttons for object nouns or you can use your own nouns	قم بإنشاء 3 جمل باستخدام الفعل (---) يمكنك الضغط على الأزرار الخضراء للاستماع للأمثلة المقترحة للفاعل أو الأزرار الزرقاء للاستماع للأمثلة المقترحة للمفعول به أو يمكنك استخدام امثلتك الخاصة
7	Prompt	Record a sentence with () and expand it with information about where, when, and why	كون جملة باستخدام الفعل (---) ، سجلها، ثم قم بالاسترسال بإضافة معلومات حول أين ومتى ولماذا حدث هذا الفعل

	Left column	Where- where might this happen? When- when might this happen? Why- why might this happen?	أين قد يحدث ذلك؟ متى قد يحدث ذلك؟ لماذا قد يحدث ذلك؟
	Right column	In في On على From من To إلى At عند Into داخل Because لأن So that عشان	
8	Prompt	Try to remember the 3 sentences you created with () and record them again	حاول أن تتذكر الجمل الثلاث التي أنشأتها باستخدام الفعل (---) وسجلها مرة أخرى
9	Prompt	The next verb is ( ) press forward arrow to continue	الفعل التالي هو (---) اضغط على السهم للأمام للمتابعة

VNeST verbs

- Record the (subject column) on the right side buttons with the colour green and the (object column) on the left side buttons with the colour blue.
- You can keep the same order of words for the subject column; however, for the object column record the words out of order (switch the order randomly for each verb).
- Ignore the column labeled (expansion)
- Pronounce the words in the central dialect (لهجة المنطقة الوسطى)
- If you ran out of pages, you can duplicate more pages:
  Home screen →workbook options →Edit this workbook →copy →paste
  or/ home screen →workbook options →duplicate this workbook → rename this workbook (then name it Level 2-Phase B – part 2 of 2)

	expansion	Object	verb	Subject
1	من الكمبيوتر	البيانات	**مسح**	المبرمج
	بالكحول	الجرح		الطبيب
	بالمساحه	السبورة		الطالب
	بيده	على شرابه		المصلي-المسلم
	بالمكنسة	الارض		العامل
2	للبيت	الرسالة	**وصل**	ساعي البريد
	للمطار	الزبون		سواق التاكسي
	بالتلفزيون	الأسلاك		الكهربائي
	للإذاعه	الخبر		المراسل
	للمحكمة	القضية		المحامي
3	بالفرشة	صورة	**رسم**	الفنان
	على السبورة	خريطة		المعلم
	للبناء	مخطط		المهندس
	على الكراسة	سمكة		الولد
	للنجاح	خطة		المستشار
4	في البحر	الصنارة	**رمى**	الصياد
	في السلة	الكورة		اللاعب
	في الزبالة	الورقة		الولد
	على الارض	اللعبة		الطفل
	للغريق	الحبل		المنقذ
5	في ملف	الورق	**جمع**	السكرتير
	لاجتماع	الموظفين		المدير
	لحل المسألة	الارقام		الطالب
	من المزرعة	الفواكه		المزارع
	من موقع الجريمة	الأدلة		الشرطي
6	في الخشب	المسمار	**دق**	النجار
	على الموعد	الباب		الضيف
	بالمطرقة	الحديد		الحداد
	في الحفلة	العود		الموسيقي
	للقائد	تحية		العسكري

