A usage-based approach to language processing and intervention in aphasia

Claudia Bruns (née Heilemann)

University College London

A thesis submitted for the Degree of Doctor of Philosophy

September 2018

Declaration

I, Claudia Bruns, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

C. Bruns

Abstract

Non-fluent aphasia (NFA) is characterized by grammatically impoverished language output. Yet there is evidence that a restricted set of multi-word utterances (e.g., "*don't know*") are retained. Analyses of connected speech often dismiss these as stereotypical, however, these high-frequency phrases are an interactional resource in both neurotypical and aphasic discourse. One approach that can account for these forms is usage-based grammar, where linguistic knowledge is thought of as an inventory of constructions, i.e., form-meaning pairings such as familiar collocations ("*wait a minute*") and semi-fixed phrases ("*I want* X"). This approach is used in language development and second language learning research, but its application to aphasiology is currently limited.

This thesis applied a usage-based perspective to language processing and intervention in aphasia. Study 1 investigated use of word combinations in conversations of nine participants with Broca's aphasia (PWA) and their conversation partners (CPs), combining analysis of form (frequency-based approach) and function (interactional linguistics approach). In study 2, an on-line word monitoring task was used to examine whether individuals with aphasia and neurotypical controls showed sensitivity to collocation strength (degree of association between units of a word combination). Finally, the impact of a novel intervention involving loosening of slots in semi-fixed phrases was piloted with five participants with NFA.

Study 1 revealed that PWA used stronger collocated word combinations compared to CPs, and familiar collocations are a resource adapted to the constraints of aphasia. Findings from study 2 indicated that words were recognised more rapidly when preceded by strongly collocated words in both neurotypical and aphasic listeners, although effects were stronger for controls. Study 3 resulted in improved connected speech for some participants. Future research is needed to refine outcome measures for connected speech interventions. This thesis suggests that usage-based grammar has potential to explain grammatical behaviour in aphasia, and to inform interventions.

Impact statement

Aphasia, an acquired communication difficulty following brain damage such as stroke, results in difficulties with processing of words, multi-word expressions and sentences. It often impacts on the ability to use language in everyday conversations. Previous aphasia research has largely been motivated by a words-and-rules approach to language, focusing on single word- and sentence processing. While some fixed and semi-fixed multiword expressions such as "*you know*", "*wait a minute*", "*I don't know*" or "*I went to* ___" are often reported in clinical observations, relatively little is known about the processing of such combinations in aphasia. There is an innovative approach that provides a framework to explore and explain the processing of multiword expressions, called usage-based construction grammar. In this theory, language experience is at the core of grammatical knowledge and can be quantified by measuring the frequency with which words co-occur in language use.

This thesis adopted a usage-based perspective within a series of three empirical studies of the processing of familiar word combinations in aphasia, and explored the potential of usage-based principles to guide aphasia intervention. In study 1, use of word combinations in Broca's aphasia was characterized by employing a frequency-based tool. This frequency-based analysis was complemented by a qualitative analysis of use of the prominent phrase "*I don't know*". Study 2 combined corpus-based and psycholinguistic methods to examine real-time processing of common multiword expressions in participants with aphasia. The third study piloted a novel computerised usage-based intervention program for individuals with non-fluent aphasia.

The materials developed result in innovations for research and clinical practice. The word monitoring task, a reaction-time experiment devised for use in study 2, represents a new way of examining multiword processing in different speaker groups, with the potential for application to future research. A refined version of the intervention developed in study 3 has the potential to become an evidence-based resource for speech and language therapists who wish to work with their clients on connected speech, with high relevance to everyday conversations. Since the intervention is computerised, it allows individuals with aphasia to practice use of familiar phrases and variations of these in their own time at home, which is a way of increasing therapy dose. Selected results from study 1 are published in Aphasiology:

Bruns, C., Varley, R., Zimmerer, V.C., Carragher, M., Brekelmans, G., & Beeke, S. (2018). "I don't know": a usage-based approach to familiar collocations in non-fluent aphasia. *Aphasiology*. DOI: 10.1080/02687038.2018.1535692

Contents

List of Tables

List of Figures

Glossary

AAT	Aachen Aphasia Test (Huber, Poeck, Weniger, & Willmes, 1983)
AD	Alzheimer's Disease
A.D.A	Action for Dysphasic Adults Auditory Comprehension Battery (Franklin, Turner, & Ellis, 1992)
AIQ-21	Aphasia Impact Questionnaire-21 (Swinburn et al., 2018; https://www.aiq-21.net/)
AoA	Age-of-acquisition
AoS	Apraxia of speech
AQ	Aphasia Quotient (indicator of aphasia severity)
BF	Bayes Factor
Bigram	two-word combination (e.g., "*good morning*"; "*it's*")
BNC	British National Corpus
BNT	Boston Naming Test (E. Kaplan, Goodglass, & Weintraub, 2001)
CAT	Comprehensive Aphasia Test (Swinburn, Porter, & Howard, 2004)
Collocation	Chunk of words that co-occur more frequently than expected by chance (e.g., "*a little bit*"; "*don't know*")
CP(s)	Conversation partner(s)
ERP	event-related potential
FE(s)	formulaic expression(s)
FLAT	Frequency in Language Analysis Tool (Zimmerer & Wibrow, 2015)
HELPSS	Helm Elicited Language Program for Syntax Stimulation (Helm-Estabrooks, 1981)
IDK	"*I don't know*"
ILAT	Intensive Language-Action Therapy (Stahl, Mohr, Dreyer, Lucchese, & Pulvermüller, 2016)
IRR	inter-rater reliability
JASP	'Jeffrey's Amazing Statistics Program' (https://jasp-stats.org/; Goss-Sampson, 2018)
MI	pointwise mutual information; measures the strength of the association of the component words of a bi- or trigram

MIT	Melodic Intonation Therapy (Helm-Estabrooks, Nicholas, & Morgan, 1989)
MR	Matrix Reasoning (subtest of the WASI-II)
MWEs	multi-word expressions
NFA	non-fluent aphasia
OANB	Object and action naming battery (Druks & Masterson, 2000)
PALPA	Psycholinguistic Assessments of Language Processing in Aphasia (Kay, Coltheart, & Lesser, 1992)
PWA(s)	participant(s) with aphasia
R1 / R2	rater 1 / 2 (in the context of inter-rater reliability)
REST	Reduced Syntax Therapy (Schlenck, Schlenck, & Springer, 1995)
RH / RHD	right hemisphere / right-hemisphere damaged
RT(s)	reaction time(s)
RUs	recurrent utterances
SPPA	Sentence Production Program for Aphasia (Helm-Estabrooks & Nicholas, 2000)
SWORD	Sheffield WORD – Structured speech therapy (Whiteside et al., 2012)
t-score	a measure of the association of the component words of a bi- / trigram, sometimes referred to as a measure of collocational significance
Trigram	three-word combination (e.g., "*good morning everyone"; "it's nice"*)
TROG-2	Test for Reception of Grammar – Version 2 (Bishop, 2003)
TTR	type-token ratio: number of types divided by tokens (traditionally used at the single-word level)
VAC	verb argument construction
WAB	Western Aphasia Battery (Kertesz, 2007)
WASI-II	Wechsler Abbreviated Scale of Intelligence – Second Edition (Wechsler, 2011)
WGT	'Words that go together' test (Dąbrowska (2014b)
WMG	Word monitoring game
WMT	Word monitoring task

Acknowledgements

Most importantly, I wish to express my deepest gratitude to my supervisors Suzanne Beeke, Rosemary Varley and Vitor Zimmerer who were always there to provide guidance and support, and who spent many hours discussing my project and reading and commenting on numerous drafts. I have learned *a great deal* about aphasia research and academic writing from you, and I am very thankful for such a wonderful and inspiring supervisory team.

I would also like to thank all participants for taking part in my studies, and Suzanne Beeke and Marcella Carragher for providing me with access to their fascinating conversation data.

Further thanks go to Rosemary Varley, Francina Clayton and Mickey Dean for volunteering to record various stimuli for my experimental tasks, and to Mike Coleman who programmed part of the computerised intervention. A special thanks goes to Gwen Brekelmans who was ‘rater 2’ in study 1 and 2, to Anna Volkmer for her input in designing acceptability questionnaires, and to the UCL Communication Clinic team for their input for study 3. I would also like to thank Yana Arkhipova for her help with transcriptions and Andrew Clark who supplied me with a microphone during the Intervention study.

Moreover, I am deeply grateful to my fellow PhD students who I shared an office with. Many thanks for supporting me along the way of this PhD journey. In particular, I would like to thank Anna Volkmer, Gwen Brekelmans, Lena Blott and Vanessa Meitanis for being such knowledgeable, fun and fantastic colleagues and friends.

During my Upgrade, I received valuable input from John Swettenham and Wendy Best. Moreover, I would like to extend my thanks to Wendy Best not only for this input but also for giving me the opportunity to work with the brilliant ‘Better Conversations with Aphasia’ team before I moved on to do this PhD.

This research was supported by a training grant awarded by the Economic and Social Research Council. Thanks to this grant, I could pursue my doctoral studies, had the opportunity to learn about new statistical methods and to disseminate my work at

various conferences. I am also grateful to the Studienstiftung des Deutschen Volkes for offering me access to their doctoral programme during my doctoral studies.

Finally, I cannot thank Freddy & my family enough for their encouragement, moral support, love and patience.

1. Introduction and aim of the project

It is fascinating to observe residual multi-word utterances such as “*don’t know*” and “*I went to* ____” in the speech of people with aphasia who otherwise produce grammatically impoverished, non-fluent language output. While these utterances often have a relatively complex structure, inconsistent with the rest of the speaker’s output, they are produced fluently, with good prosody and without articulatory effort. One interpretation of such structures is that they are residual constructions that remain available despite aphasia. This thesis aims to investigate how individuals with aphasia use and recognize such fixed and semi-fixed constructions and how these might inform aphasia rehabilitation.

The notion of constructions is grounded in usage-based or emergentist approaches to language (Bybee, 2010). Unlike rule-based, generative theories of grammar, which differentiate between language performance and competence, usage-based approaches assume that grammar emerges from language use. Linguistic knowledge is thought to be organised in constructions, pairings of linguistic form and semantic-pragmatic meaning (e.g., “*Where is X?*” to seek information with regard to place or location). These form-meaning pairings are central to language acquisition and processing. Since linguistic experience is at the core of this framework, variables such as the frequency with which items occur together, or type frequencies of certain items in constructions, determine representation of a construction as well as its productivity (Bybee, 2010). Linguistic experience can be quantified by measuring the frequency of occurrence (raw frequency) or observed relative frequency (e.g., occurrences per million words) of a word or word combination in a corpus. High-frequency word combinations, often linked to formulaic language, have been found to be processed more easily by native speakers, compared to structures with lower phrase frequency (Conklin & Schmitt, 2012).

Frequently repeated co-occurrences of words such as “*some of the*” or “*a little bit*” are known as collocations, and are examples of usage-based linguistic patterns. Word combinations can be more or less strongly associated (“*a little bit*” versus “*a lovely bit*”), as quantified by association measures such as mutual information (MI) and t-scores (Church, Gale, Hanks, & Hindle, 1991; Durrant & Doherty, 2010;

Gablasova, Brezina, & McEnery, 2017; Gries, 2010; Hunston, 2002). This is referred to as collocation strength. The association measure reported in this thesis is the t-score, a variable which is computed based on the assumption that the parts of a multi-word utterance occur together in a corpus only by chance. Higher t-scores point to more strongly associated/collocated words. More specifically, if a t-score is close to 0 or negative, words are weaker or not collocated, while a positive t-score means that there is more certainty about the presence of a collocation (e.g., "*a little bit*" has a t-score of 41, while "*a lovely bit*" has a t-score of 1).[1] Collocation strength is indicative of formulaicity (Zimmerer, Newman, Thomson, Coleman, & Varley, 2018).

Although usage-based approaches have been successfully applied to language processing and acquisition, specifically to early syntactic development (e.g., Lieven, Salomo, & Tomasello, 2009; Tomasello, 2003), the usage-based perspective is relatively new to language processing and grammatical behaviour in aphasia (Gahl & Menn, 2016). Since fixed phrases and semi-fixed constructions are often retained in aphasia, usage-based approaches offer potential to reveal new insights into grammatical behaviour in aphasia. The overall aim of this thesis is to investigate aphasic language processing from a frequency-based perspective, and to explore the impact of a novel intervention for aphasia motivated by usage-based principles.

1.1.Main research questions

Three overarching research questions are addressed in this thesis. Each relates to a different aspect of grammatical behaviour in aphasia:

1) What are the frequency-based properties of aphasic language output within naturalistic interactions?

2) Do individuals with aphasia show sensitivity to the usage-based factor of collocation strength in an on-line processing task?

3) Can the usage-based perspective inform intervention for aphasia?

[1] These t-scores are based on the spoken subsection of the British National Corpus (BNC, 2007).

1.2. Structure of this thesis

This thesis will report a series of studies that apply usage-based principles to language processing in post-stroke aphasia. It comprises an introductory chapter, three empirical chapters and a discussion chapter which synthesises and critically evaluates the results and describes opportunities for future research.

Chapter 2, ‘A usage-based perspective on language and language pathology’, introduces the main usage-based concepts and terms, followed by a description of formulaic expressions (FEs), highlighting the relevance of FEs to usage-based approaches, and the overlap between FEs and the frequency-based perspective applied in this thesis. The application of a usage-based perspective on aphasic language processing is described. After introducing the language profiles associated with aphasia, the chapter outlines research in aphasiology that is consistent with a usage-based theory.

Chapter 3, ‘Study 1: Exploring the use of familiar collocations in individuals with Broca’s aphasia’, is the first of three empirical chapters, in which the main focus is the usage of familiar collocations. Using a frequency-based analysis, spoken output produced by nine individuals with Broca’s type aphasia and their conversation partners is compared with regard to degree of association or collocation strength, followed by a comparison of these frequency-based characteristics across naturalistic conversations and semi-structured interviews in aphasia. Finally, this chapter presents an interactional linguistic analysis of one prominent collocation, “*I don’t know*”, drawn from naturalistic conversations.

Chapter 4, ‘Study 2: The recognition of familiar collocations in aphasia’, reports a study on recognition of words in more/less associated word combinations. An on-line word monitoring task was devised to test whether participants with aphasia show faster reaction times in response to a word in a strongly collocated phrase, compared to the same word embedded in a weaker collocation. Performance of a group of participants with aphasia is compared with an age-matched control group. Associations between task performance and other off-line tasks are presented, as well as an analysis of age effects across younger and older neurotypical groups.

Chapter 5, ‘Study 3: A case series report of a novel intervention targeting flexible use of familiar phrases in NFA’, addresses the application of usage-based principles to the rehabilitation of chronic post-stroke aphasia. A novel intervention is piloted employing a case series design with five individuals with non-fluent aphasia, and its impact is presented with a focus on changes in connected speech and the acceptability of the intervention to the participants.

Chapter 6, ‘Discussion and future directions’ draws the results from the empirical chapters together. The findings are evaluated, study limitations are addressed, and directions for future research are presented.

2. A usage-based perspective on language and language pathology

This chapter begins by outlining a usage-based perspective on language processing and its underlying assumptions and mechanisms (Section 2.1), presenting empirical evidence from three areas of research which support these assumptions: children's first language acquisition, second language learning, and adult experimental research. Section 2.2 addresses terminology and functions of FEs and their role within usage-based grammar as well as parallels and overlaps with a frequency-based approach to language. Section 2.3 introduces the linguistic features of aphasia, followed by an overview of aphasia research on FEs, and a review of studies on aphasia rehabilitation and aphasic sentence processing that are consistent with usage-based theory. The final section (2.4) presents a chapter summary.

2.1. Constructions: the building blocks of language?

When we converse, we combine linguistic units into larger packages such as multi-word utterances and sentences. Traditionally, at a grammatical level, individual morphemes such as "*I*", "*like*" or "*it*" have been regarded as basic building blocks of language which are combined into phrases and sentences by using abstract grammatical rules. This is known as a 'words-and-rules approach' (Pinker, 1999; Ullman, 2001; Ullman et al., 2005). In contrast, a constructionist, usage-based approach proposes that grammatical building blocks can be larger than single words. For instance, familiar multi-word utterances (e.g., "*I like it*", "*the other day*"), and more abstract patterns (e.g., V_{base} [$NP_{non\text{-}subject}$]! as in "*Open the door!*"; Diessel, 2015) could constitute a construction (A. E. Goldberg, 2003). Constructions are form-meaning pairings. For example, the imperative construction (V_{base} [$NP_{non\text{-}subject}$]!) can serve as a command ("*open the door*"), request ("*please pass me the salt*"), instruction ("*melt the butter in the saucepan*"), warning ("*be careful*"), permission ("*go on there*"), or good wishes ("*have a great birthday*"; Diessel, 2015). Constructions emerge from language usage. Consequently, whether or not a linguistic pattern is represented as a construction depends on an individual's language experience. Bybee (2010, p. 10) explains: "In usage-based theory, where grammar is directly based on linguistic experience, there are no types of data that are excluded from consideration because they are considered to represent performance rather than competence". Hence, usage-based grammar utilizes pre-collected, pre-transcribed spoken and written language as identified in large corpora such as the British National Corpus (BNC, 2007). Since linguistic experience is regarded as a critical variable for language processing, usage-based approaches are concerned with the frequency of occurrence of individual lexical items and constructions, and thus are in line with a "frequency-of-use perspective" (Conrad & Biber, 2004, p. 57).

Insights from Cognitive Linguistics have led to a family of usage-based approaches. The main idea of such approaches is that language is grounded in domain-general cognitive processes. Fillmore (1988), Lakoff (1987) and Langacker (1987, 1991) were among the first to propose specific theoretical underpinnings such as low- and high-level schemas, and entrenchment. Constructional schemas emerge from particular expressions (Langacker, 1991, 2008). Low-level schemas, "frames with slots" (Dąbrowska, 2014a, p. 619) are relatively concrete form-meaning pairings (e.g., "*Where is X*?"), while high-level schemas represent abstract templates (e.g., "*wh*-word *be*-TENSE NP"). Entrenchment describes a mechanism where speakers are claimed to have stronger representations for constructions that they come across more often (Langacker, 1987; Schmid, 2016). A consequence of linguistic entrenchment is that several units frequently occurring together fuse into a unit or chunk and might thus be stored holistically in memory. Chunking is an underlying learning mechanism for memory organization (Ellis, 2008a; Miller, 1956). Common examples of chunks are collocations such as "*and so on*", and phrases with articulatory reduction such as "*gonna*" (instead of "*going to*").

Other underlying domain-general mechanisms are categorization (identifying tokens by comparing them to previously established categories) and analogy (mapping a novel expression onto an existing schema on the basis of similarities of form or function; Bybee, 2010). Schematic slots (e.g., "*drive someone [ADJ]*") result in exemplar categories of constructions, as the more frequent member of a slot becomes the prototype (e.g., "*drive someone <u>crazy</u>*"), while new expressions (e.g., "*drive* *someone wild*") are formed based on analogy with the more frequent member (Bybee, 2010).

Constructions are distinguished based on their size, complexity and abstractness. A simplified illustration of this idea is shown in Figure 2.1. Linguistic knowledge is assumed to be represented in a "network of constructions: a 'construct-i-con'" (A. E. Goldberg, 2003, p. 219), where lexicon and grammar are represented on a continuum, i.e., there is no separation of lexicon and grammar (Boas, 2010). This idea is referred to as syntax-lexicon continuum (e.g., Stefanowitsch & Flach, 2016).

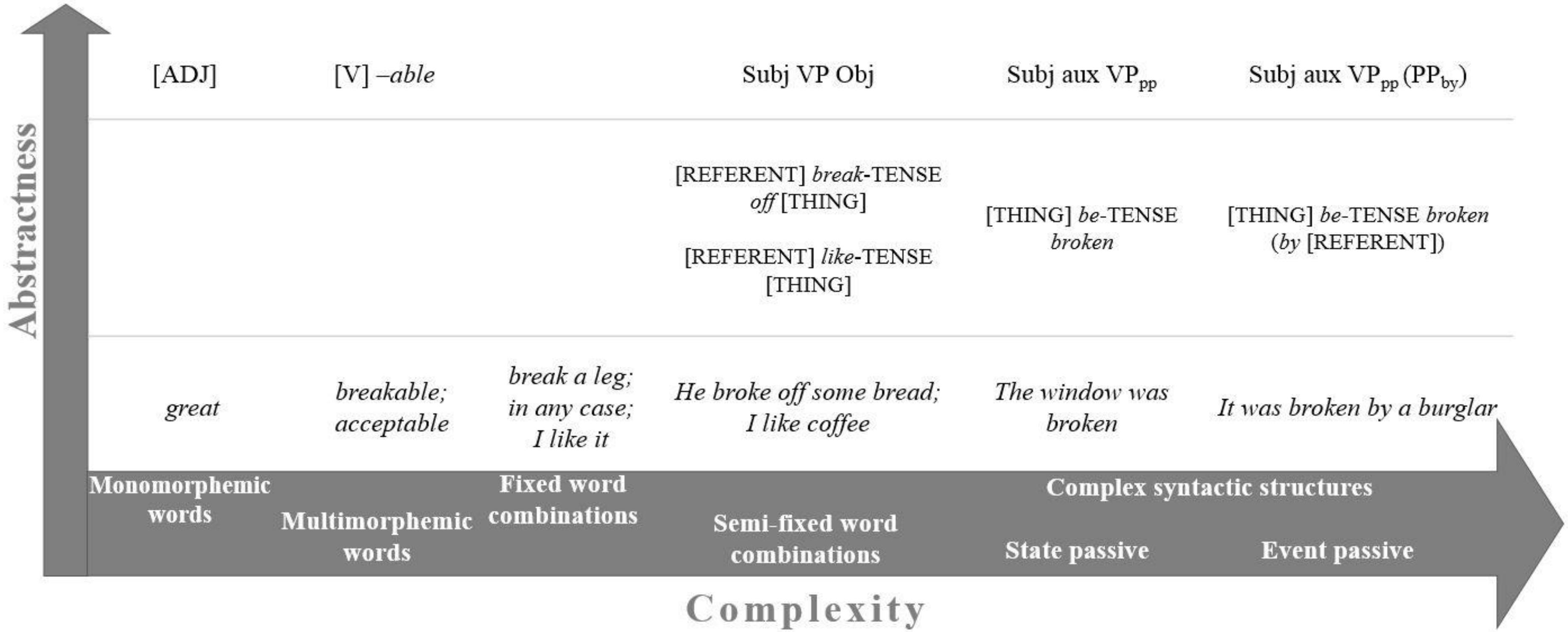


Figure 2.1: Examples of potential types of constructions based on their complexity and degree of abstractness; adapted and expanded from Stefanowitsch & Flach, 2016.