القرد		الشجرة	في الحديقة	
الحرامي		السور	للبيت	
الرياضي	**تسلّق**	الجبل	لكسب التحدي	7
الولد		الصخرة	في المنتزه	
المجتهد		سلم النجاح	بهمته	
الطباخ		الخضار	للسلطة	
القرد		الموزة	تحت الشجره	
الولد	**قشّر**	البيضة	للفطور	8
الرجال		التفاحة	بالسكين	
الحمّال		العربية	في المطار	
السواق		السيارة		
ساعي البريد	**دف**	الصندوق	الى الباب	9
الطالب		زميله	في الفسحه	
الممرض		السرير	الى غرفة العمليات	
الصحفي		مقالة	في الجريدة	
الطالب		الاجابة	على الورقة	
المعلم	**كتب**	السؤال	على السبورة	10
الجد		الوصية	لأولاده	
الطبيب		الوصفة	للمريض	
الخياط		القماش	قبل القص	
الممرض		الحرارة	في العيادة	
النجار	**قاس**	الخشب	في الورشة	11
المفتش		المعايير	للمشروع	
الطباخ		الكمية	من المقادير	
المحامي		عن موكله	في المحكمة	
المتهم		عن نفسه	قدام الشرطي	
الولد	**دافع**	عن صاحبه	في المدرسة	12
الباحث		عن فكرته	في المؤتمر	
المزارع	**قطف**	التفاح	من الشجرة	

القرد		الموزة	من الشجرة	13
الولد		الوردة	من الحديقة	
الطباخ		كزبرة	من النبتة	
المضيف		عصير	للضيوف	14
المتطوع		مساعدة	للمحتاجين	
الطالب	**قدّم**	بحث	للتخرج	
الرجال		هدايا	لاطفاله	
الاستاذ		التهاني	للمتفوقين	
الحارس		الكورة	في المباراة	15
الطفل		اللعبة		
الشرطي	**مسك**	الحرامي		
الأب		يد ولده	في التقاطع	
الصياد		الفريسة		
الكهربائي		الجهاز	بعد تصليحه	16
السواق		السيارة	في الجراج	
المدير	**شغّل**	الفريق	في مشروع جديد	
الطباخ		الفرن	في المطبخ	
الرجال		الراديو	في الصباح	
المسافر		الشنطة		17
الطباخ		الوجبة		
البياع	**حضر**	الطلب		
المعلم		الدرس		
الصيدلي		الدواء		
الطبيب		المريض	في العيادة	18
الميكانيكي		السيارة	في الجراج	
المحقق	**فحص**	الادلة	في موقع الجريمة	
الاخصائي		العينة	في المختبر	
المبرمج		البيانات	على الكمبيوتر	
الطباخ	**خلط**	المقادير	في القدر	19
الرسام		الالوان	على اللوحة	

الصيدلي		الدواء	حسب الوصفة	
الكيميائي		المحاليل	في الانبوب	
الجاهل		الأمور	ببعضها	
العميل		الموظف	عن الارشادات	
الطالب		الاستاذ	عن المسأله	
المريض	**سأل**	الطبيب	عن العملية	20
الولد		الأم	عن الغداء	
الزائر		المريض	عن احواله	
الأب		بالمستقبل	لأولاده	
المحامي		بالقضية		
العالم	**فكر**	بالتجربة		21
الطالب		بالإجابه		
الجندي		بخطة	للدفاع	
المريض		ملفه	من المستشفى	
الرجال		الشكوى	عن جاره	
الطالب	**سحب**	الدفتر	من الشنطة	22
الجرسون		المفرش	من على الطاولة	
السواق		المفتاح	من القفل	
الطفل		النصيحة	من والده	
الصايم		الآذان	للافطار	
الطالب	**سمع**	الجرس	للفسحة	23
الجندي		الانفجار	في الحرب	
المسافر		النداء	للرحلة	
الولد		في الحديقة		
الرياضي		في السباق		
الشرطي	**ركض**	لموقع الجريمة		24
الكلب		لصاحبه		
المذيع		الخبر	للجمهور	
الطبيب	**أكّد**	التشخيص	للمريض	25
القائد		الخطة	للجنود	

المسافر		الحجز	للرحلة	
البياع		السعر	على الزبون	
الرجال		شكوى	على جاره	
العامل	**رفع**	الطوب	بالرافعة	26
المُخْرِج		الستار	في المسرح	
الطباخ		الغطا	عن القدر	
الأب		نصيحة	لولده	
الأستاذ		المحاضرة	للطلاب	
الشرطي	**أعطى**	مخالفة	للسائق	27
المدير		إجازة	للموظفين	
الزبون		الفلوس	للبياع	
البياع		الطلبية	للزبون	
الرجال		على الضيوف	في المجلس	
المدير	**سلم**	المهمة	للموظفين	28
الطالب		الواجب	للمعلم	
المؤمن		الأمر	لربه	
المعلم		الاختبار	للطلاب	
المذيع		الخبر	للجمهور	
المحرر	**صحح**	المقال	للصحفي	29
المحامي		العقد		
البياع		الفاتورة		
الحمال		الشنط	للبيت	
العامل		أدوات البناء	للموقع	
الأب	**شال**	الطفل	على كتفه	30
الجرسون		الصينية	من المطبخ	
العلاج		الخطر	عن المريض	

Appendix 5 Research Ethics Committee Approval

Kingdom of Saudi Arabia
Ministry of Health
King Fahad Medical City
(162)

وزارة الصحة
مدينة الملك فهد الطبية
King Fahad Medical City

المملكة العربية السعودية
وزارة الصحة
مدينة الملك فهد الطبية
(١٦٢)

IRB Registration Number with KACST, KSA: H-01-R-012
IRB Registration Number with OHRP/NIH, USA: IRB00010471
Approval Number Federal Wide Assurance NIH, USA: FWA00018774

February 5, 2019
IRB Log Number: 19-071E
Department: External - University of Manchester
Category of Approval: EXEMPT

Dear Nourah Alohali, Paul Conroy and Stefanie Bruehl,

I am pleased to inform you that your submission dated January 29, 2019 for the study titled '**Extending the Scope of Sentence Therapy in Aphasia Across English and Arabic**' was reviewed and was approved according to Good Clinical Practice guidelines. Please note that this approval is from the research ethics perspective only. You will still need to get permission from the head of department or unit in KFMC or an external institution to commence data collection.

We wish you well as you proceed with the study and request you to keep the IRB informed of the progress on a regular basis, using the IRB log number shown above.

Please be advised that regulations require that you submit a progress report on your research every 6 months. You are also required to submit any manuscript resulting from this research for approval by IRB before submission to journals for publication.

As a researcher you are required to have current and valid certification on protection human research subjects that can be obtained by taking a short online course at the US NIH site or the Saudi NCBE site followed by a multiple choice test. Please submit your current and valid certificate for our records. Failure to submit this certificate shall a reason for suspension of your research project.

If you have any further questions feel free to contact me.

Sincerely yours,

Institutional Review Board
Approved
Date: 05 FEB 2019

Prof. Omar H. Kasule
Chairman, Institutional Review Board (IRB)
King Fahad Medical City, Riyadh, KSA
Tel: + 966 1 288 9999 Ext. 26913
E-mail: okasule@kfmc.med.sa

Kingdom of Saudi Arabia
King Saudi University (034)
p.o. Box 7805 Riyadh 11472
Tel: +966 11 467 00 11
Fax: +96611 467 19 92
http://medicalcity.ksu.edu.sa

المملكة العربية السعودية
جامعة الملك سعود (٣٤.)
ص.ب ٧٨٠٥ الرياض ١١٤٧٢
هاتف: +٩٦٦١١٤٦٧٠٠١١
فاكس: +٩٦٦١١٤٦٧١٩٩٢

جامعة الملك سعود
King Saud University
المدينة الطبية الجامعية

05.08.2019 (04.12.1440)
Ref. No. 19/0865/IRB

To:	**Ms. Nourah Alohali** Department Of ENT King Saud University College of Medicine King Saud University Medical City PHD Student at , The University of Manchester - UK Email: n.ohali@hotmail.com Principal Investigator
CC:	Paul Conroy, Stefanie Bruehl Co-Investigators
Subject:	**Approval of Research Project No. E-19-4018**
Study Title:	"Extending the Scope of Sentence Therapy in Aphasia across English and Arabic"
Type of Review:	Expedite
Date of Approval:	05 August 2019
Date of Expiry:	05 August 2020