Constructions are thought to be strongly interconnected by various associative links (Dąbrowska, 2014a; Diessel, 2015; A. E. Goldberg, 1995) which can be vertical and horizontal. Vertical instance links are thought to exist as relations between a constructional schema (e.g., "[REFERENT] *like*-TENSE [THING]") and a lexically specific instance of it (e.g., *"I like coffee"*). Links at the same level of abstractness, horizontal subpart links, refer to the relationships between constructions that are similar with regard to their forms or meanings. Hilpert & Diessel (2016) propose that the expressions "*John wrote a letter*" and "*John wrote Mary a letter*" both represent an abstract subject-predicate construction and would thus be connected via a horizontal link.

An important characteristic of language production is creativity, or the potential to create novel utterances. In order to do so, a speaker needs a certain degree of productivity. The productivity of a construction is determined by the degree of abstractness of its schema. A requirement of productivity is the presence of at least one open slot in which different subunits can occur and can be inserted (e.g., "*I like NP*"). In language acquisition, this is based on processes such as analogy or schematization (e.g., Theakston, Ibbotson, Freudenthal, Lieven, & Tomasello, 2015). However, it is not just the presence of an open slot, but also its semantic properties that influence the degree of productivity of a construction. In the "*I want X*" construction, for example, the verb "*want*" determines the semantic constraints for the types that can appear in "*X*", for instance all 'wantable' things (Theakston et al., 2015). Boas' (2003) corpus-based analysis of the 'drive-crazy' construction, with data from the BNC, revealed that speakers show a high degree of productivity, even when an open slot presents with semantic constraints. While the 'crazy'-slot in the 'drive-crazy' construction always denotes a negative change of mental state, the corpus-derived list shows a number of creative alternatives to "*crazy"*, including "*wild*","*nuts*", "*mad / to madness",* and whole clauses such as "*(...) drove him* <u>*to drink more than usual*</u>".

Within a constructionist framework, the question is how novel expressions and ultimately an inventory of constructions is built. Dąbrowska & Lieven (2005) approached this question within child syntactic development, examining two constructionist operations: juxtaposition (linking two units, e.g. "*Tom" + "do you want a coffee?*") and superimposition (combining or fusing a unit with a frame, e.g., "*Do you want* <u>*coffee / tea / water*</u>*?*"). They analysed question constructions produced by two children at two time points, at 2;0 and 3;0 years old, using a corpus of mother-child interactions (around 30 hours of recordings per child at each time point). They used a traceback procedure (Lieven et al., 2009), where the corpus was divided into a test corpus (the last few transcriptions of the larger collection) and a main corpus (the remaining transcriptions), to identify how much novel language (in the test corpus) was based on what has previously been encountered (as identified in the main corpus). Results revealed that at 2;0, 55-66% of question types produced could be derived by one operation (e.g., superimposition into a [THING]-slot), while at 3;0, there were considerably fewer utterances (25-43%) that required one operation. Instead, more constructions needed more than one operation and the open slots were semantically more variable compared to at 2;0. Importantly, the majority (around 90%) of novel questions was based on lexical units that either the child or the mother had previously produced, including exact matches of previous utterances. Thus, one main conclusion of Dąbrowska & Lieven (2005) is that children start building more abstract constructions (e.g., "*Can* [THING] [PROCESS]?") based on generalizations over actual utterances (e.g., "*Can I sit?*") and frames with slots ("*Can I* [PROCESS]*?*").

That children's grammatical knowledge evolves from a bottom-up process, with constructions developing from formulas (e.g., "*I like it*") through frames with slots ("*I like X*") into more abstract schemas (SVO) with a wider and more abstract range of slots, has been supported by Lieven et al. (2009). They examined all multi-word utterances produced by four 2;0-year-old children in a 6-week corpus of mother-child interactions (28-30 hours per child). In line with Dąbrowska & Lieven (2005), they concluded that most of the novel utterances were strongly related to previously experienced utterances.

A growing body of experimental work with adults supports the claim that linguistic experience is critical to language processing. Dąbrowska (2012), for example, suggests that education, and thus amount of linguistic experience (especially with written language), shapes the knowledge of even the most basic constructions of a speaker's native language. Ellis, O'Donnell, & Römer (2014) examined verb argument constructions (VACs, e.g., "*he/she/it* <u>*[e.g., thinks / draws]*</u> *about the…*") in adult native speakers of English. Participants were required to complete phrases such as “*s/he* _____ *about the*” with a missing verb (first experiment) or all the words that came into their mind within one minute (second experiment). Findings demonstrate that the statistics of usage, such as the frequency of a verb in a particular VAC, shapes verb generation frequency in these off-line tasks (fill-in-the-blank task and verbal fluency task). This finding was replicated in a series of on-line experiments (Ellis, 2016). Moreover, Ellis (2016) proposed that greater semantic activation might add to the processing advantage of verbs which represent more prototypical meanings of a particular VAC (e.g., *about*-VAC: “*move about*” = semantically more protoypical than “*lie about*”).

Studies in second language learning also show that common multi-word utterances, as derived from neurotypical language usage, are a main feature of fluent, native-like language processing. Forsberg (2010) analysed interview data from six groups of speakers with varying language proficiency in French (adult beginners, high school students, advanced university students, very advanced L2 users, and two native speaker groups) to investigate the relationship between proficiency and use of common phrases such as “*no problem*”. The study found that the proportion of such phrases increased with proficiency level. In line with this finding, second language teaching has been influenced by learning materials including common expressions (Meunier, 2012).

Ellis (1996) argues that syntactic rules alone cannot explain how native speakers of a language communicate, but that that a large part of language learning can be characterised as sequence learning. For instance, the familiar expression “*I want to marry you*” is preferred by native speakers over grammatically correct, but less often encountered sentences such as “*I wish to be wedded to you*”. In other words, native-like communicative competence is characterized by frequently used, conventionalized form-meaning pairings. Ellis (1996) highlights individual differences in the ability to learn sequences (measured by nonword repetition or digit span, for example). Moreover, he assumes that chunking ability constitutes a major process in second language acquisition (see also Ellis, 2015).

As has been found in young children’s language production, juxtaposition and superimposition seem to represent useful mechanisms to analyse and explain adult spoken sentence production in conversation. Using traceback analysis of an adult’s child-directed speech, Dąbrowska (2014a) demonstrated that two thirds of the speaker's novel utterances could be derived from their previously encountered speech by recombining units with these two operations. One third represented exact repetitions or one-operation derivations, and another third was derived by combining units and chunks using two to four operations involving superimposition and/or juxtaposition. Dąbrowska (2014a, p. 642) calls this mechanism of creating novel expressions the "recycling mode" and suggests it may co-exist with an "analytic mode" of adult language production, involving more abstract linguistic knowledge.

Despite the accumulating evidence for usage-based effects in child and adult language processing, it is likely that language users switch between frequency-of-use and words-and-rules processing depending on communicative drivers in a particular situation. Conrad & Biber (2004) argue that an advantage of retrieving holistic units such as multi-word sequences is fluency that can be maintained even when time constraints are experienced, for example in real-time conversation. In line with this, Dąbrowska (2014a) proposes that adult speakers in informal conversation have a preference for the frequency-of-use or 'recycling' mode, that is, they mostly rely on low-level schemas rather than applying abstract rules to smaller chunks. Moreover, the assumption that linguistic knowledge is represented as constructional schemas is in accordance with the alternation between an idiom principle and an open choice principle, first suggested by Sinclair (1987) and also discussed by Erman & Warren (2000). Van Lancker Sidtis' (2004) dual-process model refers to these two modes as holistic and compositional processing. Overall, the alternation between an idiom- and open choice principle suggests that speakers have a repertoire of semi-pre-constructed expressions (idiom principle), and subunits such as individual words are selected to fill open slots (open choice principle) only if necessary. As Ellis (1996, p. 116) notes: "the syntactic rules abstracted from the 'linguistic environment' may be used in comprehension/production of new or difficult structures, whereas idiomatic analysis operates as a default: linguistic analysis is as deep as necessary and as shallow as possible". Hence, both processing modes might be valid, with their use depending on the communicative setting and situational factors.

2.2.Formulaic expressions

Fixed expressions and semi-fixed constructions (frames with slots) are useful to achieve native-like, fluent and error-free production. Those (semi-) fixed phrases that are frequently used and/or represent strongly associated word combinations are often referred to as FEs. FEs are familiar to native speakers and many of them are constructions and constructional schemas with a high frequency of use in neurotypical discourse. Thus, there is an overlap between FEs, high-frequency phrases and familiar collocations. The properties of FEs will now be discussed in more detail.

FEs are an essential part of daily conversation, and are frequently found in other discourse contexts such as weather forecasting or auctions, where time pressure explains their presence (Kuiper, 2000). Moreover, because of their familiarity to language users, FEs ensure rapid and successful comprehension by listeners (Wray & Perkins, 2000). FEs, apart from their more or less fixed form, typically convey a certain literal (e.g., "*at the moment"*) or more conventional (e.g., "*look forward to"*) meaning (Van Lancker, 1993). FEs such as "*I see*" or "*I don't think so*" are typically used to shape conversation (e.g., "*I see*" is used as a backchannel- or feedback signal, Erman & Warren, 2000), or to signal information in an efficient and understandable way (e.g., "*mind the gap*" = "*be aware of the gap between the train and the platform*"), amongst other functions (Schmitt & Carter, 2004). Thus, they are frequent in conversation, familiar to native speakers (e.g., Janssen & Barber, 2012; Jiang & Nekrasova, 2007; Tremblay & Baayen, 2010) and associated with discourse functions (Conrad & Biber, 2004; Schmitt & Carter, 2004).

However, there is a body of largely unstandardized terminology around FEs (see Table 2.1 for a selection of terms used in the literature; see also Wray & Perkins, 2000). 'Formulaic language' serves as an umbrella term. Typically, this covers both idiomatic and nonidiomatic expressions. Idioms such as "*once in a blue moon*" are strongly associated with FEs and are traditionally regarded as the prototype of multi-word sequences being stored holistically, i.e., as one large unit. The meaning of the idiom is usually described as figurative or "highly conventional" (Cacciari et al., 2006, p. 1305). This separation between the literal and figurative meaning of the idiom distinguishes it from most nonidiomatic FEs such as "*no problem*". Idioms vary in the extent to which their meaning can be derived from single constituents (e.g., "*she changed her mind*" versus "*she is on cloud nine*"; Gibbs, Nayak, & Cutting, 1989, as cited in Cacciari et al., 2006). Furthermore, while idioms are salient, they are characterised by a low frequency of use in typical conversation (e.g., Conklin & Schmitt, 2012). Therefore, findings from research on idiomatic expressions will stay in the background in the current thesis, where the focus is on nonidiomatic high frequency word combinations.

Table 2.1: Selection of terms and descriptors used in the literature to describe fixed and semi-fixed expressions

<table>
<tr><th colspan="3">Formulaic Language</th></tr>
<tr><th colspan="2">Nonidiomatic FEs</th><th>Idiomatic FEs</th></tr>
<tr><th>Terms used in literature</th><th>Examples</th><td rowspan="10">Idioms
(e.g., once in a blue moon)</td></tr>
<tr><td>Conventional expressions
(Edmonds, 2014)</td><td>no problem</td></tr>
<tr><td>Fixed units
(Beeke, Wilkinson, & Maxim, 2007a)</td><td>I suppose; I mean</td></tr>
<tr><td>Lexical bundles
(Tremblay, Derwing, & Libben, 2009)</td><td>in the middle of the</td></tr>
<tr><td>Low-level schemas / frames with slots
(Dąbrowska, 2014a)</td><td>I like [NP]</td></tr>
<tr><td>Memorized sequences of language and lexicalized sentence stems
(Ellis, 1996)</td><td>I’m sorry to keep you waiting
NP be-TENSE sorry to keep-TENSE you waiting</td></tr>
<tr><td>Multi-word expressions (including binomials and collocations)
(Siyanova-Chanturia, Conklin, Caffarra, Kaan, & van Heuven, 2017)</td><td>knife and fork</td></tr>
<tr><td>Partially lexically filled phrasal patterns
(A. E. Goldberg, 2003)</td><td>the Xer the Yer</td></tr>
<tr><td>Prefabs
(Erman & Warren, 2000)</td><td>[sb] look forward to [sth]</td></tr>
<tr><td>Speech interaction formulas
(Van Lancker Sidtis, 2012)</td><td>thank you</td></tr>
</table>

Psycholinguistic work with adult participants has shown that nonidiomatic FEs are processed more efficiently compared to matched, less frequent counterparts (e.g., “*kind of*“ (FE) versus “*sense of*”; “*knife and fork*” (FE) versus “*spoon and fork*”; “*to sum up*” (FE) versus “*to climb up*”; “*don’t have to worry*” (FE) versus “*don’t have to* *wait*", Conklin & Schmitt, 2012; Siyanova-Chanturia et al., 2017). As outlined in Siyanova-Chanturia & Martinez (2015) this processing advantage is seen both in recognition (e.g., Arnon & Snider, 2010; Sosa & MacFarlane, 2002; Tremblay, Derwing, Libben, & Westbury, 2011) and in production (Arnon & Cohen Priva, 2013; Janssen & Barber, 2012). Moreover, young children (2- to 3-years old) are sensitive to the frequency of language input at the level of four-word utterances (Bannard & Matthews, 2008). Bannard & Matthews (2008) showed that children were significantly faster and more accurate in repeating high- as compared to low frequency utterances (phrase frequency was determined by a large corpus of child-directed speech). Ambridge, Kidd, Rowland, & Theakston (2015) provide a review of frequency-related effects in child language acquisition at different levels of abstractness. These findings highlight the importance of FEs to fluent language processing both in children and in adults.

For most psycholinguistic research, the criterion of frequency of occurrence plays an important role: the more frequent a multi-word expression, the more likely it is to achieve formulaic status, as it is either thought to be processed as a single unit (Bybee, 2006) or represented in the lexicon (Siyanova-Chanturia & Martinez, 2015). Corpora, serving as representative sources of typical language use, are utilized for deriving stimuli for experimental research into the processing of familiar collocations. Importantly, one can measure the frequency of occurrence of a word combination (i.e., raw frequency or relative frequency with which a combination occurs in a corpus) or the frequency of co-occurrence with which two or more words are found in a corpus, relative to what would be expected by chance. The latter is a main criterion for labelling a word combination as a collocation (Gries, 2010). To determine whether an utterance might be a collocation, measures of association strength such as MI- or t-scores are used (Church et al., 1991; Gries, 2010). The higher the MI- and t-score, the more associated the units are, and thus, the more likely a word combination represents a collocation (Durrant & Doherty, 2010). Vilkaite & Schmitt (2017), for instance, operationalised the criterion of a collocation as at least 50 occurrences in the BNC and a minimum MI-score of 3. In this way, they ensured that these utterances would be used recurrently and their constituents would be strongly associated. Association measures should be regarded as relative, and can be ranked within each individual study as indicators of whether the elements of a word combination are more or less strongly associated.

Although research into FEs such as binomials or collocations is often dominated by usage frequency, the methods of studying FEs are influenced by other criteria regarded as relevant to formulaic status – and there are many that potentially affect formulaicity. Conrad & Biber (2004) list six characteristics which are typically used to identify FEs. They stress that depending on the aim of a study, a number of, or all of the following features – presented in Table 2.2 – are useful to consider: fixedness; idiomaticity; frequency; length of sequence; completeness in syntax, semantics or pragmatics; and intuitive recognition by native speakers of a language community. Adopting the perspective of communication sciences and disorders, Van Lancker Sidtis (2004) suggests the following properties for FEs: stereotyped form, conventionalized meaning, association with social context, inclusion of attitudinal and affective valence, and familiarity-recognition by native speakers. Some of these overlap with the features described in Conrad & Biber (2004). It is noteworthy that most of the criteria in Table 2.2 represent continuous rather than categorical variables. This makes strict differentiation between FEs and nonformulaic expressions challenging (Read & Nation, 2004).

Table 2.2: Comparison of properties of FEs suggested by Conrad & Biber (2004) and Van Lancker Sidtis (2004)

Properties of FEs / multi-word sequences	***Conrad & Biber, 2004**** *Corpus-based, frequency-driven approach*	***Van Lancker Sidtis, 2004*** *Communication sciences-approach*
	Fixedness	**Stereotyped form** of words and intonation
	Idiomaticity / Compositionality	**Conventionalized meaning** (also true for some nonidiomatic FEs)
	Frequency	-
	Length of sequence	-
	Completeness in syntax, semantics or pragmatics	Meanings are **associated with social context**
	Intuitive recognition by native speakers of a language community	**Familiarity** – **recognition** by native speakers (especially true for speech formulas such as greetings)
	-	Inclusion of **attitudinal and affective valence** (not true for some types of FEs, e.g., “*salt and pepper*”)

** Conrad & Biber (2004) term the linguistic material of interest “multi-word sequences”.*

The overlap between formulaicity, frequency and familiarity becomes evident from Table 2.2. Siyanova-Chanturia & Martinez (2015, p. 565) who use the term ‘multi-word expressions’ (MWEs) to highlight their frequency-based approach, explain that “the interplay between frequency and familiarity has not yet been fully understood; however, our stance is that frequency leads to familiarity and hence should be deemed as a primary characteristic of MWEs”.

The frequency with which language users encounter linguistic elements can affect language processing not only at the single word- and phrase level, but also in complex constructions such as the passive. As shown by Gries & Stefanowitsch (2004), verbs such as *"use"*, *"publish"*, and *"compare"* are passive bias verbs, i.e., they occur more frequently in the passive construction than what would be expected based on their overall frequency in a corpus. Verb bias, specifically, the frequency with which verbs occur in participle structures, has been found to reduce processing difficulties in ambiguity resolution (Trueswell, 1996).

2.3. A usage-based perspective on aphasia

Aphasia is an acquired language disorder caused by brain damage, most commonly as a result of stroke. Around one third of stroke survivors present with aphasia (Stroke Association State of the Nation report, 2018).[2] Individuals with aphasia typically show deficits in word- and sentence processing of oral and written language across modalities (Hallowell, 2017). A common distinction is made between fluent aphasia and non-fluent aphasia (NFA) according to characteristics of spontaneous speech (Goodglass, Kaplan, & Barresi, 2001b).

One of the classic subtypes of non-fluent language production is Broca's aphasia. Production in individuals with Broca's aphasia consists of simplified syntactic structures characterised by a lack of inflectional markers, prepositions, verbs and copulas (Goodglass et al., 2001b). Language production is telegraphic, that is, it mainly consists of content words and short phrases (Hallowell, 2017). This symptom complex is known as agrammatism. Other non-fluent subtypes of aphasia are global, motor transcortical and mixed transcortical aphasia (Potagas, Kasselimis, & Evdokimidis, 2013). Fluent aphasia, on the other hand, is classically associated with Wernicke's aphasia. As outlined in Goodglass, Kaplan, & Barresi (2001a, p. 67), "the varieties of fluent aphasia share the 'family resemblance' of having frequent uninterrupted runs of five or more words that are well articulated and grammatically

[2] Retrieved from https://www.stroke.org.uk/resources/state-nation-stroke-statistics (last accessed on 28/08/2018).

coherent”. One of the main characteristics of fluent aphasia is severe difficulty in word-finding (Goodglass et al., 2001b) and erroneous syntactic structure and/or grammatical morphology (Ruiter, Kolk, Rietveld, & Feddema, 2013), with individuals tending to produce neologisms (e.g., “*bring me a trunket*” instead of “*bring me a drink*”; Hallowell, 2017, p. 160). Although sentence construction may seem good for individuals with Wernicke’s aphasia (relative to NFA), it falls below neurotypical controls on a number of characteristics such as the production of fewer complex or well-formed sentences (Edwards & Tucker, 2006). Other subtypes of fluent aphasia are sensory transcortical, conduction and anomic aphasia (Potagas et al., 2013).

The test batteries used to determine the language profiles of individuals with aphasia typically consist of a spontaneous speech sample such as an interview with a clinician or researcher, naming objects or actions, describing pictured situations, and other decontextualized tasks such as reading and writing single words or matching words and sentences to pictures. However, production tasks such as elicited monologues do not reflect situations encountered in everyday life (Carragher, Sage, & Conroy, 2015). For instance, Beeke, Wilkinson, & Maxim (2003) revealed that the language output produced by speakers with agrammatism in conversations has a different grammatical structure to that of elicited speech. As a result, they suggest that targeting the grammar of everyday conversations in assessments as well as therapy may be more effective than targeting decontextualised language. This suggestion points to a performance-based approach to aphasia assessment and intervention to capture ‘everyday’ grammar – an idea that overlaps with usage-based assumptions, where language performance is at the core.

2.3.1. FEs in aphasia

The presence of short, familiar phrases in aphasic language production has stimulated a number of studies focusing on the forms, use and processing of FEs. Wray (2002a), for example, referring to previous investigations (e.g., Code, 1982), concluded that two- or three-word long casual interjections and idiosyncratically repeated phrases such as “*wait a minute*” seem to be relatively unaffected by aphasia

(Wray, 2002a, pp. 219–220). She listed five general categories of FEs commonly observed: memorized phrases (e.g., prayers or poems), conventional phrases (e.g., "*take care*"), idiosyncratic phrases (e.g., "*that's mine*"), involuntarily repeated phrases, and pause fillers (e.g., "*yes*"). Furthermore, Wray (2002a) proposed that formulaic frames with slots occur particularly in the language output of people with fluent aphasia. Based on an initial analysis of excerpts from two individuals with fluent aphasia, Wray (2002a) concluded that the slots of the frames mainly remain open, due to the speakers' word finding difficulties.

FEs are not only a communicative tool for neurotypical speakers (section 2.2), but they can also be interactionally beneficial for individuals with aphasia. McElduff & Drummond (1991) explored the functional use of FEs (which they call 'automatic utterances') based on monologic (picture description) and conversational speech samples from four speakers with NFA (all were in the acute phase). They demonstrate that FEs often have discourse functions, for example to comment or provide an answer. They highlight the fact that conversational speech was more suitable for eliciting and analyzing FEs with regard to their communicative functions, as compared to elicited monologues. That FEs can be a conversational resource in aphasia has also been shown by Simmons-Mackie & Damico (1997). In an analysis of conversational data from two speakers with NFA, they note that one individual used the phrase "*I don't know*" to yield the conversational floor, and the second speaker used the phrases "*very nice*" and "*all the time*" to express agreement and magnitude, respectively. Simmons-Mackie & Damico (1997) documented these FEs as one example of a compensatory strategy. Other studies found that the phrases "*I think*" and "*I suppose*" were used by a speaker with agrammatism to contribute to conversations (Beeke, 2003; Beeke et al., 2007a). More evidence for the communicative value of FEs comes from Simmons-Mackie, Kingston, & Schultz (2004), where a speaker with severe NFA, Pam, combined "*look*" with the number sequence "*1 2 3*" to get attention and imply quantity, which successfully conveyed a situational meaning. Thus, "Pam was able to strategically choose and place the words in contexts that helped the listener interpret her intents" (Simmons-Mackie et al., 2004, p. 120). While FEs can be a conversational resource, overusing such expressions can be a barrier to conversation. Wray (2012a) discusses the negative effects of over-reliance on FEs on interactions between individuals with Alzheimer's Disease (AD) and their carers.

The FEs available to individuals with aphasia are often partitioned out from analyses of elicited speech. Referred to as stereotypes, emotional utterances, and automatic speech, they are associated with pathological behaviour (Blanken & Marini, 1997; Rodrigues & Castro-Caldas, 2014). As Van Lancker Sidtis (2012, pp. 68–69) argues, “these terms are misleading because the utterances are often standard conventional expressions used intentionally to communicate (…)” and “the preserved utterances constitute remnants of natural competence for the very large repertory of formulemes”. Studies adopting this contemporary view of FEs, namely that such expressions are parts of a constructional inventory, have focused on the neural basis of FEs (Van Lancker & Kempler, 1987; Van Lancker Sidtis & Postman, 2006). Experimental investigations into the processing of FEs have largely employed idiomatic FEs (Lum & Ellis, 1999; Van Lancker & Kempler, 1987; Van Lancker Sidtis & Yang, 2017) and not explored the full range of idiomatic and nonidiomatic FEs, many of which are characterized by high frequency of usage.

As outlined in section 2.2, FE identification can vary depending on each study’s objectives, and criteria such as familiarity to native speakers or frequency of a phrase often represent continuous variables rather than binary distinctions (e.g., more or less frequent phrases, more or less familiar phrases). The studies conducted by Van Lancker Sidtis and colleagues (Bridges & Van Lancker Sidtis, 2013; Van Lancker Sidtis, 2012; Van Lancker Sidtis & Postman, 2006; Van Lancker Sidtis & Rallon, 2004) employ intuitive recognition of FEs by two or three native speakers. Those expressions which are agreed between raters are considered to be formulaic and are further analysed. Disadvantages of having raters decide about formulaic status include subjectivity, the time-consuming and labour-intensive features associated with this procedure, and most importantly the binary nature of FE classification, rather than a degree of familiarity or formulaicity.

The combination of corpus- and frequency-based approaches together with a functional analysis in discourse has great potential in providing a multifaceted and rich analysis of speaker behaviour. As suggested by Read & Nation (2004), the gold standard is to combine quantitative with qualitative approaches in order to achieve a comprehensive account of FEs. Recent developments provide novel methods to approach degree of formulaicity from a quantitative, frequency-based perspective

(Zimmerer et al., 2018; Zimmerer & Wibrow, 2015). These frequency-based methods operate at the multiword level and enable identification of more or less strongly associated word combinations, demonstrating that groups of speakers with different types of aphasia can be distinguished from neurotypical controls and participants with RHD, based on the frequency profiles of word combinations. Chapter 3 will discuss these methods in more detail, and how these can be complemented by a qualitative, interactional linguistic analysis to explore functional use of specific word combinations.

2.3.2. Usage-based approaches to aphasic sentence processing

Language processing in aphasia is known to be affected by usage-based variables such as age-of-acquisition (AoA; that is, the age with which a word is learned) and frequency of words (Brysbaert & Ellis, 2016; Kittredge, Dell, Verkuilen, & Schwartz, 2008; Nickels & Howard, 1995), and it might also be affected by syllable frequency (Aichert & Ziegler, 2004; Laganaro, 2005; Laganaro & Alario, 2006). However, frequency exerts an influence beyond the syllable- and single word level.

A number of investigators have researched the influence of verb transitivity bias on aphasic sentence processing (DeDe, 2013; Gahl et al., 2003; Knilans & DeDe, 2015). Gahl et al. (2003), for example, by testing comprehension of sentences in eight participants with varying subtypes of aphasia. They administered an off-line plausibility judgement task and assigned the stimuli to four syntactic frames: active transitive, passive, intransitive-undergoer subject, intransitive-agentive subject. For each of these constructions, verbs that occur frequently and verbs that occur infrequently in the given frame were tested (e.g., passive-bias verbs such as "*elect*" in a passive sentence versus active transitive-bias verbs, e.g., "*disturb*", in a passive sentence). One finding indicated that syntactic structure alone does not account for sentence comprehension difficulties, but that the likelihood of a verb appearing in a certain syntactic structure plays an important role. Thus, passive sentences were better understood when they included a passive bias verb. Moreover, DeDe (2013) extended these findings, using an on-line self-paced reading task, demonstrating that participants with aphasia are sensitive to whether a verb most commonly occurs in a transitive or intransitive construction (e.g., "*dance"* is intransitively biased, as in "*the couple danced every night*", while "*call"* is transitively biased, as in "*the agent called the writer from overseas to make an offer*"). The aphasic group (N = 10) showed disrupted reading times when there was a mismatch between a verb's transitivity bias and the sentence structure, for example, when the transitive bias verb "*call*" appeared in an intransitive sentence (e.g., "*the agent called from overseas to make an offer*"). These investigations represent empirical evidence of an influence of usage-based factors such as frequency cues (i.e., verb biases) on sentence processing in aphasia.

Gahl & Menn (2016) argue that future aphasia research should consider effects of an individual's linguistic experience, and therefore language usage. One such example is a recent eye-tracking study by Huck, Thompson, Cruice, & Marshall (2017), who explored the influence of target word frequency and contextual predictability (i.e., a word's predictability based on the preceding sentential context) on sentence reading in mild aphasia. Predictability of target words was determined using a cloze task with a normative group, where participants completed a sentence with up to three possible words. Highly predictable words had a minimum cloze probability of .84, while unpredictable (but still plausible) words had a cloze probability of less than .01. This was followed by a survey, determining predictability based on ratings (rating the fit of target words using a Likert-type scale). In this way, four types of sentences were created that included high- versus low frequency words embedded in a high- versus low predictability context. Results revealed that silent reading in aphasia was slowed when a word was unpredictable from the preceding context, and that high predictability facilitated the reading of low-frequency words. This study suggests that individuals with mild aphasia rely on the preceding sentence context in silent reading. Furthermore, usage-based, constructionist theories are beginning to be explored in doctoral theses in aphasia (Anderson, 2017; Hatchard, 2015).

2.3.3. Usage-based approaches to aphasia rehabilitation

To date, only few speech and language therapy interventions have been linked to usage-based principles. Riches (2013) piloted a usage-based intervention designed for children identified as having 'specific language impairment' (SLI; as it was then termed). Compared to traditional interventions for this population which often target morphemes (e.g., past tense *–ed*), Riches trained passives using two usage-based principles. The first was constructional grounding (Israel, Johnson, & Brooks, 2001) by which a complex construction such as an event passive "*The potatoes were mashed by the chef"* was grounded in a simpler construction such as a state passive "*the potatoes are mashed*". The second was construction conspiracy (Abbot-Smith & Behrens, 2006), whereby learning was based on the overlap between constructions. For example, the full passive "*The sausages were chopped by the dog*" is thought to be acquired by the prior acquisition of similar constructions with regard to form and function (e.g., "*I want my sausages chopped*). Using these assumptions, the 6-week intervention involved a stepwise treatment of passive structures, starting with state passives (e.g., "*the vase was broken*"), followed by practicing ambiguous passives (e.g., "*the vase was broken*" which can also be interpreted as an event passive), and finally producing event passives (e.g., "*the vase was broken by the dog*"). The intervention was piloted with two children, and outcomes were evaluated using a sentence repetition task and a comprehension task. Both children improved significantly in producing and comprehending passive constructions. While this study highlights the relevance of usage-based theory to speech and language therapy, application to aphasia is still in its early stages (Gahl & Menn, 2016).

Since most constructions are acquired during childhood, first language acquisition is an important research field for usage-based grammar (Diessel, 2015). However, usage-based principles are rarely used to study how constructions may be (re-)learned in populations with acquired language disorders such as aphasia. One exception is the research group around Stahl (Stahl, Mohr, Dreyer, Lucchese, & Pulvermüller, 2016) who have begun to make explicit links to usage-based phenomena such as (semi-)fixed FEs in the context of Intensive Language-Action Therapy (ILAT, Stahl et al., 2016; based on CIAT / CIATplus; Meinzer, Djundja, Barthel, Elbert, & Rockstroh, 2005).

ILAT is a therapeutic approach aiming to increase functional communication by training participants with NFA to use verbal requests within an interactive barrier task (for example, asking for matching picture cards using "*I want the X*", "*could I please have the X*"; rejecting requests; asking clarifying questions, and using FEs such as "*thank you*" / "*you're welcome*"). While the main aim is to name objects shown on the cards, Stahl et al. (2016) note that this intervention is motivated by usage-based theory (referring to Tomasello, 2005). The interactive structure of the language game encourages participants to practice fixed and semi-fixed phrases (Stahl et al., 2016). A recent crossover randomized control trial, comparing ILAT to a naming therapy (where barriers were removed and cards were described by using "*This is a X*" / "*Here I can see a X*"), showed better outcomes for participants in the ILAT group (Stahl et al., 2016). This indicates that training constructions within the dialogic setting of requesting picture cards is a powerful format to improve participants' expressive abilities. However, improved language production was evaluated by using a combined score of repetition and naming subtests of the Aachen Aphasia Test (AAT; Huber, Poeck, Weniger, & Willmes, 1983), including structures such as single words, complex words and sentences. Outcomes at the connected speech level were not reported separately, and everyday conversational data were not collected as part of the evaluation. Thus, it remains unclear to what degree ILAT stimulates improved phrase or sentence production in NFA.

The potential benefit of harnessing FEs in interventions for aphasia has also been highlighted by Stahl & Van Lancker Sidtis (2015). They argue that including such phrases in intervention programmes may have positive effects on a participant's motivation, well-being and quality of life, or with respect to ILAT (Stahl et al., 2016), may improve turn-taking. This suggestion is based on the finding that FEs represent a resource for individuals with aphasia as they may be processed in the right hemisphere (RH; Van Lancker Sidtis & Postman, 2006; Van Lancker Sidtis & Yang, 2017) and on studies investigating the role of phrases such as "*thank you*" in the success of Melodic Intonation Therapy (MIT; Helm-Estabrooks, Nicholas, & Morgan, 1989). Stahl, Henseler, Turner, Geyer, & Kotz (2013) and Stahl & Kotz (2014) for example, discussed the role of FEs in MIT, highlighting the motivating effect that FEs can have in the therapy process. According to Stahl et al.'s (2013) model of speech recovery, a combination of both ‘novel’, propositional speech and FEs may increase effectiveness of intervention for NFA and thereby maximise conversational abilities.

While other aphasia intervention studies, aimed at improving expressive abilities in NFA, were not directly motivated by or discussed in light of a usage-based approach, they include elements that can be reinterpreted from a usage-based perspective. The following paragraphs address such overlaps and parallels between published intervention studies and usage-based principles.

Reduced Syntax Therapy (REST), devised by Schlenck, Schlenck, & Springer (1995), is targeted at speakers with severe NFA. This intervention program aims at training reduced utterances (elliptic constructions such as the [DOING/DONE] + [WHAT] construction, as in “*drinking coffee*” or “*washing hands*”) and increasing their complexity step by step (e.g., [WHO] + [DOING/DONE] + [WHAT]). Structures are elicited by describing pictures or story completion (Springer, Huber, Schlenck, & Schlenck, 2000). In a later stage, REST helps to expand the frames provided by these structures, by adding lexical items for temporal and local modification, such as “*yesterday*”. The acceptability, user-friendliness and effectiveness of a web-based version of REST (e-REST) was recently explored within a single-case study by Ruiter, Rietveld, Hoskam, & Van Beers (2016). The therapy principles are in accordance with adaptation theory, a framework proposed by Kolk (1995) and used for planning compensation therapy for people with agrammatism (Ruiter, Kolk, & Rietveld, 2010). Adaptation theory assumes that agrammatism represents a linguistic deficit that can be described as a limitation of computational resources, resulting in the need to compensate for this difficulty: one way to adapt to language production difficulties is to produce simplified utterances such as ellipses. Reinterpreting REST within a usage-based perspective, the elliptic constructions act as conversational schemas (frames with slots). Thus, REST is in line with usage-based principles, where lexicon and syntax are thought to be represented on a continuum. Researchers exploring the efficacy of REST, however, do not make links with frameworks such as usage-based theory that might underpin and explain the results.

Carragher et al. (2015) took the REST principles and combined these with mapping therapy principles (e.g., Rochon, Laird, Bose, & Scofield, 2005) with a focus on learning the associations of thematic roles, e.g., the agent slot, and the types of words that can be inserted into that slot. A group of nine individuals with Broca's aphasia received 8 weeks of a stepwise treatment targeted at two- and three-constituent constructions: In a first stage, two-unit constructions such as "*Emma knitting*" were trained, followed by three-word constructions (e.g., "*Mike washing car*"). In a last step, the syntactic frame was expanded (e.g., "*Today Kate posting letter*"). Viewed from a usage-based perspective, this process reflects juxtaposition. Importantly, each frame was treated as semi-fixed which means that the participant could fill the open slots of a construction with a variety of lexical items (i.e., superimposition), reinforcing production of thematic roles around a verb. The authors addressed the question whether this type of hybrid intervention can improve everyday conversation skills. While increased syntactic well-formedness in producing trained and untrained sentence types was found within elicitation tasks, there was little evidence for change in conversation. Again, no link to usage-based variables such as phrase frequency of target constructions was made.

The Sentence Production Program for Aphasia (SPPA; Helm-Estabrooks & Nicholas, 2000), a refined version of the Helm Elicited Language Program for Syntax Stimulation (HELPSS; Helm-Estabrooks, 1981), aims at improving a speaker with NFA's functional communication by training a structured set of construction types such as Imperative Intransitive (e.g., "*come here*"), Imperative Transitive (e.g., "*answer the phone*"), WH-Interrogative ("*What are you making*?"), or Declarative Transitive ("*I make coffee*"). These structures are trained within a story completion format, motivated by Goodglass, Gleason, Ackerman Bernholtz, & Hyde (1972) and Gleason, Goodglass, Green, Ackerman, & Hyde (1975). The principles used in the SPPA are repetition and delayed auditory repetition within a hierarchical structure that administers a total of eight sentence types. While the SPPA includes a number of expressions with high communicative value, it is guided by abstract grammatical structures rather than constructionist principles such as frequency of usage, juxtaposition or superimposition. While Noël (2008) published a report of a gamified intervention using SPPA stimuli, and Silagi, Hirata, & Mendonça (2014) evaluated the effectiveness of a Brazilian Portuguese version of the SPPA based on a single case report, the evidence base for SPPA is currently limited.

Another approach, called 'Voluntary control of involuntary utterances' (Helm-Estabrooks & Albert, 2004; Helm & Barresi, 1980), trains words and phrases available and relevant to an individual's own communicative inventory, using reading and naming tasks. Examples of recommended target structures are "*love*", "*okay*", "*what*", and "*I don't know*". This is compatible with the usage-based assumption that building blocks can be of various sizes. However, there is little published research on the evidence base of this program.

Script training for aphasia (Bilda, 2011; Cherney, Halper, Holland, & Cole, 2008; S. Goldberg, Haley, & Jacks, 2012; Kaye & Cherney, 2016) is clearly focused on performance- and usage-based linguistic material (words, phrases and sentences) frequently found in everyday social situations such as going to a restaurant. Bilda (2011) tested a computer-based script training for chronic aphasia within a case series investigation, training scripts embedded in dialogues. Materials included semi-fixed phrases such as "*I'd like X*" when ordering a meal. Although there was no connected speech measure to investigate change in everyday conversations, results showed improved language production within trained conversational scripts, generalisation effects in untrained scripts, and – based on participant-reported questionnaires – increased confidence and more positive ratings with regard to participation in everyday conversations. A summary of 13 script training studies by Kaye & Cherney (2016) concluded that while scripts were successfully acquired in all studies, some of which found maintenance effects, there were limited generalization effects to untrained scripts or interactions with different communication partners. One criticism of script training is that while the linguistic material is often relevant to everyday situations, the flexibility of these phrases is limited. Linebarger, Romania, Fink, Bartlett, & Schwartz (2008, p. 1403) point out that "the kinds of functional situations (e.g., service encounters) that evolve in daily life typically require creating novel messages and cannot be negotiated with a set of high-frequency utterances designed for all users". In this context, Bilda (2011) reports that participants suggested practicing more individualised scripts.

A software program in which a participant's own speech can be recorded and listened back to, called SentenceShaper® (Linebarger, 2015; Linebarger et al., 2008), is compatible with some features of usage-based theory, in that conversational building blocks can be words, phrases or sentences. SentenceShaper® was designed to create coherent, individualized multi-word utterances by providing memory support: recorded conversational building blocks are kept in memory through listening back, new words or phrases can be added to existing recordings, and building blocks can be reordered. Thus, this therapy tool, motivated by the assumption that non-fluent speech production results from processing limitations (e.g., Kolk, 1995), gives users a work space for sentence planning. It has been shown that narratives produced with SentenceShaper® were longer and characterized by better grammatical structure (as measured by listener ratings and linguistic measures; Linebarger, 2015) but as noted in Newton, Kirby, & Bruce (2017, p. 21), "the system itself if not easily used as a prosthesis in ordinary conversation".

2.4. Summary

Despite a growing body of research highlighting the potential of usage-based approaches to aphasia, it represents an under-researched field, especially compared to the available evidence of usage-based effects in neurotypical language processing. Reinterpreting existing intervention programs for NFA from a usage-based perspective shows that many elements of existing intervention programs are compatible with the idea that language use shapes the conversational building blocks which can be words, (semi-)fixed phrases and sentences. There is evidence in the literature that frequency-related variables influence aphasic language processing at the word- and sentence level.

This thesis seeks to contribute to this area of research by exploring the frequency-based characteristics and functions of multi-word utterances produced by individuals with NFA in everyday conversations (study 1, Chapter 3), studying the recognition of familiar collocations (study 2, Chapter 4), and designing and evaluating a pilot intervention aimed at improving the use of frames with slots in NFA (study 3, Chapter 5). The main focus is on familiar word combinations and collocations, and so these terms are adopted. FEs will also be referred to in the following chapters since previous studies often used this term.

3. Study 1: Exploring the use of familiar collocations in individuals with Broca's aphasia

The material presented in this chapter is published in Aphasiology:

Bruns, C., Varley, R., Zimmerer, V.C., Carragher, M., Brekelmans, G., & Beeke, S. (2018). "I don't know": a usage-based approach to familiar collocations in non-fluent aphasia. Aphasiology. DOI: 10.1080/02687038.2018.1535692

3.1. Introduction

Despite grammatically impoverished, non-fluent speech in Broca's aphasia, residual multi-word utterances (e.g., "*I don't know*") are often fluently produced and employed in appropriate situations. Such multi-word utterances often represent common word combinations, some of which are referred to as FEs. The latter term describes a prefabricated sequence of words, as well as single words such as the discourse particle "*well*" (Van Lancker Sidtis & Postman, 2006; Wray, 2002b). FEs represent a significant part of the conversational inventory, and are a feature of proficient language use. Estimates vary from 24-48% (Van Lancker Sidtis & Rallon, 2004; in unscripted telephone conversations) to 59% of typical conversation (Erman & Warren, 2000). This study applies corpus methods to establish the collocation strength, or degree of association between components of multi-word expressions used by speakers with aphasia and their conversation partners (CPs). Word combinations which are more frequent and more strongly collocated, are more likely formulaic.

While the forms and functions of familiar collocations in neurotypical talk are well studied (e.g., Conrad & Biber, 2004; Drew & Holt, 1998; Kecskés, 2000; Meunier, 2012; Schmitt & Carter, 2004), systematic investigations of their use in aphasia are relatively rare. Code (1982) analysed recurrent utterances (RUs) across subtypes of aphasia. Code found that real-word RUs mostly consisted of high-frequency words (e.g., "*I told you*", "*so so*") and typically occurred in Broca's aphasia. A salient pattern was the pronoun + verb group (e.g., "*I want to*"). These forms are often described as stereotypes or lexical automatisms (Blanken, 1991; Code, 1982; Grande et al., 2008), suggesting pathological behaviour (Blanken & Marini, 1997; Rodrigues & Castro-Caldas, 2014). However, as typical language consists of large proportions of familiar collocations, whether they should be viewed as markers of pathology in Broca's aphasia is unclear.