Dear Ms. Nourah Alohali,

I am pleased to inform you that your above-mentioned research project submitted to the IRB was reviewed and approved on 05 August 2019 (04 Dhul Al-Hijjah 1440). You are now granted permission to conduct this study given that your study does not disclose participant's identity and poses no risk to the patients.

As principal investigator, you are required to abide by the rules and regulations of the Kingdom of Saudi Arabia and the research policies and procedures of the KSU IRB. If you make any changes to the protocol during the period of this approval, you must submit a revised protocol to the IRB for approval prior to implementing the changes. Please quote the project number shown above in any future correspondence or follow-ups related to this study.

This approval is for a period of one (1) year commencing from the date of this letter. If you wish to have your protocol approved for continuation, please submit a completed request for reapproval of an approved protocol form (KSU-IRB 017E) at least 30 days before the expiry date. Failure to receive approval for continuation before the expiration date will result in automatic suspension of the approval of this protocol on the expiration date. Information collected following suspension is unapproved research and can never be reported or published as research data.

We wish you success in your research and request you to keep the IRB informed about the progress and final outcome of the study in a regular basis. If you have any question, please feel free to contact me.

Thank you!

Sincerely yours,

Prof. Abdulrahman Alsultan
Chairman of IRB
Health Sciences Colleges Research on Human Subjects
King Saud University College of Medicine
P. O. B ox 7805 Riyadh 11472 K.S.A.
Email: aalsultan1@ksu.edu.sa

College Medicine
Institutional Review Board (IRB)

/braezell

P.O. Box 64399 Riyadh 11536 Saudi Arabia
Telephone: 966 11 562 0000 Fax: 966 11 562 0236

مدينة سلطان بن عبد العزيز للخدمات الإنسانية
SULTAN BIN ABDULAZIZ HUMANITARIAN CITY

11th April 2019

MS. NOURAH ALOHALI
Primary Investigator
PhD Student, M.Sc.
University of Manchester
Neuroscience and Experimental Psychology DNEP
Biological Sciences
Zochonis Bldg., Room 3.01
Manchester, M1L3 9GB

Subject : Your Research Title: *"Extending the Scope of Sentence Therapy in Aphasia across English and Arabic".*

Dear Ms. Nourah Alohali,

Thank you for submitting your above-mentioned research proposal which the committee had made the final reviewed. I pleased to say, that the Research & Ethics Committee had approved it and with the below pointers to be followed when starting your data collections:

Effective Date of Data Collections: 1st September 2019
Total period of data collection: 4 months
Start date: 1st September 2019
End date: 31st December 2019

Kindly present this letter to the head/manager of the department/rehab ward in which data will be collected. The head/manger has the authority to do all needed arrangement to facilitate data collection. You are advised to quote the project number indicated herein in all transactions and communications. You are advised to submit a progress report, this time after three (3) months from approval of your research proposal, in relation to this research schemed we need you also to send us a final report after your research will be completed in relation to this research schemed to update the committee of its results. **You may informed that, for each participant from SBAHC, you will be strictly required to copy the signed consent form and hand it to the nurse to be added to their medical file.**

I trust you research scheme proves fruitful and beneficial to you, the patients and this institution.

Your research protocol has now been documented under:

Project Number: 040/ 2019/1st September
Series of : 2019 / 1st September

Thank you and best regards.