Van Lancker Sidtis & Postman (2006) explored the neural basis of FEs. They recorded the use of numerals, proper nouns, idioms, conventional expressions (e.g., "*as a matter of fact*"), speech formulas (e.g., "*right*"), expletives (e.g., "*damn*"), sentence stems (e.g., "*I guess*"), discourse particles (e.g., "*well*") and pause fillers (e.g., "*uh*") in spoken output of neurotypical individuals, speakers with fluent aphasia, and individuals with right hemisphere damage (RHD). They found higher proportions of FEs in speakers with aphasia compared to individuals with RHD and suggested that these expressions are represented in the RH (Sidtis, Canterucci, & Katsnelson, 2009). Other investigators have explored idiom comprehension across participants with aphasia and RHD (Van Lancker & Kempler, 1987), and the processing of 'automatic language' (counting from one to 10; completing familiar, idiomatic phrases; repeating idiomatic phrases) in fluent and NFA (Lum & Ellis, 1999). The findings indicate relatively preserved production and comprehension of these expressions, many of them idiomatic, in aphasia.

As detailed in Chapter 2, familiar collocations, due to their high usage frequency, can be explained by theories such as usage-based Construction Grammar (A. E. Goldberg, 2003). Thus, this approach has potential for exploring familiar collocations in aphasia. While familiar collocations are known to be retained in aphasia, it remains unclear whether such combinations are a main feature of aphasic language production in everyday conversations, and whether they perform specific conversational functions in aphasia. The first analysis of study 1 adopted a frequency-based perspective to word combinations. Frequency characteristics were established at the level of two-word- (bigrams) and three-word combinations (trigrams) by employing automated analysis software, the *Frequency in Language Analysis Tool* (FLAT; Zimmerer et al., 2018; Zimmerer, Wibrow, & Varley, 2016; Zimmerer & Wibrow, 2015), to analyse language produced by speakers with aphasia in different talk contexts. The FLAT determines usage frequency of every word and word combination in a test sample using the 10 million word spoken-conversation section of the British National Corpus (BNC, 2007), which represents typical language use. In addition to frequency of occurrence, the FLAT automatically computes collocation variables of bi- and trigrams. A commonly used measure of collocation strength, the t-score, was employed. For instance, the trigram "*it's alright*" has a t-score of 28 (based on the spoken BNC), whereas "*it's new*" has a t-score of 4. Higher values point to more strongly associated or collocated word combinations, reflecting a higher certainty of collocation. In this way, analysis 1 explored the following research questions: a) How strongly associated are the word combinations produced by speakers with aphasia compared to CPs in naturalistic conversations? b) Do frequency-based values vary across talk contexts in aphasia? It was hypothesized that speakers with aphasia rely on more common multi-word utterances as compared to neurotypical control speakers, and that the frequency-based values are robust across different talk contexts.

Analysis 2 focused on familiar collocations in interaction. Conversation analytic studies show that such expressions can be interactionally beneficial in aphasia. For example, the "*I suppose*" construction (Beeke, 2003) was produced by a speaker with chronic Broca's aphasia five times in a 13 minute conversation to express his opinion. Familiar collocations have also been viewed as compensatory strategies in aphasic talk. Simmons-Mackie & Damico (1997) found that set phrases like "*very nice*" and "*all the time*" – although they might not add new information to the conversation – were used as a resource by two speakers with Broca's aphasia to regulate interaction, or in the case of "*all the time*", to express magnitude (section 2.3.1). The second analysis explored the conversational uses of "*I don't know*", which, in the current dataset, was common in non-fluent aphasic talk and control participants.

3.2.Method

3.2.1. Participants

This study used pre-collected everyday conversation data (dataset 1) and semi-structured interview data (dataset 2). Conversation data stem from nine dyads, each comprising a person with post-stroke NFA (at least 6 months post-onset) and their neurotypical CP. Dyads 1-8 took part in a study by Carragher, Sage, & Conroy (2013), and dyad 9 in a study by Best et al. (2016).[3] Background information on all dyads is summarized in Table 3.1. All participants presented with Broca's aphasia. Aphasia classification was based on clinical consensus, grammatically impoverished output on picture description tasks, and performance on standardised language assessments (Table 3.1; for more details see Beeke, Maxim, Best, & Cooper, 2011, p. 228; Carragher et al., 2013, p. 852). Dyads took part in intervention studies and recorded weekly conversations prior to, during and after intervention. Analysis is based on pre-therapy recordings only and so the nature of these interventions is not relevant here. Participants gave written informed consent to long term storage of their data. Ethical approval was granted by NHS IRAS ethics (Carragher et al., 2013) and NHS Cambridgeshire 1 Research Ethics Committee (Best et al., 2016).

[3] Dyad 2 in Best et al. (2016).

Table 3.1: Naturalistic interaction participant information

Dyad	**Initials**	**Corpus**	**Speaker with Broca's aphasia**					**CP relation to speaker with aphasia**
			Gender	**Age at time of recording**	**Time post onset (months)**	**Naming objects (% correct)**	**Naming actions (% correct)**	
1	**KK**	Carragher et al., 2013	male	48	24	15	18	Wife
2	**GL**		male	47	12	32	22	Partner
3	**BL**		male	64	57	45	31	Wife
4	**DC**		male	40	72	32	30	Father
5	**JH**		female	36	8	27	43	Husband
6	**PM**		male	67	47	60	59	Wife
7	**PG**		male	66	132	65	65	Wife
8	**DM**		male	48	36	72	85	Wife
9	**SC**	Best et al., 2016	male	39	30	83	55	Wife
			Mean	**50.6**	**46.4**	**47.7**	**45.1**	
			SD	**12.1**	**38.1**	**23.2**	**22.3**	

Naming objects: Percentages are derived from the Boston Naming Test (BNT; E. Kaplan, Goodglass, & Weintraub, 2001), used for assessing lexical retrieval for dyads 1-8; for dyad 9, the percentage reflects an average across three baselines of the OANB (Druks & Masterson, 2000) noun subset.

Naming actions: Percentages are based on performance in the OANB verb subset (an average across two baselines for dyads 1-8 and three baselines for dyad 9).

Dataset 2 comprised semi-structured interviews from the online database AphasiaBank (MacWhinney, Fromm, Forbes, & Holland, 2011). AphasiaBank classifies aphasia subtype with the Western Aphasia Battery (WAB; Kertesz, 2007). Only participant 9 (SC) from dataset 1 had a WAB assessment and so AphasiaBank samples were selected that matched as closely as possible to that speaker's WAB profile (Broca's aphasia, Aphasia Quotient [AQ] of 60.7). Thirty-nine participants (13 female, 26 male) were selected from AphasiaBank. They were classified as having Broca's aphasia and their mean age was 55.7 years (SD = 11.4). The AphasiaBank group had a WAB AQ between 50.0 and 70.4, indicating a range from moderate to mild aphasia. The mean time post-onset of the AphasiaBank group was 88 months (SD = 70; range: 6 to 309 months).

3.2.2. Datasets

With regard to dataset 1, there were forty pre-therapy videotaped conversation samples. These consisted of four samples recorded over 4 weeks for dyads 1-8, and eight samples over 8 weeks for dyad 9. All conversations were recorded in the dyads' home under instruction to videotape a conversation at a time of day and on a topic of the dyad's choosing. Conversations were transcribed as part of prior studies by Carragher et al. (2013) and Best et al. (2016), or by the researcher where additional untranscribed samples existed. Sampling consisted of 5- to 25-minute segments where the participant with aphasia was in conversation with a family member. As a standard, the first 5-minute segment of a recording was not sampled, to allow participants to feel less conscious about the presence of the camcorder (for one dyad, some pre-transcribed samples started at the beginning of the recording). The videoed samples yielded 269 minutes of transcribed conversation which forms the basis for the current analysis. The average transcriptions per dyad reflected 22 min of conversation for dyads 1-8 and for dyad 9, 93 min of conversation.

Dataset 2 comprised 50 discourse samples from the 39 participants selected from AphasiaBank (one sample for 31 participants, and in order to maximise sample size, two to three samples from eight participants who were tested on multiple occasions).

These reflect semi-structured interviews, where a clinician asks questions about a participant's speech, stroke story and an important event. The average duration of samples was 4:57 minutes (SD = 2:20). In total, 247 minutes of transcribed material from AphasiaBank were analysed.

3.2.3. Analysis 1

FLAT Version 1.1 was used for analysis. The program automatically extracts uni-, bi-, and trigrams in a sample of transcribed language, determines values on a number of measures of productivity, and derives frequency-related values from the spoken BNC for words and word combinations. For example, it segments the trigram "*it's alright*" into three unigrams ("*it*", "*'s*", "*alright*") and two bigrams ("*it's*" and "*'s alright*"). Since only grammatical utterances are expected to appear in the spoken BNC, and ungrammatical utterances are atypical and would therefore represent very low frequency or non-existent combinations in a normative database, only grammatical strings are analysed by FLAT. Prior to frequency-based analysis, all transcripts were formatted in a manner consistent with the conventions of the FLAT (Zimmerer et al., 2018, 2016). For example, the utterance "*but seven days cycling*" was analysed as two separate clauses: "*but seven days*" and "*cycling*". Clause boundaries were marked with separators ("<.>"). "<fill>" was used to replace any non-lexical interjections (e.g., "*erm*"), pauses and repetitions of words other than "*yes*"/ "*yeah*", "*oh*" and "*no*".

FLAT outputs also include measures of productivity: the number of word combinations, and combination ratio (the number of trigrams produced by a speaker divided by the number of words). Speakers with higher combination ratios display more output consisting of word combinations, i.e., better ability to combine words into multi-word expressions. More traditional indicators of lexical diversity are type-token ratio and vocabulary diversity (MacWhinney et al., 2011). However, these measures investigate productivity at the single word level, while the present study investigates the amount of connected speech produced. Hence, combination ratio is reported.

To determine degree of association between words of the bi- and trigrams, t-scores were used, an association measure that indicates whether words co-occur more frequently than would be expected if all the words in the corpus were randomly distributed (Durrant & Doherty, 2010; Gries, 2010; Hunston, 2002).[4] The trigram "*it's alright*" has a t-score of 28, whereas "*it's new*" has a t-score of 4. Hence, the words in "*it's alright*" are more strongly associated or collocated. Collocation strength is one marker of formulaicity (Zimmerer et al., 2018). This study reported t-scores if a bi- or trigram had a frequency of occurrence of >1. Thus, the results are based on combinations that occur in the spoken BNC. Moreover, FLAT outputs include both type and token bigram and trigram summary statistics. Type values reflect the inventory of bigrams and trigrams, while token measures reflect how frequently individual types are used. Table 3.2 shows details of FLAT variables and their calculation.

[4] The FLAT 1.1 employs additive smoothing by which 1 is added to every unigram, bigram and trigram frequency count in order to avoid a frequency value of 0 to enable calculation of t-scores.

Table 3.2: Frequency-related variables used in the FLAT analysis

		Utterance type		
		Unigram (e.g., *it; 's; alright*)	**Bigram** (e.g., *it's*)	**Trigram** (e.g., *it's alright*)
Variables	**Observed absolute frequency**: occurrence in the spoken BNC	253864 (*it*) 199264 (*'s*) 7994 (*alright*)	68662	772
	Observed relative frequency: occurrence per million words	25386.4 19926.4 799.4	6866.2	77.2
	Expected frequency (taking into account the approximate number of words in the spoken BNC, 10,000,000)	-	$\frac{frequency_{it} frequency_{'s}}{10{,}000{,}000}$ (e.g., 5058.60)	$\frac{frequency_{it} frequency_{'s} frequency_{alright}}{10{,}000{,}000^2}$ (e.g., 4.04)
	t-score	-	$\frac{frequency_{it's} - expected\ frequency_{it's}}{\sqrt{frequency_{it's}}}$ (e.g., 242.7)	$\frac{frequency_{it's\ alright} - expected\ frequency_{it'salright}}{\sqrt{frequency_{it's\ alright}}}$ (e.g., 27.6)
	Combination ratio	Raw number of trigrams produced by a speaker divided by number of words		

3.3. Results

For group comparisons, Shapiro-Wilk tests were used to determine whether variables were normally distributed. Where variables were normally distributed, two-tailed independent t-tests were performed; otherwise, two-tailed Mann-Whitney U tests were employed. Pearson's r was calculated to report effect sizes (trivial effect: $r < 0.1$, small effect: $r = 0.1$, medium effect: $r = 0.3$, large effect: $r = 0.5$; Goss-Sampson, 2018).

Combination ratio was compared for the two speaker groups. The mean combination ratio for speakers with aphasia was .17 ($SD = .06$), and for CPs, it was .54 ($SD = .03$), representing a significant difference, $t(16) = -15.16$, $p < .001$, $r = .97$ (large effect). This confirms the agrammatic status of the speakers with aphasia, characterized by the limited ability to produce grammatical multi-word expressions.

3.3.1. How strongly associated are the word combinations produced by speakers with aphasia compared to CPs in naturalistic conversations (dataset 1)?

The t-score profiles for bi- and trigram types of both speaker groups are shown in Figure 3.1. For bigrams, the nine speakers with aphasia produced a total of 988 types (range: 30-262) and 1809 tokens (range: 101-413), and their CPs produced 7467 bigram types (range: 317-2082) and 11160 tokens (range: 414-3636). The average type-based bigram t-score for speakers with aphasia was 26 ($SD = 7$), and for CPs it was 19 ($SD = 4$), representing a significant difference between the two groups: $t(16) = 2.71$, $p = .016$, $r = .56$ (large effect). With regard to bigram tokens, the aphasic group had a higher average bigram t-score compared to the CP group ($M = 53$, $SD = 23$ versus $M = 37$, $SD = 3$, respectively). This difference was not significant, $t(8.24) = 2.091$, $p = .069$, $r = .59$ (large effect). A higher bigram t-score in the aphasic group, however, indicates that the constituents of the bigrams were more strongly associated with each other. The large effect size combined with a non-significant difference could be due to the variability within the aphasic group.

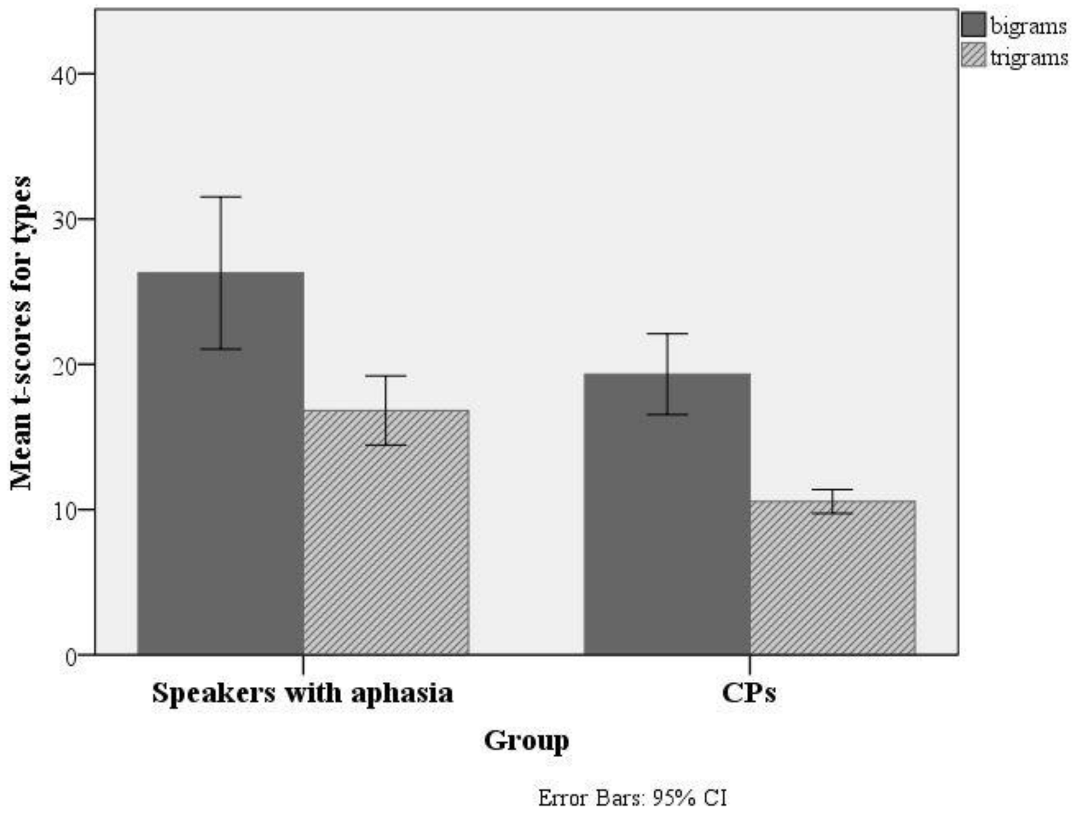


Figure 3.1. Average type-based bi- and trigram t-scores for speakers with aphasia and CPs.

At the level of trigrams, speakers with aphasia produced 469 types (range: 11-172) and 698 tokens (range: 14-196). The CP group produced 5705 trigram types (range: 210-1751) and 6644 tokens (range: 236-2160). The effects observed at the bigram level were stronger for trigrams. For trigram types, the aphasic group displayed a higher average t-score of 17 (SD = 3) compared to 11 in the CP group (SD = 1). The difference was significant, $t(9.864) = 5.74$, $p < .001$, $r = .88$ (large effect). The mean token-based trigram t-score was 39 (SD = 29; median = 26) for the aphasic group, and 14 (SD = 1.9; median = 14) for the CP group. The difference was significant, $U = 8.00$, $z = -2.87$, $p = .004$, $r = -.68$ (large effect), indicating more strongly collocated word combinations in speakers with aphasia compared to neurotypical speakers. However, there was high inter-subject variability in the number of trigrams produced and average trigram t-scores in the aphasic group, with one speaker (GL) presenting with average token- and type-based trigram t-scores that were in the normative range (see Appendix A.1).

The trigram types most frequently used by the nine speakers with aphasia are shown in Table 3.3. The "*don't know*" construction is a phrase available to seven out of nine speakers with aphasia (exceptions: DC, GL) and no other trigram was as widely used. In comparison, the three most frequently used trigrams in the CP group were: "*I don't*" (72 tokens, 1.08% of all CP tokens, used by all nine speakers), "*don't know*" (67 tokens, 1.01% of all CP tokens, used by all nine speakers), and "*do you want*" (26 tokens, 0.39% of all CP tokens, used by six speakers).

Table 3.3: Most frequently produced trigram types and tokens, everyday conversational data (speakers with aphasia)

Trigram type	**Total tokens produced**	**Proportion out of all trigrams**	**Trigram t-score**	**BL**	**DM**	**DC**	**GL**	**JH**	**KK**	**PM**	**PG**	**SC** (8 samples)
don't know	77	9.5%	95.01	21	3	0	0	21	17	1	1	13
I don't	72	8.8%	136.32	22	0	1	0	22	15	1	1	10
going to do	11	1.4%	26.94	0	0	0	0	0	0	11	0	0
one two three	10	1.2%	30.19	0	0	0	4	1	2	0	2	1
two three four	7	0.9%	26.44	0	0	0	3	1	1	0	2	0
it's alright	7	0.9%	27.64	0	0	0	0	0	0	0	0	7
that's it	6	0.7%	45.86	0	0	2	0	0	4	0	0	0
wait a minute	5	0.6%	16.67	0	0	0	2	0	3	0	0	0
n't know what	5	0.6%	39.33	0	0	0	0	0	5	0	0	0

3.3.2. Do frequency-based values vary across talk contexts in aphasia (dataset 1 vs. dataset 2)?

The average combination ratio of the AphasiaBank group was .25 (SD = .11), compared to .17 (SD = .06) in the conversational data, indicating that speakers in the AphasiaBank group were more successful in combining words into multi-word utterances. This group difference was significant, t(45) = -2.066, p = .045, r = .29 (small effect) which could stem either from the different conversational setting (elicited monologue in semi-structured interviews versus dyadic talk) or from varying aphasia severity across the two speaker groups.

AphasiaBank participants produced 1955 different bigram types and 4743 tokens. However, there was a high inter-individual variation: Participant #37 only produced 9 tokens, while participant #10 produced 535 tokens. Average type-based bigram t-scores in the AphasiaBank group (M = 31, SD = 13) did not differ significantly from the naturalistic data (M = 26, SD = 7), t(24.186) = -1.66, p = .109, *ns*, r = .32 (medium effect). The average token-based bigram t-scores in semi-structured interviews was 48 (SD = 26), compared to 53 (SD = 23) in conversational data. Again, a comparison of bigram t-scores revealed no significant difference, t(46) = .464, p = .645, *ns*, r = .07 (small effect). This indicates that the frequency-based characteristics with regard to collocation strength in the two settings are similar.

With regard to trigrams, the AphasiaBank group produced 1708 different trigram types and 2596 tokens. One participant (#37), however, did not produce any three-word combinations. A comparison of type-based trigram t-scores in the two speaker groups (AphasiaBank: M = 22, SD = 20, median = 16; naturalistic data: M = 17, SD = 3, median = 16) was not significant, U = 159.00, z = -.324, p = .746, *ns*, r = -.05 (small effect). The token-based trigram t-scores did not differ across datasets (everyday conversations: M = 39, SD = 29; semi-structured interviews: M = 30, SD = 27), U = 128, z = -1.16, p = .255, *ns*, r = -.17 (small effect). Again, the t-score measures appear robust across contexts, and t-scores from constrained elicitation conditions show good ecological validity.