Sincerely yours,

11/4/19

MR. DAKEEL ABDULRAHMAN AL JEDAIE
Acting Chairperson of the Research & Ethics Committee

NHS

National Research Ethics Service

North West 5 Research Ethics Committee - Haydock Park

North West Centre for Research Ethics Committees
3rd Floor - Barlow House
4 Minshull Street
Manchester
M1 3DZ

Telephone: 0161 625 7819
Facsimile: 0161 237 9427

03 March 2010

Professor Matthew A Lambon Ralph
Neuroscience and Aphasia Research Unit (NARU)
School of Psychological Sciences
Zochonis Building
Oxford Road
Manchester M13 9PL

Dear Professor Lambon Ralph

Study title:	**Neuropsychological investigation of memory and language problems with patients with brain damage: programme of research**
REC reference:	**01/8/094**

Thank you for your letter of 23 February 2010, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair (Dr D Manning – Consultant Paediatrician).

Mental Capacity Act 2005

The members of the committee present approved the supplementary application on the basis described in the documentation submitted. I confirm that the committee has approved this research project for the purposes of the Mental Capacity Act 2005. The committee is satisfied that the requirements of section 31 of the Act will be met in relation to research carried out as part of this project on, or in relation to, a person who lacks capacity to consent to taking part in the project.

Confirmation of ethical opinion

The research continues to have a favourable opinion from this committee. It should continue to be conducted on the basis previously approved by the committee, as amended by this supplementary application. The conditions of approval issued with the committee's original favourable opinion continue to apply.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Covering Letter: From Professor M.A Lambon Ralph		13 January 2010
Covering Letter: From Professor M.A Lambon Ralph		18 November 2009
Notice of Amendment		13 January 2010
MCA 1 Form		13 November 2009
Protocol	2	13 January 2010
Participant Information Sheet	2	22 June 2007
Participant Consent Form	3	13 November 2009
Participant Information Sheet: MRI Scanning	3	13 November 2009
Participant Consent Form: MRI Scanning	3	13 November 2009
Participant Information Sheet: For Consultees	1	13 November 2009
Participant Consent Form: Consultee Declaration Form	1	13 November 2009
Response to Request for Further Information		23 February 2010
Participant Information Sheet: MRI Scanning - For Consultee	2 (22_02_10)	22 February 2010
Participant Consent Form: MRI Scanning - Consultee Declaration Form	2 (22_02_10)	22 February 2010

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

Feedback on the application process

Now that you have completed the application process you are invited to give your view of the service you received from the National Research Ethics Service. If you wish to make your views known please use the feedback form available on the NRES website at:

https://www.nationalres.org.uk/AppForm/Modules/Feedback/EthicalReview.aspx

We value your views and comments and will use them to inform the operational process and further improve our service.

01/8/094	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

P.P **Dr Donal Manning**
Chair

Appendix 6 Digit Span (from Wechsler Memory Scale) (Wechsler, 1987)

Instructions

For each list recalled score 1 point. Test is terminated when patient fails both lists. The "span" indicates the longest length at which **one** list was recalled correctly

Forward

2.	4 7	1 8
3.	6 2 9	3 7 5
4.	5 4 1 7	8 3 9 6
5.	3 6 9 2 5	6 9 4 7 1
6.	9 1 8 4 2 7	6 3 5 4 8 2
7.	1 2 8 5 3 4 6	2 8 1 4 9 7 5
8.	3 8 2 9 5 1 7 4	5 9 1 8 2 6 4 7

Span =

Backward

2.	5 1	3 8
3.	4 9 3	5 2 6
4.	3 8 1 4	1 7 9 5
5.	6 2 9 7 2	4 8 5 2 7
6.	7 1 5 2 8 6	8 3 1 9 6 4
7.	4 7 3 9 1 2 8	8 1 2 9 3 6 5

Span =

Appendix 7

Table: Participants' performance on VAST subtests before and after therapy (raw scores)