3.3.3. Interim summary

The frequency-based analysis showed that speakers with aphasia, in comparison to their neurotypical CPs, employed significantly fewer word combinations. Furthermore, higher trigram t-scores (both type- and token based) in the aphasic group indicated that the words within combinations were more strongly associated. Use of more strongly collocated trigrams in speakers with aphasia appears stable across conversational settings. Combination ratio was lower in conversational samples than in semi-structured interviews. This might be due to more severe aphasic impairments in the dyadic group or the influence of probe questions designed to elicit extended monologue in the semi-structured interviews.

Despite high inter-subject variability with regard to specific constructions used in conversations, there was a small common subset across the nine speakers with aphasia including bigrams such as "*I know*", "*no no*", "*it's*", and the trigram "*don't know*". Moreover, the AphasiaBank samples included a further 89 "*don't know*" tokens. "*don't know*" was also produced at least once by each speaker in the CP group. The second analysis addresses conversational functions of these "*I don't know*" (IDK) expressions.

3.4. The functions of IDK

The word combinations produced by speakers with aphasia are more strongly collocated compared to those of their CPs. Analysis 2 investigates whether or not familiar collocations were used in a functionally typical way, based on one construction available to most speakers in the aphasic group: IDK. Typically, IDK or its reduced variant "*I dunno*", is associated with an inability to supply information (Hesson & Pichler, 2016). It often occurs in reply to a question, and is sometimes followed by a complement, as in "*I don't know* <u>*where he went*</u>". However, use of the phrase is not restricted to this prototypical meaning (e.g., Diani, 2004; Pekarek Doehler, 2016; Tsui, 1991). IDK can also function to avoid commenting, disagreement, or commitment to an answer. Other functions are marking uncertainty (hedging), and minimising compliments (Grant, 2010).

There is currently no systematic investigation of the use and functions of IDK in aphasia. Hesson & Pichler (2016) analysed IDK use by speakers with dementia during cognitive assessment, and showed that severity of cognitive impairment was positively associated with the prototypical use of IDK. Mikesell (2009) explored its use in a single case study of an individual with frontotemporal dementia. She found that IDK often functioned as an appropriate answer to *wh*-questions, but sometimes occurred as a strategy to withdraw from a conversation and in other situations reflected memory difficulties. In the current analysis, the primary research question was: What are IDK usage patterns in speakers with aphasia when engaged in dyadic conversation, and do these patterns differ between aphasic and CP groups?

The analysis was based on dataset 1. A higher proportion of isolated IDK tokens in aphasia was expected, but no other predictions regarding similarities and differences in the two speaker groups were made. Hence, the following analyses are exploratory.

3.5. Analysis 2

IDK instances used by speakers with aphasia and CPs in dyads 1-9 were identified. Instances of "*don't know*" or "*dunno*", co-occurring with an explicit or implicit first-person pronoun "*I*", were included in the analysis. CP data were used as a normative sample of IDK usage. Video clips were extracted of each IDK example and one to two turns before and after the token to allow coding of conversational function.[5] IDK tokens were analysed separately for each speaker group.

All IDK tokens were coded for phonetic form, syntactic variation and conversational function using a rating system adapted from Pichler & Hesson (2016), Hildebrand-Edgar (2016) and Diani (2004). Four phonetic forms, the full "*don't know*" and the reduced "*dunno*", both with and without pronoun, were distinguished. For

[5] Unless the token occurred at the beginning or end of a sample or was followed by a long pause.

syntactic variations, there were five categories: isolated, IDK with *wh*-word, complement, co-occurrence with a discourse marker (e.g., "*well*", "*so*"), and 'other' to capture constructions that could not be assigned to any of these categories. Conversational functions were coded using five main categories: lack of knowledge (LOK), interpersonal (INT), turn-constructional (TC), multifunctional (M; a combination of any of the mentioned functions) and 'unclear function' (U) which was added for instances where there was not enough context or evidence to assign a category. Appendix A.2 provides an overview and examples of the function rating system. For INT and TC, sub-categories were assigned to enable documentation of more specific functions such as avoiding commenting (sub-function of INT), or yielding the conversational floor (sub-function of TC; see Appendix A.2). However, the five main categories were used to quantify the distribution of conversational functions.

Following the steps in Pichler & Hesson (2016), rater 1 (R1; the researcher) coded the video clips for conversational functions of all IDK tokens, while a second rater (R2; a PhD student with a strong background in phonetics) independently coded all tokens for phonetic form and syntactic variation. Inter-rater reliability (IRR) was determined based on a random selection of a subset of 23 CP tokens (37% of data) and 29 tokens from speakers with aphasia (38% of data). These were coded by the other rater, i.e., R1 coded these for form and syntax, and R2 coded these for function.

IRR was established on data from both groups. During the process of calculating IRR, tokens where one rater assigned a multifunctional code whereas the other rater assigned one function, were considered a match if the assigned single function was one of the functions subsumed under the multifunctional rating. Furthermore, raters agreed a revised definition of 'hedging' during the IRR process (see Appendix A.2 for more details on this). Across both speaker groups, percentage agreements for form (87% for the CP group, 86% for the aphasic group) and syntax (96% for the CP group, 93% for the aphasic group) were higher than for function (70% for the CP group and 72% for the aphasic group). While IRR for form and syntax was higher than for function, all IRR figures were within an acceptable range, considering the IDK rating scheme was a novel instrument applied to aphasic discourse in dyadic exchanges (Kopenhaver Haidet, Tate, Divirgilio-Thomas, Kolanowski, & Happ, 2009).

The Fisher's exact test was performed with SPSS to determine whether there was an association between speaker group and distributions of IDK forms, syntactic variations and functions.[6] To report effect size, Cramer's V was employed (Field, 2009).[7] Adjusted residuals (z-scores) with a cut-off of +/-2 were used to identify which cells deviated from average values.

To illustrate functional patterns of IDK usage, representative examples of IDK functions were selected and analysed further following an interactional linguistic approach (Couper-Kuhlen & Selting, 2001), using existing conversation analytic (CA) transcriptions from Carragher et al. (2013) and Best et al. (2016). Appendix A.3 includes information on CA transcription symbols used in these extracts.

3.6. Results

The following sections characterize the frequencies of IDK phonetic forms, syntactic variations and conversational functions separately for the two speaker groups.

3.6.1. What are IDK usage patterns in speakers with aphasia when engaged in dyadic conversation, and do these patterns differ between aphasic and CP groups?

The dataset yielded 62 CP IDK tokens and 77 IDK tokens produced by speakers with aphasia. All nine CPs contributed at least one IDK token (range: 1-9), whereas seven out of nine speakers in the aphasic group produced between one and 22 IDK

[6] Due to frequencies below 5 in some cells, the chi-square test could not be applied. In contingency tables larger than 2x2, SPSS uses the Fisher-Freeman-Halton Exact Test (see: http://www-01.ibm.com/support/docview.wss?uid=swg21479647, last accessed on 06/09/2018).

[7] Conventions for Cramer's V with df = 3: .06 = small effect, .17 = medium effect, .29 = large effect; Cramer's V with df = 4: .05 = small effect, .15 = medium effect, .25 = large effect; Goss-Sampson, 2018.

tokens. Figure 3.2 shows the distribution patterns of IDK phonetic forms. The Fisher-exact test showed that the proportions of IDK forms differed across the two groups (p = .032, V = .250, medium effect). Inspecting the adjusted residuals of individual cells, this association was driven by the "*don't know*" (adjusted residuals: +/-2.8) and "*I don't know*" (adjusted residuals: +/-2.1) categories. Frequencies for "*don't know*" were considerably lower in the aphasic group (5% compared to 21% in the CP group) than what would be expected if counts were independent of speaker group. Frequencies in the "*I don't know*" category were considerably higher in the aphasic group (77% compared to 60% in the CP group). The least frequent form in the CP group was "*dunno*" (7%), which was also rarely produced by the aphasic group (8%). "*I dunno*" accounted for 13% of CP tokens and 10% of tokens produced by speakers with aphasia.

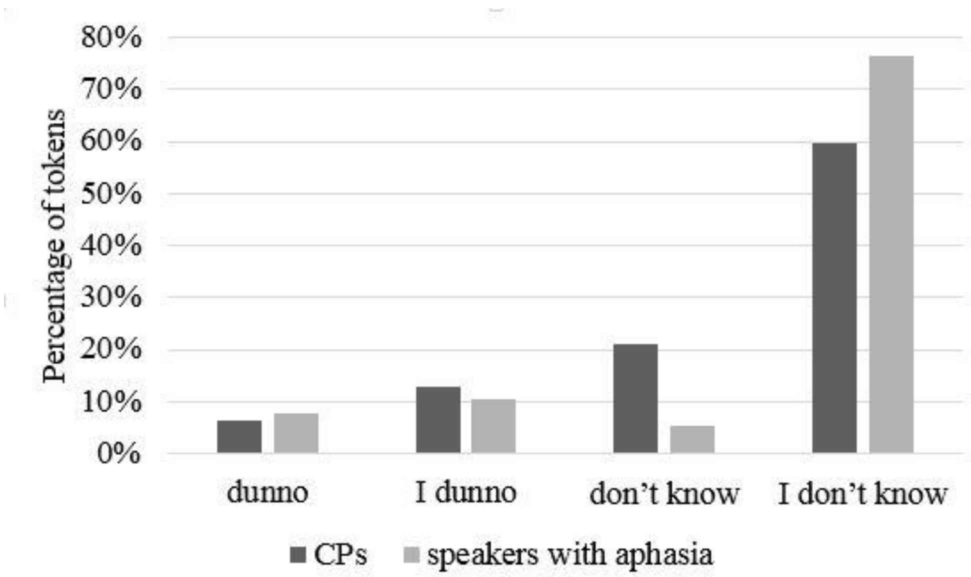


Figure 3.2. IDK phonetic forms in CPs and speakers with aphasia.

Figure 3.3 shows the distribution of syntactic variations. In order to calculate the Fisher's exact test, IDK tokens needed to be assigned to mutually exclusive categories. This was not the case when including counts for IDK + discourse marker, as these tokens could have co-occurred with other syntactic variations. Hence, this category was excluded from the analysis. Based on the remaining four syntactic variations, the distribution differed significantly between the two speaker groups (p < .001, Fisher's exact test, V = .574, large effect). Adjusted residuals indicated that this effect was largely driven by IDK tokens with complement (adjusted residuals: +/-6.1), but also by isolated IDK tokens and 'other' (adjusted residuals: +/-3.6 and +/-3.7, respectively).

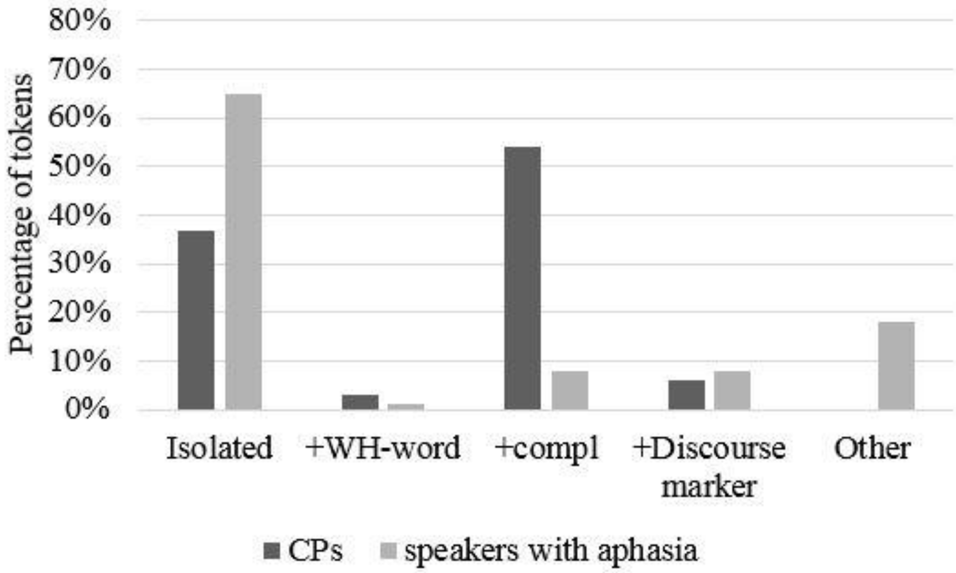


Figure 3.3. Syntactic variations of IDK tokens in CPs and speakers with aphasia.

In the CP group, more than half of the tokens (35 out of 65; 54%) were IDK with complement, whereas speakers with aphasia rarely added a complement (6 out of 77; 8%), reflecting their language difficulties. Isolated IDK tokens were more common in the aphasic group (50/77 tokens; 65%) compared to the CP group (24/65 tokens; 37%). One speaker with aphasia, BL, made use of an unusual syntactic variation captured via the syntactic category 'other' (18% of all tokens in the aphasic group): 14 out of 21 IDK tokens produced by BL represent the set phrase "*I don't know forget*" or "*I don't know for*". IDK tokens were accompanied by a discourse marker four times in the CP group (6%) and six times in the aphasic group (8%). In both groups, tokens with a *wh*-word were rarely produced (one token in the aphasic group; 1%, and two tokens in the CP group; 3%).

Appendix A.4 shows an overview of profiles of IDK functions for each individual. Figure 3.4 shows that IDK was used for a variety of functions by speakers with aphasia. However, there was a significant difference between the speaker groups with regard to the overall distribution of functions (p = .031, Fisher's exact test, V = .276, large effect). Inspection of adjusted residuals revealed that this effect was due to the lower number of multifunctional tokens in the aphasic group (4%; adjusted residual: -2.5) compared to the overall average (9%).

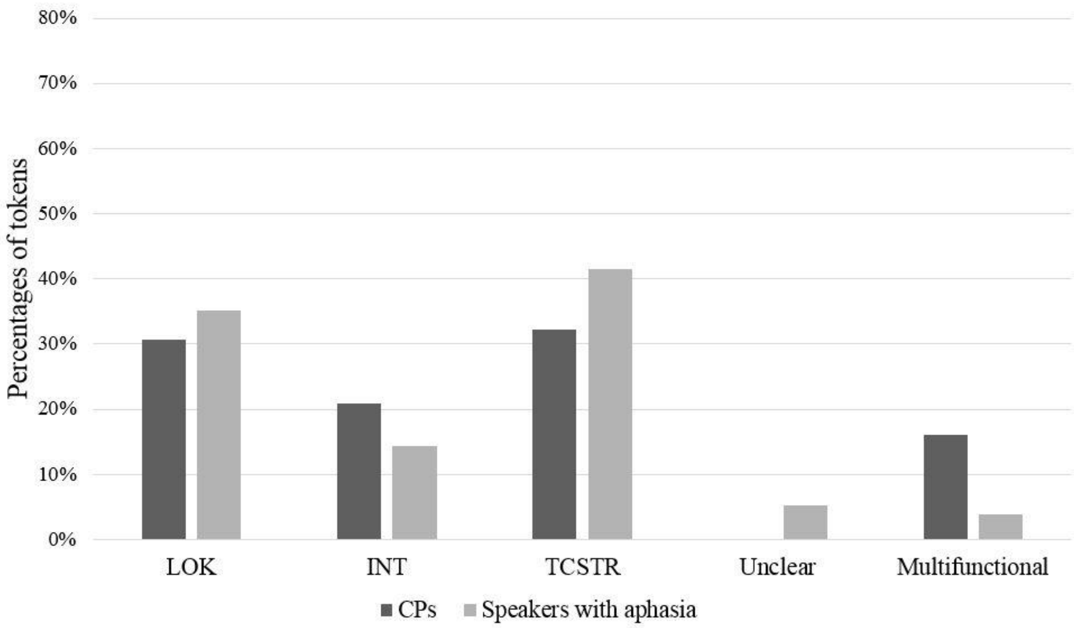


Figure 3.4. Functions of IDK tokens in CPs and speakers with aphasia.

In both speaker groups, IDK frequently had a turn-constructional function (CPs: 32%, speakers with aphasia: 42%) or was used to indicate lack of knowledge (CPs: 31%, speakers with aphasia: 35%). The third most common function in both speaker groups was interpersonal (CPs: 21%, speakers with aphasia: 14%), for example avoiding commenting or hedging. There were a number of unclear IDK tokens in the aphasic group (5%), all produced by one speaker (BL).

A closer look at the relationship between syntactic variation and function revealed differences between speaker groups with regard to the functions of isolated IDK tokens. A large part (46%, 11/24) of CP isolated tokens were turn-constructional, and 13% (3/24) were coded as lack of knowledge. The remaining tokens were coded as either interpersonal (21%, 5/24) or multifunctional (21%, 5/24). In the aphasic group, on the other hand, isolated IDK tokens were used equally to indicate lack of knowledge (37%, 18/50), and to serve a turn-constructional function (37%, 17/50). The remaining isolated tokens were rated as interpersonal (22%, 11/50), multifunctional (4%, 2/50) and unclear (4%, 2/50). The higher percentage of isolated IDK tokens with lack of knowledge function in the aphasic group compared to the CP group might reflect a difficulty with constructing longer conversational turns. It could also be explained by the tendency of CPs to ask speakers with aphasia questions. In the aphasic group, 73% (19/27) of all IDK lack of knowledge tokens followed a CP question (and out of these, 14 were isolated IDK tokens, versus 32% (6/19) of CP tokens). While some examples suggest that these IDK tokens indicate an inability to answer a CP question due to aphasia ("*What's the name of the hotel, can you remember?*" – *"dunno"*, DM_4.2_1_2), others may show a more typical use of IDK, i.e., a genuine lack of knowledge ("*Have they got them in here?*" [referring to item in catalogue] – "*don't know",* SC_2_3). In the CP group, similar lack of knowledge uses were found, for instance when a speaker with aphasia asks his CP "*[...] time is it?*" – "*I don't know*" (PM_1.3_1_2_3).

Within the turn-constructional category, both speaker groups used IDK tokens to hold or to yield the conversational floor. However, these functions were considerably more frequent in the aphasic group, in which these two sub-functions combined accounted for 81% of all TC tokens, compared to only 45% in the CP group. In the aphasic group, holding the conversational floor was often related to word finding difficulties, as in this example about refurbishing a conservatory: "*em em ((sings melody while gestures painting something)) em (0.3) I don't know em blue no em white no brown? brown? yeah.*" (JH_2.3_4). Extract 1 shows an example of a turn-yielding IDK from a speaker with aphasia. Here IDK is combined with a pause during which the speaker with aphasia looks at the CP (line 14) emphasizing the use of IDK to signal speaker change.

Extract 1 (JH_4.3_1)

	01	CP:	Tuesday (.) anticoagulant clinic
	02	PWA:	oh (for) flip sake (1.2) em, em, right so ⌈Northfields em (1.0) ⌉
	03		| ((draws circle on table with | ⌊finger, looks at CP)) ⌋
	04	CP:	mmm, ⌈mmm ⌉
	05	PWA:	⌊ ((looks away from CP)) ⌋
	06		⌈what (will) ⌉
	07		⌊((interrogative gesture with left hand, palm up)) ⌋ >em, em, em<
	08		⌈ (1.0) ⌉
	09		⌊((breathing out, lowers left arm with palm facing down)) ⌋
	10		right, ⌈or there ⌉
	11		⌊((raises left arm with palm facing down)) ⌋
	12		⌈(no) em, how much ⌉
	13		⌊((interrogative gesture with left hand, palm up)) ⌋
→	14		⌈em, em, **I don't know** ⌉ ⌈ (2.3) ⌉
	15		⌊((interrogative gesture with left hand, palm up)) ⌋ ⌊ ((looks at CP)) ⌋
	16	CP:	what, the reading?
	17	PWA:	Yes

CP examples of using IDK to hold and yield the conversational floor include utterances such as "*and if it's not raining it's sports day (1.4) but (0.9) I th- (0.7) I dunno (1.7) if it rains it could be off won't it, NAME'll let us know anyway*" (PM_3.2_8, where *"I dunno"* is followed by a lengthy lapse in the talk where either speaker could take a next turn).

Some speakers in the CP group showed a turn-constructional pattern that was not observed in the aphasic group; that of other-initiated repair (4/20 TC tokens). That is, the CP explicitly stated that he or she had trouble understanding the meaning of the speaker with aphasia's previous turn(s). Examples are "*I don't know what you want*" (CP; KK 2.2_2) or "*well I don't know what you mean love*" (CP; PG_4.2_5_6_7). This function was clearly related to expressive difficulties caused by aphasia.

Interpersonal and multifunctional IDK tokens were found in both speaker groups, however, they were less frequent in the aphasic group (14% and 4%, respectively, compared to 21% and 16% in the CP group). Extract 2 illustrates two IDK tokens with interpersonal functions used by a speaker with aphasia (one of which was coded as multifunctional). The conversation is about a mutual acquaintance who is attending car maintenance classes. The CP, at line 1, brings up the possibility of the speaker with aphasia attending a similar class.

Extract 2 (KK 3.2_1_2_3)

	01	CP:	you could go and do something like that
→	02	PWA:	hmm: no **I don't know** no hahahah
	03	CP:	be too hard cos of your speech?
→	04	PWA:	er: nothing (0.3) an (0.9) **I don't know**
	05		(1.1)
	06	CP:	°hmm° (.) bit more difficult though int °e?°

The first IDK at line 2, softens the speaker with aphasia's disagreement with this suggestion. This interpersonal function is emphasized by the accompanying laughter, a sign of a delicate issue or a dispreferred response (see Tsui, 1991). After the CP wonders whether such a class might be too hard, the speaker with aphasia makes a statement with a turn-final IDK (line 4). This was classified as multifunctional, as it a) serves as a turn-yielding device (the conversation lapses for 1.1 seconds afterwards), and b) signals avoidance of a commitment to the previous answer "*no*" (line 2).

A variety of functions (TC, M, LOK, and INT) were assigned to IDK + discourse marker combinations in both speaker groups. While the number of these items was low (six tokens in aphasic group, four tokens in CP group), "*well*" was the most common co-occurring discourse marker (three in the aphasic group and two in the CP group).

Finally, one speaker with aphasia (BL) made use of atypical IDK constructions, "*I don't know forget*" and "*I don't know for"*, sometimes combined with a gesture. These variations were observed in 67% of BL's IDK tokens (14/21). Two of these were assigned an 'unclear' function, whereas the remaining 11 represented a turn-constructional function, four of which were used in order to take a turn. Only one other speaker with aphasia (KK) used IDK to take a turn, on one occasion. Extract 3 illustrates this idiosyncratic turn-taking function of IDK in BL's talk. Prior to this extract, BL pointed at the camera as the CP was leaving the room. At line 4, the CP sits back down on the couch.

Extract 3 (BL 1.8_8_9_10)

	01	CP:	OOH (0.5) it's <u>cold</u> when you move(d)
	02	PWA:	(ælə?)
	03		⌈ (1.4) ⌉
	04		⌊((CP sits back down)) ⌋
	05	CP:	what?
	06		(0.7)
→	07	PWA:	**<u>I</u> don't <u>know</u> for⌈ <u>get</u>** ⌉
	08		⌊((demonstrative gesture towards camera)) ⌋
	09		(2.5)
→	10	PWA:	⌈**<u>I</u> don't <u>know</u> for<u>get</u>** ⌉
	11		⌊ ((demonstrative gesture towards camera)) ⌋
	12		⌈ (2.5) ⌉
	13		⌊ ((CP folds her arms)) ⌋
	14	PWA:	⌈yes? ⌉
	15		⌊ ((looks at CP)) ⌋
	16		(0.5)
	17	CP:	°yeah what°?
	18		(0.6)
→	19	PWA:	⌈**I don't ↑know** ⌉(2.7) camera, (0.3) shut, (0.9) the, (0.4)⌈door. ⌉
	20		⌊((shakes head))⌋ ⌊((looks at CP)) ⌋
	21		⌈ (1.7) ⌉
	22		⌊ ((CP looks at PWA)) ⌋

BL produces an unintelligible turn at line 2 which is followed by a pause of 1.4 seconds before the CP indicates that she has trouble understanding, asking "*what?*" (line 5). This initiation of repair is followed by a pause and the two IDK tokens of interest here, at lines 7 and 10.

The IDK at line 7, combined with a demonstrative gesture towards the camera, appears to aid BL to take a turn. After a pause of 2.5 seconds, BL repeats his turn with identical intonation and gesture (line 10). Again, the IDK token appears to be a strategy to take a verbal turn, with semantic content added via the demonstrative gesture. It may be a comment about the fact that they were in the middle of making a video recording when the CP left the room. At line 14, after a significant pause, BL checks the CP's understanding with "*yes?*". However, the CP again initiates repair ("*yeah what*?", line 17), after which another turn-constructional IDK can be observed at line 19. This time the IDK token appears to hold the conversational floor, as the turn then continues with a comment about the camera. This reinforces the view that BL's IDK comments at lines 7 and 10 have been about the topic of video recording.

3.6.2. Interim summary

In summary, IDK was common to both speaker groups and available to all but two of the individuals with aphasia. The main difference between the two speaker groups was the proportion of isolated IDK tokens versus IDK tokens with complement. While instances of all function categories were found in neurotypical speakers as well as speakers with aphasia, findings indicate group-specific usage patterns including a higher proportion of turn-constructional IDK tokens in the aphasic group. In addition, IDK was used as a turn-taking resource exclusively by two speakers with aphasia. By contrast, IDK as a resource for initiating repair was found to be a CP-specific function.

3.7. Discussion

This study explored whether residual constructions in aphasic conversation consisted of high frequency, familiar word combinations. Data from nine dyads recorded within everyday conversations were analysed, as well as a larger sample of semi-structured interviews with speakers with Broca's aphasia. Moreover, this study presented usage patterns of IDK in everyday conversations of speakers with Broca's aphasia and their CPs.

3.7.1. Frequency-based analysis

The frequency-based analysis showed that association measures such as t-scores are an effective way of quantifying aphasic language output. They are robust across conversational settings (everyday conversations versus semi-structured interview data) and across individuals with varying degrees of aphasic impairment. Collocation strength as measured by t-scores can be used as an estimate of the degree of formulaicity (Zimmerer et al., 2018), with higher scores indicating a higher likelihood that a word combination represents (part of) a FE. There were more strongly collocated combinations in the aphasic group compared to neurotypical CPs – a finding that is consistent with previous research (Zimmerer et al., 2018), and, if seen as indicators of formulaic language, also consistent with studies using different methods (Van Lancker Sidtis & Postman, 2006). The current study shows that increased reliance on familiar collocations in aphasia also extends to everyday conversational settings. Another contribution is the analysis of both types and tokens as well as bigrams and trigrams. At the level of bigrams, type- but not token-based inventories distinguish the aphasic group from CPs. However, at the level of trigrams, groups were distinguished based on both types and tokens. Effect sizes were large for all type- and token-based comparisons both at the level of bi- and trigrams, and the preserved inventory of common phrases in aphasia appears to be influenced by usage-based factors such as frequency and collocation strength (DeDe, 2013; Knilans & DeDe, 2015). These strongly collocated residual utterances require less combinatorial effort and may be processed as holistic units (Zimmerer et al., 2018), which makes them resilient to aphasia, particularly where combinatorial mechanisms are disrupted.

The stability of frequency-based profiles across different talk contexts suggests ecological validity of elicited speech tasks when taking association measures. Such tasks represent more controlled settings compared to everyday conversations and allow easier comparisons between individuals/groups as there is less variability as to the content/nature of the interaction. Future investigations could use association measures in tasks such as narrative production for comparisons across individuals or speaker groups, or to investigate further the influence of aphasia severity on the reliance on familiar collocations.

It should be noted that frequency of use is a complex variable, intercorrelated with other measures such as AoA (Baayen, Milin, & Ramscar, 2016). Just as AoA may help to explain why some single words remain accessible despite aphasia (Brysbaert & Ellis, 2016), it could be a confounding variable of frequency-related effects at the multiword level. A recent study with neurotypical participants provided evidence of multiword AoA effects on language processing (Arnon, McCauley, & Christiansen, 2017). Future studies addressing the processing of familiar collocations in aphasia might consider AoA as well as association measures.

3.7.2. IDK analysis

Despite this frequency-based difference between the aphasic and CP group, and the variability in the types of word combinations accessible to speakers with aphasia, one shared construction, IDK, was found in the inventories of both speaker groups. IDK stands out in Broca’s aphasia because its syntactic structure is atypical of agrammatic output. However, and in contrast to studies of automatic speech, recurrent or stereotyped phrases that imply this is pathological language (Blanken & Marini, 1997; Rodrigues & Castro-Caldas, 2014), results suggest that familiar collocations such as IDK may be ‘stereotypes’ at a formal, but not at a functional level.

IDK was most commonly realized as the full form “*I don’t know*” in the aphasic and control group. There was a small difference of overall distribution of IDK phonetic forms across the two groups, however, a larger effect size was seen when comparing the group level syntactic variations of IDK tokens. While speakers with aphasia most often used isolated IDK tokens, CPs produced more IDK tokens with a complement. This finding was expected since speakers with aphasia have difficulties in combining smaller linguistic units into longer, grammatically well-formed utterances. At the same time, this suggests that IDK is represented as a relatively fixed unit that may not require grammatical processing (Beeke, Wilkinson, & Maxim, 2007b), a finding that supports the claim that familiar collocations like IDK may be processed as a formula. On the other hand, combinations of IDK and discourse markers such as "*well*" could be observed in the aphasic group with similar frequency to the CP group, indicating that combining IDK with pragmatic elements may be easier than adding lexical elements. However, the overall number of these cases was small.

IDK fulfilled a variety of functions in both CPs and speakers with aphasia. When comparing the frequencies of IDK functions across the two speaker groups, the findings reveal a group difference with a large effect size. This difference stems from the relatively low number of multifunctional IDK tokens in the aphasic group. However, this result needs to be interpreted with caution given that IRR for function was lower than for phonetic form and syntax. The function most commonly associated with IDK, namely indicating a lack of knowledge, was observed in both CPs and speakers with aphasia. However, the number of isolated IDK tokens with a lack of knowledge function was higher in the aphasic group compared to CPs. In the current dataset, question-answer sequences initiated by a CP were common, and this may account for the high number of IDK tokens produced by speakers with aphasia that served to signal lack of knowledge. Both the amount of questioning and the response type reflects the presence of aphasia; CPs use questions to initiate conversation with a person with aphasia, and speakers with aphasia use IDK to provide legitimate answers to such questions whilst signalling the presence of aphasic language difficulties.

IDK was frequently used by speakers with aphasia for turn-constructional functions such as turn yielding, a finding supported by Simmons-Mackie & Damico (1997). Turn-constructional functions were also found in CPs which is in line with previous studies with neurotypical speakers (Pekarek Doehler, 2016). The higher proportion of tokens assigned to turn holding/yielding in the aphasic group is likely to stem from difficulties associated with aphasia. A speaker with aphasia may rely on the use of an island of fluency such as IDK as a resource for opening the conversational floor when grammatical and word finding difficulties make turn construction challenging, or as in extract 3, to regulate turn-taking. The use of familiar collocations to aid turn construction has also been previously reported in Broca's aphasia (Beeke, 2003). Another finding directly related to aphasia was a CP-specific pattern, namely IDK as a method of initiating repair, i.e., signalling a lack of understanding of a prior aphasic turn.

In addition, some speakers with aphasia used IDK for interpersonal functions such as to avoid commenting and, in a small number of cases, IDK turns could even be described as multifunctional (e.g., with mixed turn-constructional and interpersonal functions). Despite the relatively low frequency of such tokens, such multiple functions – associated with typical IDK usage – have not been documented in speakers with aphasia to date. Hence, the present study suggests that IDK is a conversational building block in Broca's aphasia that extends beyond the reported turn-constructional function (Simmons-Mackie & Damico, 1997) and the more fundamental lack of knowledge function. This suggests that such building blocks may be invaluable linguistic structures that could routinely be considered in language assessments and could be harnessed in speech and language therapy interventions (Stahl & Van Lancker Sidtis, 2015). For instance, Helm & Barresi (1980) advocated the use of residual utterances in intervention for severe NFA.

An idiosyncratic function of IDK was observed for one speaker with NFA in this dataset, namely to simply take a turn. For BL, IDK appears to be a resource to contribute to a conversation, where – similar to other functions in the turn constructional category – it regulates turn-taking rather than adding semantic content to a topic. However, in some instances, semantic weight was added to BL's turns constructed using IDK by the addition of a gesture. It should be noted that not every instance of BL's IDK tokens could be assigned to a function. There were some tokens where the function was unclear. Hence, although the present study indicates that IDK is a useful conversational resource for speakers with NFA, it may be difficult to assign a function to some uses of IDK, especially in more severe aphasia.

3.8. Conclusion

The observation of more strongly collocated word combinations in speakers with aphasia supports a usage-based view of language processing. Just as many speakers with aphasia are able to retrieve familiar single words, they can also retrieve familiar multi-word utterances such as IDK.

This study suggests that identifying common word combinations in aphasic talk, with help of a frequency-based approach, is useful to characterize and evaluate the grammatical behaviour of individuals with aphasia. IDK was used in different communicative situations as a relatively fixed, yet effective conversational tool with a functional profile that seems to be adjusted to the turn construction difficulties associated with aphasia.

4. Study 2: The recognition of familiar collocations in aphasia

4.1. Introduction

Psycholinguistic research with adults demonstrates that patterns such as familiar collocations are building blocks that are processed more efficiently compared to less familiar word combinations. Conklin & Schmitt (2012) reviewed studies that investigated the processing of high-frequency phrases and collocations. Results revealed that speakers recognize FEs faster and recall them more accurately relative to matched, less frequent counterparts (section 2.2). This has been found especially for native speakers, indicating that neurotypical adults are sensitive to frequency and predictability information at the phrase level (Siyanova-Chanturia & Martinez, 2015). More recent studies have replicated and expanded the findings reviewed by Conklin & Schmitt (2012). An eye-tracking study by Sonbul (2015), using an off-line task (typicality rating) and on-line reading measures, found collocation effects in both native and non-native speakers of English and in both paradigms. Using eye-tracking, Vilkaite & Schmitt (2017) tested whether non-adjacent verb-noun collocations (e.g., "*provide <u>some of the</u> information*" instead of "*provide information*") elicit facilitatory effects. Native speakers showed a processing advantage for both adjacent and non-adjacent collocations, whereas non-native speakers showed facilitation only for the adjacent collocations. Moreover, Siyanova-Chanturia et al. (2017) demonstrated that binomials (e.g., "*knife and fork*") are processed differently from matched, less frequent but semantically associated noun pairings (e.g., "*spoon and fork*"). Importantly, this difference was only found when the binomials were presented with the conjunction "*and*", but not without it, suggesting that the prefabricated form and language experience determines this effect. Novel evidence of AoA effects on processing times for multiword sequences (early- versus late-acquired trigrams) in adults comes from Arnon et al. (2017). Thus, there is a growing evidence base of the effects of usage-based variables on language processing (both production and comprehension, see Siyanova-Chanturia & Martinez, 2015) across different neurotypical speaker groups.

The reported studies address underlying mechanisms for the processing advantage of high-frequency phrases and collocations, including speeded computation versus holistic processing (e.g., Tremblay et al., 2011), semantic relatedness versus frequency effects (Durrant & Doherty, 2010; Sonbul, 2015), spreading activation between the first and second element of a collocation (Collins & Loftus, 1975; Vilkaite & Schmitt, 2017), and easier semantic integration of familiar phrases as well as pre-activation of a mental template of a phrase (Siyanova-Chanturia et al., 2017). However, the cognitive basis of familiar collocations is yet to be fully determined.[8]

Usage-based grammar theories can serve as a framework to explain these facilitatory effects (Dąbrowska, 2014b; Vilkaite & Schmitt, 2017). Since collocations emerge through repeated use, greater linguistic experience is linked with greater collocational knowledge. Dąbrowska (2014b) designed an instrument, the 'Words that go together' test (WGT), to assess individual speakers' collocational knowledge, using a multiple choice format. Participants were asked to select the most familiar collocation from five similar phrases (e.g., "*odd remark*", "*peculiar remark*", "*queer remark*", "*unnatural remark*", "*weird remark*"). Stimuli were derived from a collocations dictionary, and had an overall frequency of at least 5 in the BNC and a minimum MI-score of 4. Based on a group of 80 adult participants, performance in the WGT was strongly correlated with measures of language experience (education level, vocabulary size, self-reported reading). Unexpectedly, collocational knowledge was found to peak at around 32 years of age, and to decrease from approximately 50. Moreover, Dąbrowska (2014b) correlated performance in the WGT with corpus-derived measures of association strength (including MI- and t-scores), but did not find a significant relationship with any corpus variables. While these findings support usage-based assumptions, i.e., that more linguistic experience links to greater collocational knowledge, the lack of a relationship between association measures and WGT performance raises questions about whether corpus-derived collocations are represented in a speaker's mind.

[8] It should be noted that the reported studies investigated the processing of different types of FEs, including high-frequency phrases, binomials, lexical bundles and collocations. Thus, the stimuli represented a heterogeneous set regarding properties such as length and transparency of meaning.

Despite the interest in the processing of familiar collocations in neurotypical participants, most of the reported studies tested younger adults only. In contrast, little work has been conducted to study the processing of strongly collocated phrases in older adults or participants with aphasia (PWA). Thus, this study seeks to shed light on facilitatory effects of familiar collocations in three participant groups: younger adults, older adults and PWA. Specifically, it will be tested whether sensitivity to collocation strength is seen in recognition.

While familiar collocations have attracted little attention in aphasia research, there is evidence that usage-based variables play an important role in single word- and sentence processing (Brysbaert & Ellis, 2016; DeDe, 2013; Gahl et al., 2003; Huck et al., 2017; Kittredge et al., 2008; Knilans & DeDe, 2015). For instance, Knilans & DeDe (2015) investigated sensitivity to structural frequency during sentence reading. They found that PWA were sensitive to the frequency of syntactic structures, as measured by eye-tracking during reading (more frequent) subject- and (less frequent) object clefts (e.g., "*It was the father that entertained the baby during the party last week*" versus "*It was the baby that the father entertained during the party last week*"). This finding was based on longer reading times at the second noun phrase of object cleft sentences. These results suggest that PWA are sensitive to frequency information during sentence comprehension, and supports previous findings on verb bias in aphasia (e.g., DeDe, 2013, see also Chapter 2). Moreover, Mondini, Jarema, Luzzatti, Burani, & Semenza (2002) explored the processing of compounds, i.e., units that are spelled as two words and thus reflect a collocation with a unitary meaning (e.g., "*red cross*"). The study examined noun-adjective gender agreement in non-compound structures like "*vecchia donna*" (old woman), in comparison to agreement in compounds like "*prima donna*" in two Italian-speaking participants with NFA. Three off-line tasks were administered: a completion task, a reading task and a repetition task, with and without embedding of compound and non-compound structures into a sentential context. Both participants performed better on compounds than on non-compound structures/novel combinations. While the authors suggested that compounds can be accessed as one large unit, this whole-word retrieval could not explain all results. Thus, their model is in line with a processing framework involving two routes, i.e., whole-word- and morphemic constituent access (Mondini, Luzzatti, Saletta, Allamano, & Semenza, 2005). There is some suggestion that PWA, as compared to individuals with

RHD, perform better on the recognition of idioms (e.g., "*while the cat's away the mice will play*") in an off-line, auditory picture-matching task (Van Lancker & Kempler, 1987). Study 1 of this thesis indicated that the inventory of residual bi- and trigrams in speakers with Broca's aphasia may (at least partially) be shaped by usage-based variables such as phrase frequency and collocation strength. These investigations point to an influence of frequency and familiarity of multiword utterances and relatively complex structures (e.g., compounds, idioms) on language production and comprehension in aphasia.

Taken together, there is evidence suggesting that structural frequency influences sentence processing in aphasia and that familiar phrases such as idioms may be processed differently by PWA. However, it is unknown whether listeners with aphasia show a processing speed advantage of strongly collocated word combinations over matched control phrases. To examine the sensitivity to collocation strength on auditory word recognition, the current study developed a word monitoring task (WMT).

The word monitoring paradigm is an example of an on-line task. In contrast to off-line tasks (for instance employed by Mondini et al., 2002, and Van Lancker & Kempler, 1987), on-line experiments enable real-time analysis of language processing (Shapiro, Swinney, & Borsky, 1998). In the word monitoring paradigm, the participant reacts to a pre-specified target word as quickly as possible once it is encountered in a sentential context, by pressing a button (Tyler, Moss, Patterson, & Hodges, 1997; for an overview of the paradigm, see Kilborn & Moss, 1996). This implicit, computerised reaction-time task which taps into real-time processing, allows investigators to study whether participants are sensitive to words embedded in collocations, as compared to the same words appearing in less collocated or familiar structures (e.g., Sosa & MacFarlane, 2002).

A WMT consists of two parts. First, the participant is presented with a target word, followed by a sentential context in which the target word is embedded (e.g., Marinis, 2010; Tyler et al., 1997). The participant is asked to press a response key as soon as the target word is encountered in an auditorily-presented sentence. Thus, precise determination of the target onset is required to ensure accurate measurement of monitoring latencies. Words that start with a plosive, a fricative or an affricate are best suited for accurate determination of onset points in a sentence (Phonetics for Word

Monitoring manual; Brekelmans & Meitanis, 2016, unpublished manuscript). The task itself can include three types of sentences: experimental sentences, filler sentences and catch trials (Marinis, 2010). Experimental sentences include the target words of interest (e.g., "*cake*" → "*They baked a cake on mother's day*"). Filler sentences include distractor items, so that the participant is unaware of the purpose of the research (e.g., "*soon*" → "*Let's go home soon and relax*"). Catch trials are sentences in which a target word does not occur, which means that the participant should not press a button at all (e.g., "*lunch*" → "*He was thrilled to be invited*"), to ensure that participants pay attention to the sentences rather than pressing a button without carefully listening.

Since the WMT can be used with participants with normal (or corrected) sight/hearing, but is focused on spoken language, it is suitable for analysis of auditory processing in a number of populations, including children (Tyler & Marslen-Wilson, 1981), neurotypical adult speakers (Sosa & MacFarlane, 2002), adults with progressive aphasia (Tyler et al., 1997), and adults with non-progressive, stroke-related aphasia (Baum, 1989; Friederici, 1983, 1985; Tyler, 1985; Tyler & Cobb, 1987; Wayland, Berndt, & Sandson, 1996). In the past, WMT studies with PWA focused on word recognition based on different syntactic cues, or the disruption of language processing within ungrammatical sentence contexts.

This study investigated the main research question: Are PWA, as compared to an age-matched control group, sensitive to collocation strength within trigrams? A second objective was to explore whether younger and older adults show a similar degree of sensitivity to collocation strength. Collocations for the WMT were derived from a normative corpus, the spoken BNC. The main prediction was that neurotypical adults (both younger and older) and PWA should show a processing speed advantage for strong collocations when compared to their weaker collocational counterparts. This prediction was based on the presence of frequency effects in other areas of aphasic language processing (e.g., Kittredge et al., 2008), and frequency and familiarity effects at the sentence level during comprehension and recognition in PWA (Knilans & DeDe, 2015; Van Lancker & Kempler, 1987). A third objective of this study was to explore the relationships between on-line processing of collocations and performance on standard clinical off-line tasks such as the Test for Reception of Grammar – Version 2 (TROG-2, Bishop, 2003) and the Boston Naming Test (BNT, E. Kaplan et al., 2001).

In addition to the implicit WMT, an explicit cloze task was devised for use in this study with the aim to quantify a participant's ability to predict the final word of a familiar collocation, based on a constrained semantic context. Cloze tasks require a participant to complete an incomplete sentence. For a detailed explanation of the development of cloze tasks and their relevance to psycholinguistics, see Shaoul, Baayen, & Westbury (2014, p. 4 ff). In the current study, the cloze task mainly acted as a filler task between two word monitoring lists (section 4.2.3.4). In addition, it allowed for comparison of collocation processing in on-line and off-line environments and across input and output modalities. The WMT and cloze task were piloted based on a convenience sample of neurotypical adult participants to explore the feasibility of the protocol and to ensure that the WMT conditions elicited robust effects. This pilot work is outlined in section 4.2. Next, the experiment was made accessible for PWA. Section 4.3 presents the main study, where two group comparisons took place: younger versus older adults, and PWA versus an age-matched control group.