Participant No. and initials		Verb comprehens ion		Sentence comprehens ion		Grammatica lity judgment		Action naming		Object naming		Fill-in verbs in sentences		Sentence production		Semantic association	
Max. score		48		45		30		41		45		19		192		19	
		Pre-*	Post -	Pre-	Post -	Pre-	Post -	Pre-	Post -	Pre-	Post -	Pre-	Post -	Pre-	Post -	Pre-	Post -
1	NS	39.5	43	33.0	41	20.0	24	24.0	30	30.0	39	15.5	18	98.5	175	15.5	16
2	SA	47.5	48	41.0	45	27.5	29	31.5	36	41.5	44	18.0	19	172. 5	187	18.0	18
3	SM	46.0	48	44.5	45	25.0	24	38.5	41	43.5	45	18.0	20	177. 5	183	18.5	19
4	ND	45.0	46	39.5	43	16.5	16	10.0	33	25.5	30	4.0	12	76.5	164	18.0	19
	Mean	44.5	46.3	39.5	43.5	22.3	23.3	26.0	35.0	35.1	39.5	13.9	17.3	131. 3	177. 3	17.5	18.0
	SD	3.5	2.4	4.8	1.9	4.9	5.4	12.2	4.7	8.8	6.9	6.7	3.6	51.4	10.1	1.4	1.4

*Pre-therapy score represents the average baseline score (baseline 1 + baseline 2)

Table: Participants' performance on WAB subtests before and after therapy (raw scores)

Participant No. and initials		Yes/No Qs		auditory word recognition		sequential commands		repetition		Object naming		word fluency		sentence completion		responsive speech	
Maximum score		20		60		11		15		20		20		5		5	
		Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
1	NS	15	20	55	56	9	7	6	11	15	16	1	5	4	5	5	5
2	SA	19	19	56	56	6	6	11	11	18	19	9	12	5	5	5	5
3	SM	20	19	60	58	10	10	12	13	19	20	9	13	4	5	5	5
4	ND	18	14	48	53	7	7	6	11	8	14	13	8	2	4	5	4
	mean	18	18	54.8	55.8	8.0	7.5	8.8	11.5	15.0	17.3	8.0	9.5	3.8	4.8	5.0	4.8
	SD	2.2	2.7	5.0	2.1	1.8	1.7	3.2	1.0	5.0	2.8	5.0	3.7	1.3	0.5	0.0	0.5

Table: Cookie theft discourse's word count per participant at four testing time points

Participant's initials	Baseline 1	Baseline 2	Average baseline	Interim 1	Interim 2	% increase B to Int1	% increase B to Int2
NS	2	6	4	11	13	+175.0%	+225%
SA	14	18	16	21	36	+31.3%	+125%
SM	15	20	17.5	15	20	-14%	+14.3%
ND	8	7	7.5	10	17	+33.3+	+126.7%
Mean (SD)	9.75(6.0)	12.75(7.3)	11.25 (6.5)	14.25(5.0)	21.5(10.1)	+26.7%	91.1%

*Abbreviations: %= percentage, B= average baseline, Int1= Interim 1, and Int2=Interim 2. **Formula used to calculate percentage of increase in scores: (Scores post therapy/ scores at baseline)*100

Table: Dinner Party discourse's word count per participant at four testing time points

Participant's initials	Baseline 1	Baseline 2	Average baseline	Interim 1	Interim 2	% increase B to Int1	% increase B to Int2
NS	4	22	13	16	27	+23.1%	+107.7%
SA	28	60	44	46	68	+4.5%	+54.5%
SM	67	45	56	49	67	+12.5%	+19.6%
ND	14	26	20	37	49	+85.0%	+145.0%
Mean (SD)	28.25(27.6)	38.25(17.6)	33.25(20.2)	37 (14.9)	52.75(19.3)	111.28	+58.6%

*Abbreviations: %= percentage, B= average baseline, Int1= Interim 1, and Int2=Interim 2. **Formula used to calculate percentage of increase in scores: (Scores post therapy/ scores at baseline)*100