4.2. Pilot study

The main aims of the pilot study were to find out whether:

a) stronger collocations elicit a facilitatory effect as measured by a shorter reaction time (RT) in stronger compared to weaker collocations of trigram pairs;

b) monitoring latencies are correlated to t-score difference of a trigram pair (i.e., whether monitoring latencies vary depending on how strongly associated a trigram is compared to its counterpart);

c) the cloze task elicits acceptable cloze probabilities.

4.2.1. Study design

The pilot study used a within-subject, repeated measures design with one dependent variable (normalized RT to target words) and one independent variable (experimental condition: first- versus second-word manipulations, see section 4.2.3.1). The performance in an explicit task of the production of familiar collocations (cloze task) was also investigated.

4.2.2. Participants

Ten native speakers of English (7 female, 3 male) were recruited via opportunity sampling. Mean age of the participants was 28.7 years (SD = 5.5; age range: 21 to 38 years). All participants gave written informed consent prior to taking part in the study. Approval for this study was given by the UCL Division of Psychology and Language Sciences Research Ethics Committee (LC/2013/05).

4.2.3. Materials

An implicit WMT and an explicit cloze task were developed. In the following sections, the development and structure of each will be outlined.

4.2.3.1. WMT stimuli

Trigrams with a final noun starting with /p/, /b/, /t/, /d/, /k/, /g/, /s/, /z/, /f/, /v/, /tʃ/ or /dʒ/ were derived from the spoken subcorpus of the BNC. All trigrams occurred within the 500 most frequent trigrams ending with a noun and thus would be likely to be familiar to native speakers of English. Next, the trigrams (representing stronger collocations) were paired with weaker collocational counterparts, following the steps outlined in Table 4.1. In sum, 36 trigram pairs were generated. Half of these represented first-word-, the other half second-word manipulations (e.g., "*<u>from</u> this point*" versus "*<u>at</u> this point*"; "*a <u>great</u> deal*" versus "*a <u>fair</u> deal*"). The underlying phrase structure of the stronger and weaker collocations was identical in 33 of 36 trigram pairs. For three pairs, structural changes had to be made (see Table 4.1). Appendix B.1 gives an overview of phrase structures. All stimuli represented grammatically well-formed and semantically plausible phrases. In first-word manipulations, 14 of the 18 critical words represented prepositions (78%), and the remaining were one verb (6%), two determiners (11%) and one noun (6%), while in second-word manipulations, 12 of 18 (67%) critical words represented determiners, one word was a preposition (6%), and in the remaining five pairs (28%), adjectives were manipulated (out of the adjectives, two represented structural exceptions, see Table 4.1, step 3d).

Table 4.1: Steps to derive WMT stimuli

Step	Description	Examples
1)	500 most frequent trigrams (=stronger collocations) with **noun** as a final constituent (i.e., [any word] + [any word] + [noun]) derived from the spoken subcorpus of the BNC	*at the moment* *a long time* *on the table*
2)	Target word starts with a **plosive**, **fricative** or **affricate**	*a long <u>t</u>ime*
3)	Find weaker collocation counterpart by modifying <u>first</u> or <u>second</u> constituent of the trigram	*a <u>long</u> time* vs *a <u>nice</u> time;* *<u>on</u> the table* vs *<u>up</u> the table*
3a)	Raw phrase frequency of counterpart (weaker collocation) below 20 Exception: "*a nice time*" (weaker collocation), raw frequency = 22	
3b)	Stronger and weaker collocation matched for number of syllables of modified constituent	
3c)	**First / Second** constituent matched for unigram frequency category: NB arbitrary frequency categories (1000-10000; >10000)	
3d)	Word class of modified constituent matched Exceptions: • "*a little bit*" — "*a lovely bit*": article + quantifier + noun — article + adjective + noun • "*the other day*" — "*the only day*" article + determiner + noun — article + adjective + noun • "*go to sleep*" — "*have to sleep*" verb + preposition + noun — verb + *to*-infinitive + verb	

Trigram t-scores ranged from 9.3 to 41.4 (median = 12.9) in stronger collocations and 1.4 to 4.6 (median = 2.2) in weaker collocations, representing a highly significant difference, W = 1296, $p < .001$, rank-biserial correlation, $r_B = 1.0$ (large effect).[9] The following variables were controlled across first- and second word manipulations:

[9] For the Mann-Whitney U test, JASP reports Wilcoxon's W and uses rank-biserial correlation as a measure of effect size, with the following conventions: $r_B < 0.1$ (trivial effect), $r_B = 0.1$ (small effect); $r_B = 0.3$ (medium effect); $r_B = 0.5$ (large effect; Goss-Sampson, 2018).

- The t-score differences of the 18 pairs of first-word manipulations (mean trigram t-score difference = 11.62, SD = 4.56) and second-word manipulations (mean trigram t-score difference =14.34, SD = 8.04) did not differ: W = 137.5, p = .448, *ns*, r_B = - .151 (small effect).
- Target word frequencies (based on the spoken BNC) in the two experimental conditions (first- versus second-word manipulations) did not differ significantly, as indicated by a Mann-Whitney U test: W = 113, p = .126, *ns*, r_B = -.302 (medium effect).
- Position of target word in the sentences: Every trial consisted of two sentences, a 'lead-in' or 'scene-setting' sentence (Marslen-Wilson & Tyler, 1981) creating the context, and a test sentence including the trigram of interest. The target word always occurred in the second sentence (in line with Marinis, 2010; Marslen-Wilson & Tyler, 1980), but varied in its position. Considering all words across scene-setting and test sentences (ranging from 11 to 23 words), target words ranged from position 7 (i.e., 7th word) to 17. When comparing first- and second-word manipulations, there was no significant difference in target position (t(34) = -.45, p = .65, *ns*, d = .15, trivial effect). The positions of target words in the test sentences varied from position 3 to 10. There was no significant difference in the position of the target word in the test sentences when comparing first- and second-word manipulations (t(34) = -.53, p = .602, *ns*, d = -.18, trivial effect). Between stronger and weaker collocations (across first- and second-word manipulations), the position of the target word was held constant.

The frequency of the critical word (e.g., "*long*"/"*nice*" in "*a long time*" / "*a nice time*") was controlled. As described in Table 4.1, arbitrary frequency categories were defined. A post-hoc analysis of the log word frequencies of the critical words in the stronger versus weaker collocations using Mann-Whitney U test, however, suggests a significant difference between the two lists in first-word manipulations, W = 226.5, p = .042, r_B = .398 (medium effect), with a mean log frequency of the critical words in stronger collocations of 4.85 (SD = .58) and 4.57 (SD = .63) in weaker collocations. This was likely driven by the highly frequent function word "*in*", which was the critical word in seven of the 18 stronger collocations. In second-word manipulations, the log frequency of critical words did not differ significantly between stronger (M = 4.73, SD

= .79) and weaker ($M = 4.35$, $SD = .69$) collocations, $W = 199.5$, $p = .240$, *ns*, $r_B = .231$ (small effect).

In addition to the 36 trigram pairs, a number of distractor items were added. Binomials such as “*salt and pepper*” or “*fish and chips*” were identified in the spoken BNC. Next, for each binomial ($N = 12$), a less common but semantically related counterpart was created (e.g., “*juice and pepper*” or “*bread and chips*”). The reason for choosing binomials was to shift the attention of the participant to those items, as they are more salient compared to the trigrams in first- and second-word manipulations (i.e., to make participants unaware of the purpose of the experiment). There were insufficient frequency data available for these pairs because many of the less common noun pairings did not occur in the spoken BNC.

In total, the WMT consisted of 48 sentence pairs (96 trials), divided into 18 pairs with first-word-manipulations (37.5% of all trials), 18 pairs with second-word-manipulations (37.5% of all trials) and 12 binomials pairs (25% of all trials). Four representative examples of each condition have been chosen to illustrate the nature of the sentence pairs. These are presented in Table 4.2. A full list of the stimuli can be found in Appendix B.2.

Table 4.2: Examples of sentences in each condition.

Experimental items				**Distractor items**	
First-word-manipulations (n=18 pairs)		**Second-word-manipulations (n=18 pairs)**		**Binomials (n=12 pairs)**	
Stronger collocation	*Weaker collocation*	*Stronger collocation*	*Weaker collocation*	*Binomial*	*Less common noun pairing*
It's a tricky situation. The lorry round the CORNER of the side street was parked there since yesterday.	It's a tricky situation. The lorry near the CORNER of the side street was parked there since yesterday.	We went to the restaurant yesterday. We had a little BIT of fish and some potatoes.	We went to the pub yesterday. We had a lovely BIT of fish and some potatoes.	Peter looked great. He had his best shirt and TIE on.	Andrew looked great. He had his best hat and TIE on.
I'm interested in wildlife photography. There is a great exhibition in the COUNTRY next to Brazil.	I'm interested in ceramic sculptures. There is a great exhibition from the COUNTRY next to Peru.	Your plan looks good. Six weeks is a long TIME for travelling in Italy.	Your plan looks good. Eight weeks is a nice TIME for travelling in Australia.	It's a great place. I love the fish and CHIPS served in this pub.	It's a great place. I love the bread and CHIPS served in this pub.
I think your shopping trip was a success. The chair in the GARDEN is comfortable.	I think your shopping trip was a success. The chair for the GARDEN is beautiful.	I have checked the radiator. I would say there is a hundred PERCENT chance that it works.	I have checked the radiator. I would say there is a twenty PERCENT chance that it works.	I don't like this picture. The bride and GROOM both look tired.	I don't like this painting. The horse and GROOM both look tired.
The new hotel looks massive already. At this POINT, I even think it is higher than the church.	The new building looks massive already. From this POINT, I even think it is higher than the city hall.	We might move to South London. There was a balcony in the first PLACE that we saw yesterday.	We might move to Southampton. There was a balcony in the one PLACE we saw yesterday.	This is a great barman. I want to go and get gin and TONIC for all of us.	This is a great barman. I want to go and get lime and TONIC for all of us.

Target words are highlighted in capitals; trigrams are underlined.

Across the three sentence conditions, the positions of targets were varied to avoid participants developing expectations regarding target position. The position of target words in all 48 pairs ranged from word 3 to word 10 (in the test sentence). Figure 4.1 shows the distribution of the target word positions across all conditions.

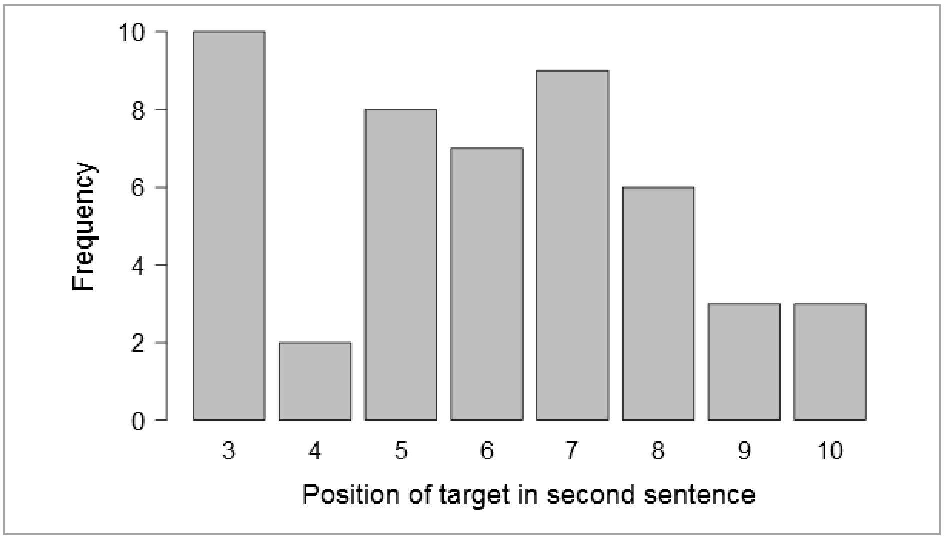


Figure 4.1. Distributions of target word positions in test sentences (i.e., the second sentence in each trial) across all 48 WMT pairs.

The words and sentences were recorded in a sound treated room at UCL, using the program ProRec (version 2.2, Huckvale, 2016; downloaded in June 2016). A female native speaker of English produced the target words and the sentential contexts of both the experimental and the distractor sentences with natural prosody. The subsequent *.wav*-files were separated using the Audio Segmentation tool implemented in ProRec. Files were converted to mono sounds using Praat, and then each sentence was analysed in Praat to detect the target words' onset point within each sentence. Most of the target words (N = 32) started with a plosive, 12 target words started with a voiceless fricative, three with an affricate, and one item started with the voiced fricative [v]. The onset point was defined as the release burst of the initial plosive (Grosjean, 1980) or the start of turbulence (i.e., frication) for the fricatives/affricates. Both the waveform and the spectrogram were used for determining onset points. Figure 4.2 illustrates this procedure for the plosive-initial target "*table*". In a last step, each recording was set to the same intensity (70.0 dB). The principles outlined in the Phonetics for Word Monitoring manual (Brekelmans & Meitanis, 2016) were followed during this procedure.

IRR of onset points was determined with the help of a second rater (R2 from study 1). Ten stimulus pairs (20.8%) were randomly selected (using the RANDOM function in Excel). R2 determined the onset points for the 20 target words, applying the same conventions as rater 1 and blind to rater 1's values. There was an average difference between raters' onset points of 1.51 ms (SD = 1.68, range: -3.10 to 3.06). An independent t-test (two-tailed) revealed a non-significant difference between the onset points determined by the two raters ($p = 0.996$).

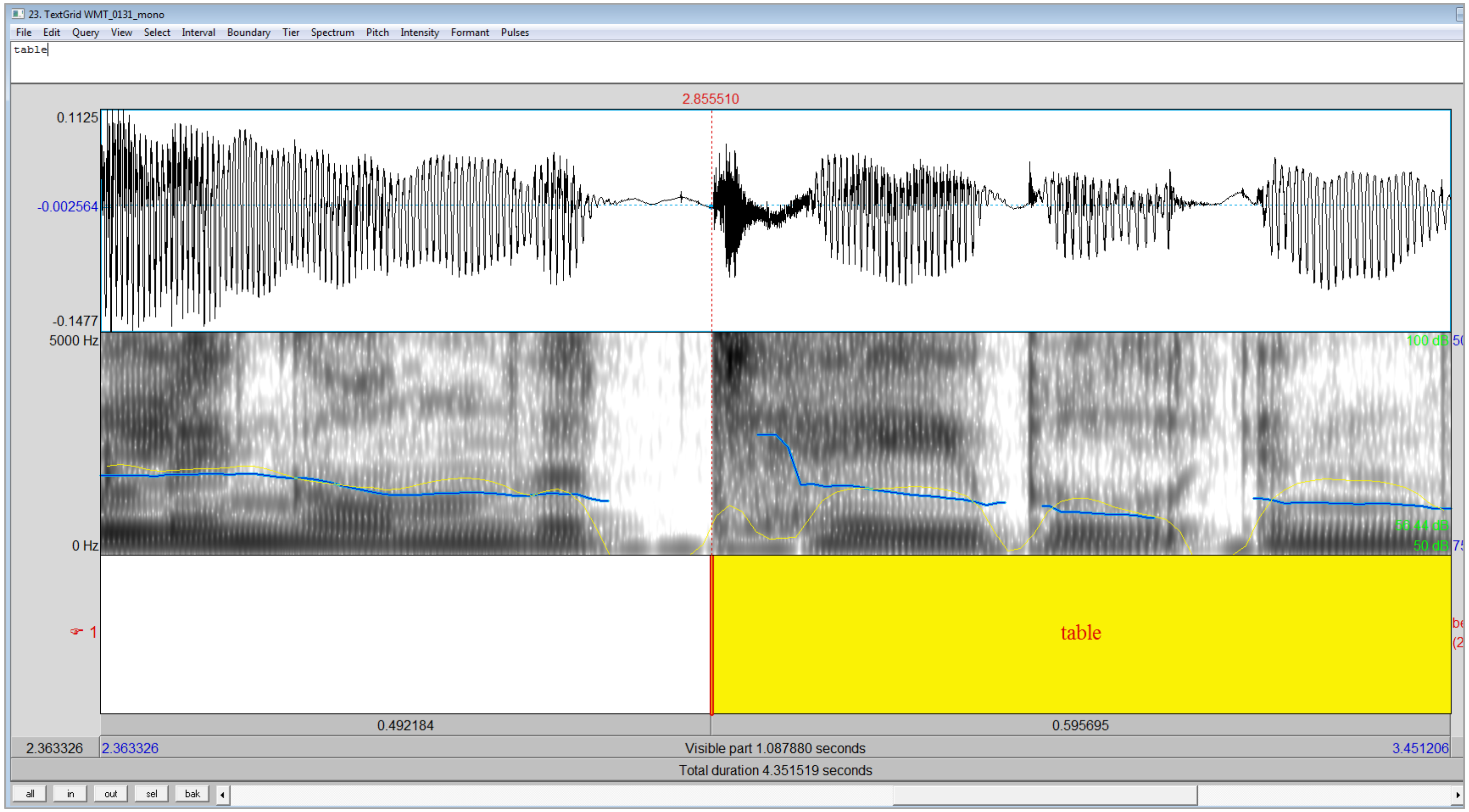


Figure 4.2. Determining the onset point of the target word "table" in: "I've seen a spider. It was crawling on the table in the living room." The red line visualises the onset of the target word.

The 96 stimuli were counterbalanced into two lists. Each list contained the same number of stronger and weaker collocations. The trigram and binomial pairs were allocated to the different lists so that members of a pair never occurred in the same list. Similarly, half of the first- and second-word manipulations, respectively, were allocated to one list and half to the second. Sentences within each list were presented in pseudorandom order. Participants were presented with both stimulus lists, but in counterbalanced order with half hearing List 1 first (participant 1, 3, 5 etc) and half List 2 (participant 2, 4, 6 etc). Before the experimental items in each list, participants were presented with four practice items. These items were taken from Zimmerer et al. (2017) and probed constructions which were not tested in the current study.

4.2.3.2. WMT procedure

The WMT was programmed with DMDX software (version 5.1.3.4, Forster & Forster, 2003). The experiment was run on a Lenovo laptop. Participants were presented with the following written instruction (centred) on the laptop screen:

"*You will first see and hear a word. Keep that word in mind as you listen to the sentences that follow. Click the mouse as soon as you hear the word in the sentence. Be as quick and accurate as possible*."

The written form of the target word, together with its auditory form, was presented to the participant. This was followed by a 2 second blank screen inter-stimulus interval after which the stimulus sentence was presented in auditory form. Participants were asked to use their dominant hand and press a response button (mouse button on laptop keyboard) when the target word was detected. RT was measured in milliseconds from the beginning of the scene setting sentence. After the auditory string, the screen went blank. The inter-trial interval was set at 3 seconds after the button press. Timeout criterion was set at 30 seconds.

After the four practice items, participants were presented with 48 trials. After half of these, the participant could choose to take a self-timed break.

4.2.3.3. Cloze task stimuli

The stimuli for the cloze task were derived from Van Lancker Sidtis & Rallon (2004, p. 22 ff), who extracted FEs from the American screenplay ‘Some Like It Hot’ by Wilder and Diamond (1959; see Wilder & Diamond, 1990). A total of twenty familiar collocations, representative of everyday conversational formulas, were selected. For each of those, a sentential context was generated, in which the final word was omitted. For example, “*Caroline just finished reading a bedtime story to her children. When she leaves their room, she says: Good* ________” *[target: NIGHT]*. More example items are presented in Table 4.3. The full stimulus list can be found in Appendix B.3.

Table 4.3: Example cloze task items

Item number	**Sentences**	**Target word**
1	The train is about to depart. The station master shouts at the last few passengers: “All _________!”	*aboard*
10	Mary asks Hannah what time the concert starts. Hannah says: “I’m not quite _____________.”	*sure*
15	Anna arrives at the birthday party and meets Jane for the first time. She says: “Pleased to meet ___________.”	*you*
17	Carol asks the policeman if she should go left or right. He points to the right and says: “This ___________.”	*way*

The stimuli were recorded under the same conditions as the WMT stimuli (Section 4.2.3.1). A female native speaker of English read out the sentences without the final word, but with prosodic contour indicating continuation.

4.2.3.4. Cloze task procedure

The cloze task served as a filler task to reduce awareness of the paired nature of the WMT. It was designed to elicit a one-word response and aimed to assess the participant's ability to complete familiar collocations, given a constrained and plausible context. After listening to the incomplete sentence, the participant was asked to say the first word that comes to mind that completes the sentence. Appendix B.3 includes the cloze task score sheet. The instructions were:

"*You will now listen to sentences. In the final sentence that you hear, the final word is missing. That means it is incomplete. I would like you to complete this sentence with the first word that comes to your mind that makes the sentence complete and logical. It is important that you only say <u>one</u> word. Are you ready?*"

For an answer that matched the expected target word, the participant received a score of 1. The same score was given for the expected word after self-correction. If the participant responded with an unexpected word or gave an "*I don't know*"- or no answer, a score of 0 was given. Repetitions were given if the participant asked for them, if there was no response after approximately 5 seconds, or if the participant responded with more than one word.

4.2.4. Pilot procedures

Participants were tested individually and, after they were briefed on the purpose of the study, they gave informed consent. Next, one WMT list was presented, followed by the cloze task. Finally, the second WMT list was presented. At the end of the session, the participant was asked what they thought the experiment was evaluating in order to determine degree of explicit awareness of the stimuli within the WMT. The whole procedure took between 25 and 30 minutes.

4.2.5. Pilot data analysis

For the WMT, it was predicted that stronger collocations should decrease monitoring latencies (i.e., speed up the button press) in comparison to weaker collocations. The dependent variable, normalized RT difference (z-score difference) of a trigram pair, should be larger than zero (= facilitatory effect). Both first- and second-word manipulations should elicit a facilitatory effect. Similarly, for distractor items, monitoring latencies should be shorter for binomials compared to their less common counterparts. The participants' answers in the cloze task were predicted to represent the expected target words. Various definitions of acceptable cloze probabilities were found, varying between .4 and .9 (Block & Baldwin, 2010). For the current study, a minimum cloze probability of .8 was regarded as acceptable.

Raw RTs from the DMDX output files and the determined onset points for each target word were copied into an Excel spreadsheet and further processed as illustrated in Table 4.4. Extreme values were removed based on the intra-individual average RT (RTs greater than 2 SDs from the individual mean RT, in line with Baum, 1989), and the RTs were standardised for each subject by transforming raw RTs into z-scores (Table 4.4).

The dependent variable in the subsequent analysis was z-score difference per trigram pair (see Table 4.4 for detail on their calculation). After the z-score differences were generated, further analyses were made with the program JASP (versions 0.7.5.6, 0.8.0.0 and 0.8.1.2; JASP Team, 2017, https://jasp-stats.org/), a statistical program that can be used for traditional, frequentist analyses such as t-tests and ANOVAs, but that was mainly developed for performing Bayesian tests using Bayes Factors (BFs; for an introduction to JASP, see Wagenmakers et al., 2017, and Goss-Sampson, 2018; for an overview of the differences between Bayesian and frequentist statistical procedures, see Dienes, 2011).

A BF describes how likely data are to arise from one compared to another model. As explained in Marsman & Wagenmakers (2017, p. 550), "when the Bayes factor BF_{10} equals 20, the data are 20 times more likely under H_1 than under H_0". A BF_{10} of larger than 3 supports evidence for H_1, BFs_{10} ranging between 1 and 3 can be regarded as anecdotal evidence for H_1, a BF_{10} around 1 points to equal support for both hypotheses (i.e., no evidence), and a BF_{10} of less than 1/3 indicates that there is evidence for H_0 (Wagenmakers et al., 2017).

As JASP is still under development, some features (for example building specific graphs) were not yet available. In these cases, SPSS (version 21 and 22) was used. Facilitatory effects were determined based on t-tests against 0, and a t-test for paired samples was used for comparing first- and second-word manipulations. Effect sizes for t-tests will be reported by using Cohen's d (< 0.2 = trivial, 0.2 = small, 0.5 = medium, 0.8 = large effects; Goss-Sampson, 2018).

Table 4.4: Pre-processing steps for WMT data

E	F	G	H	I	J	K	L	M	N	O
Condition	List	Item	RT_Raw	Onset_time	Onset_time_final	RT_diff	RT_diff_without_early_resp	RT_diff_outl_removed	z-score	z-score difference
1	1	11	3406.89	3.0903	3090.3	316.59	316.59	316.59	-0.341361434	0.537016012
1	2	12	3209.14	2.8326	2832.6	376.54	376.54	376.54	0.195654578	
1	1	21	4605.22	4.2543	4254.3	350.92	350.92	350.92	-0.033842507	-1.172029608
1	2	22	4826.08	4.606	4606	220.08	220.08	220.08	-1.205872115	
1	2	31	3079.07	2.5847	2584.7	494.37	494.37	494.37	1.251144097	0.766692251
1	1	32	3596.76	3.0168	3016.8	579.96	579.96	579.96	2.017836348	
1	2	41	4170.65	3.8711	3871.1	299.55	299.55	299.55	-0.494001181	0.201907271
1	1	42	3580.29	3.2582	3258.2	322.09	322.09	322.09	-0.29209391	

Step	**Column**	**Explanation**
1	H	Raw RT from DMDX output (in milliseconds)
2	J	Onset point as determined in Praat (in milliseconds)
3	K	Raw RT (column H) minus target onset point (column J)
4	L	Remove early responses, i.e., values in column K that are below 0
5	*Calculate average and SD based on column L*	
6	M	Remove outliers, i.e. values that are below and above 2 SDs of individual average, as calculated in step 5
7	*Calculate average and SD based on column M*	
8	N	Calculate z-scores based on raw RT (column M), average RT and SD RT from step 7: $$z = \frac{RTraw - average\ RT}{SD\ RT}$$
9	O	**z-score difference** between weaker and stronger collocation e.g. item 12 (weaker collocation) – item 11 (stronger collocation) → positive z-score difference indicates faster RT in stronger collocation → negative z-score difference indicates faster RT in weaker collocation

4.2.6. Results

The WMT elicited facilitatory effects in each of the three conditions, a result that was in line with the prediction that z-score difference of a trigram pair should be different from zero and positive. A detailed overview of the z-score differences by trigram pair and participant is presented in Appendix B.4. Second-word-manipulations elicited an average z-score difference of .55 (SD = .34), indicating a large effect for those trigrams in which the second unit was modified (one-sample t-test against 0: $t(9) = 5.17$, $p < .001$, $d = 1.63$, large effect). A smaller, but still positive z-score difference was found for first-word-manipulations (M = .26, SD = .18). A one-sample t-test indicates that there was a significant difference from 0 ($t(9) = 4.59$, $p = .001$, $d = 1.45$, large effect). Participants' average z-score differences were analysed to investigate if facilitation was dependent on experimental condition. A paired samples t-test showed that the difference between first- (M = .26; SD = .18) and second-word manipulations (M = .55, SD = .34) was significant, $t(9) = -2.651$, $p = .026$, $d = -.838$ (large effect). In the distractor condition (binomials), the participants showed a z-score difference of .51 (SD = .44). A one-sample t-test (against 0) indicated a large effect of manipulating binomial pairs ($t(9) = 3.67$, $p = .005$, $d = 1.16$).

Further investigation of the facilitatory effect, found in the 36 pairs of first- and second-word manipulations, revealed a positive but non-significant relationship between t-score difference (i.e., the difference between the t-score of the stronger and the weaker collocation) and z-score difference, $r = .06$, $p = .369$, *ns*. A greater t-score difference was related to a slightly higher average z-score difference. However, raw trigram t-score differences were ranging between 7 and 22 and an outlier t-score difference of 40 biased the scatterplot so that it did not resemble a linear relationship. Therefore, t-score differences were ranked (using the RANK function in SPSS), and a correlation was performed with ranked t-score differences and average z-score differences, again with $r = .06$, $p = .368$. Both correlation plots are shown in Appendix B.5. A Bayesian analysis ('Bayesian Correlation Pairs' in JASP) of ranked trigram t-score differences and item-based average z-score differences revealed a BF_{10} of .28 in favour of H_1, reflecting moderate evidence in favour of H_0. A Bayesian robustness check indicates that a Beta prior width varying between 0.5 and 2 is in accordance with moderate evidence in favour of H_0.

While half of the cloze task items (n = 10) elicited a cloze probability of 1, and 7 items had cloze probabilities of ≥ .8, there were 3 items with a cloze probability below .8. These were "*help yourself*", "*get out of here*", and "*that's alright*". For "*get out of here*", only four participants answered with the expected word, and only two produced the expected answer for "*that's alright*". For item #6 ("*help yourself*"), the answers varied between the expected words "*yourself*" / "*yourselves*" (N = 7), and the unexpected word "*me*" (N = 3). Appendix B.6 gives an item overview across the group.

4.2.7. Discussion

A WMT was developed to investigate real-time processing of collocations, and piloted with a small group of neurotypical adults. The results showed that in both experimental conditions, word recognition was facilitated in stronger compared to weaker collocations. This effect was relatively small in first-word manipulations and became significantly stronger for second-word manipulations. A large effect was observed in the distractor condition, binomials. This was an expected finding as these items are more salient compared to the experimental items. Seven of the 10 participants reported at the end of the testing session that they thought the experiment was investigating effects around common noun pairings, which confirmed that binomials drew away participants' attention from first- and second-word manipulations.

While these results suggest that the current WMT is appropriate for use with a larger sample, some methodological refinements were introduced in order to make the task more suitable for PWA. One refinement was to include pictures accompanying the WMT instructions on the computer screen, to make them easier to understand. It was expected that PWA may miss pressing the button more frequently than the participants tested in the pilot study, due to sentence processing difficulties. As a consequence, the duration of the timeout interval was shortened to ensure the experiment did not become too lengthy in the case of missing button presses. A fixation cross during intertrial intervals, and ellipses during auditory presentation of the sentences were added to the WMT in the main study to maintain attention to the screen.

Half of the cloze task items elicited a cloze probability of 1. However, three items (“*help yourself*”, “*it’s alright*”, “*get out of here*”) had a cloze probability below .8. Therefore, a larger sample of normative controls is needed to test whether these patterns are robust and if they are, ways of improving the task or scoring procedure should be developed. The items with relatively low cloze probabilities might indicate that these items represent semi-fixed formulas with more than one expected final word.

4.3. Main study

The main study used the pilot WMT and cloze task materials and procedures in order to answer the following research questions:

- Is the degree of sensitivity to stronger collocations dependent on experimental condition (first- versus second-word manipulations)?
- Are there age effects in the sensitivity to collocation strength (younger versus older adults)?
- Within older adults, is there a relationship between education and sensitivity to collocation strength?
- Do PWA show sensitivity to collocation strength (as compared to an age-matched control sample, i.e., PWA versus older adults)?
- Is sensitivity to collocation strength associated with t-score difference of trigram pairs?
- Can the cloze task patterns found in the pilot study (i.e., acceptable cloze probabilities in at least seven items) be replicated with a larger normative sample?
- For PWA, is sensitivity to collocation strength related to the ability to complete familiar collocations within a cloze task?
- For PWA, is sensitivity to collocation strength related to performance in other off-line tasks?

Based on the pilot study results, it was predicted that facilitatory effects would be stronger in second-word manipulations compared to first-word manipulations. It was predicted that facilitatory effects would be comparable across younger and older neurotypical groups, and across PWA and an age-matched control group. Moreover, there would be a positive correlation between educational level and facilitatory effects. Facilitation would also be positively correlated with t-score differences of trigram pairs. Another prediction was that a larger sample of neurotypical participants should show acceptable cloze probability levels for at least seven out of ten cloze task items. It was also predicted for the aphasic group that there would be a positive correlation between sensitivity to collocation strength in the on-line task and performance on the cloze task. However, given that the task is confounded with output abilities (retrieving / producing words), this effect may be masked by more severe output difficulties for some PWA. No specific predictions about the relationships of facilitation in the WMT and performance in other off-line tasks were made. The null hypotheses tested in the present investigation were:

1) Sensitivity to collocation strength is equal in both experimental conditions across participant groups.

2) Sensitivity to collocation strength is equal across younger and older adults.

3) In older adults, education level is not correlated with sensitivity to collocation strength.

4) Sensitivity to collocation strength is equal across PWA and the age-matched control group.

5) Sensitivity to collocation strength is not correlated with t-score difference of a trigram pair.

6) For PWA, there is no relationship between on-line processing of familiar collocations (sensitivity to collocation strength) and off-line processing of such word combinations (performance in the cloze task).

7) For PWA, there is no relationship between sensitivity to collocation strength and performance in off-line tasks of language processing.

4.3.1. Study design

The present experiment with three groups (younger adults, older adults, PWA) used two mixed designs. The dependent variable was average z-score difference per condition by participant. The first condition by group ANOVA investigated ageing effects and consisted of one within-subjects factor (experimental condition with two levels, i.e., first- versus second-word manipulations) and one between-subjects factor (participant group with two levels, i.e., younger and older adults). Relationships between education and task performance were explored for older adults. In the second condition by group ANOVA, the within-subjects factor stayed the same (condition), but the between-subjects factor ‘group’ consisted of PWA and older controls. Furthermore, relationships between on-line sensitivity to collocation strength and off-line language processing were further investigated for the aphasic group.

4.3.2. Participants

Neurotypical younger adults (age range: 18-30 years) were recruited with help of the UCL participant management system SONA. The study was advertised as a 60-minute session consisting of several tasks that involved listening to and saying words. Participants received course credits in exchange for their participation. Older adults (40 years and older) who had not had a stroke or any other neurological illness were recruited from lab-internal volunteers list, an advert placed with the London chapter of the University of the Third Age, and SONA. PWA were recruited from a university research register. The participants in the older and aphasic groups were paid an hourly rate of £7.50. All participants gave informed consent to participate in the study. Ethical approval was given under the same project ID as the pilot study (LC/2013/05).

In this way, 24 adults were recruited into the younger control group. Out of these, two were excluded as they were not mono- or bilingual speakers of English, or English was not the main language in childhood. The remaining 22 participants (18 female, 4 male) were between 18 and 30 years old, with an average age of 20.6 (SD = 3.1). Six were bilingual and the remaining 16 participants were monolingual speakers of English. All participants were undergraduate or postgraduate students at UCL (see Table 4.5 for demographic information).

Table 4.5: Younger control group (university students): participant characteristics

ID	Mono- / bilingual*	Gender	Age
YC01	B	f	20
YC03	M	m	23
YC04	B	m	23
YC05	M	f	18
YC06	M	f	21
YC07	M	f	26
YC08	M	f	18
YC09	B	m	20
YC10	B	f	19
YC11	M	f	19
YC12	M	f	20
YC13	M	f	25
YC14	M	f	18
YC15	M	m	19
YC16	B	f	18
YC17	M	f	18
YC18	M	f	19
YC19	B	f	20
YC20	M	f	21
YC21	M	f	19
YC22	M	f	30
YC23	M	f	18

** Classified based on a pre-screen check via email*

A further 24 participants were recruited into the older control group. One participant was excluded because of a (self-reported) diagnosis of Parkinson's disease. The remaining 23 participants (13 female, 10 male) were between 41 and 78 years old, with an average age of 62.7 years (SD = 10.4). One was bilingual (English and Polish), and 22 were monolingual speakers of English. The majority indicated their highest educational level to be 'graduate studies' (n = 14; 61%), followed by 'postgraduate studies' (n = 7; 30%) and 'sixth form' (n = 2; 9%). Participants reported either no hearing difficulties, or difficulties corrected with a hearing aid (see Table 4.6 for demographic information).

Table 4.6: Older control group: participant characteristics

ID	Mono- / bilingual	Gender	Age	Education		Handedness
				Highest level	Years	
OC1	M	m	72	Graduate studies	11	right
OC2	M	f	66	Postgraduate studies	19	right
OC3	M	m	62	Graduate studies	16	right
OC4	M	f	72	Graduate studies	14	right
OC5	M	f	64	Graduate studies	17	right
OC6	M	f	68	Sixth Form	12	right
OC7	M	f	75	Graduate studies	16	right
OC8	M	f	68	Graduate studies	15	right
OC10	M	f	77	Graduate studies	14	right
OC12	M	f	78	Graduate studies	13	right
OC13	M	f	70	Graduate studies	14	right
OC14	M	m	74	Postgraduate studies	20	right
OC16	M	f	62	Sixth Form	13	right
OC17	M	f	66	Graduate studies	16	left
OC18	B	f	62	Postgraduate studies	17	right
OC19	M	f	53	Graduate studies	10	right
OC20	M	m	41	Postgraduate studies	18	right
OC21	M	m	56	Postgraduate studies	17	right
OC22	M	m	48	Graduate studies	13	right
OC23	M	m	45	Postgraduate studies	18	right
OC24	M	m	53	Graduate studies	18	right
OC25	M	m	52	Postgraduate studies	21	left
OC26	M	m	58	Graduate studies	13	right

Finally, 25 participants (22 male, 3 female) were recruited for the aphasic group. One participant presented with severe comprehension difficulties and, after an attempt at the practice items of the WMT and consultation with his wife, it was decided that he was unsuitable for the study. The main reasons – apart from comprehension difficulties – were signs of frustration with the task and the technology. Another participant completed both testing sessions but the WMT data were unusable as he consistently pressed the button after the end of the final sentence instead of after the target (average RT: 2747 ms after target onset), indicating that the data did not reflect automatic processing. Furthermore, during the task, he seemed increasingly frustrated, so that the researcher decided to only present one WMT list. For a third participant, a medical procedure resulted in non-completion of the full protocol.

The remaining 22 participants (19 male, 3 female) were between 37 and 84 years old, with an average age of 60.3 (SD = 10.1). Table 4.7 summarizes demographic information for the aphasic group. There were 20 monolingual speakers of English and two multilingual speakers. All participants reported English as their main language. All participants presented with chronic, stroke-related aphasia (mean time post onset in months: 73.4, SD = 40.9, range: 16-162). The group was mixed in terms of their underlying expressive and receptive language difficulties (Section 4.3.6.8). While the aphasic and older control group were age-matched, the distribution of education levels differed in the two groups. In the aphasic group, there were 11 participants (50%) reporting ‘graduate studies’ as their highest educational level, followed by ‘secondary school’ (n = 5; 23%), ‘sixth form’ (n = 3; 14%) and ‘postgraduate studies’ (n = 3; 14%), indicating that the older control group was biased towards participants with higher educational levels. All but one participant reported either no hearing difficulties, or difficulties corrected with a hearing aid. One participant (A24) reported that he was planning to get his hearing checked.[10]

[10] In the A.D.A. test on auditory processing (see section 4.3.3.2), this participant had a score of 39 correct judgements (out of 40) which reflects a performance above average.

Table 4.7: Aphasic group: participant characteristics

ID	Mono- / multilingual	Gender	Age	TPO (months)	Education (Highest level)	Handedness pre / post stroke
A1	M	m	51	48	Graduate studies	R – R
A2	M	m	54	45	Graduate studies	R – R
A3	M	m	76	104	Sixth form	R – L
A4	M	m	63	76	Secondary school	R – R
A5	M	m	57	79	Graduate studies	R – R
A6	M	m	37	43	Graduate studies	R – L
A7	M	m	51	87	Graduate studies	R – L
A8	M	m	64	126	Graduate studies	R – R
A10	M	f	48	38	Secondary school	R – L
A11	M	m	61	16	Secondary school	R – R
A13	M	m	67	54	Sixth form	R – L
A14	M	m	55	49	Sixth form	R – L
A15	Multi	m	84	156	Graduate studies	R – R
A16	M	m	67	78	Graduate studies	R – R
A17	Multi	f	58	98	Postgraduate studies	R – L
A18	M	m	68	75	Postgraduate studies	R – L
A19	M	m	59	120	Postgraduate studies	R – L
A20	M	f	71	162	Graduate studies	R – R
A21	M	m	57	57	Graduate studies	R – R
A22	M	m	56	19	Graduate studies	L – L
A23	M	m	55	23	Secondary school	L – L
A24	M	m	67	62	Secondary school	R – R

4.3.3. Materials

4.3.3.1. WMT

The same stimuli as in the pilot WMT were used. Participants reacted to the target words by clicking the left mouse button on the laptop keyboard. Several modifications were made. The instructions were made accessible for PWA through use of shorter sentences and pictures/diagrams (Figure 4.3).

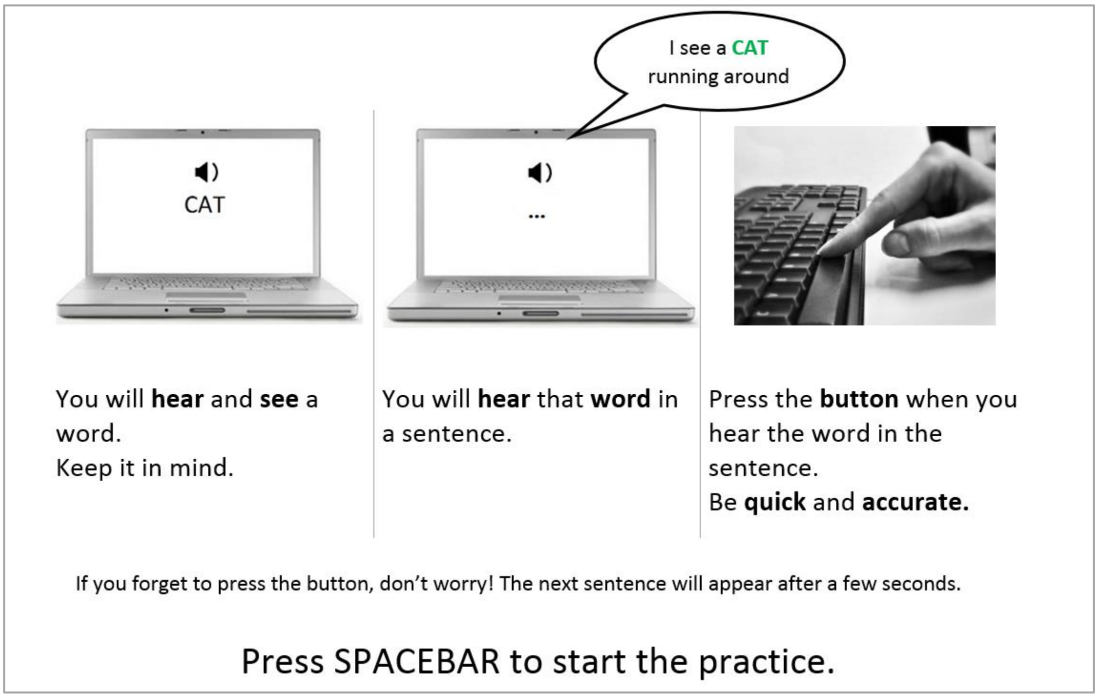


Figure 4.3. Aphasia accessible instructions for the WMT (shown at the beginning of each list).

Every trial started with a fixation cross ("+"), presented for 3000 ms, followed by the target word (visual and auditory presentation). After a 2000 ms blank screen interval, the sentential context was played, accompanied by ellipses on the screen ("…"). The fixation cross and ellipses were added to support participants' attention to the task. The task structure is represented in Figure 4.4. The timeout was set to 6000 ms to decrease the delay between the current and the next trial in the event of no response. Raw RTs were processed as in the pilot experiment. Every participant was tested individually on a Lenovo laptop. Due to technical difficulties, a different laptop (Acer) was used after the first five participants of the younger control group.

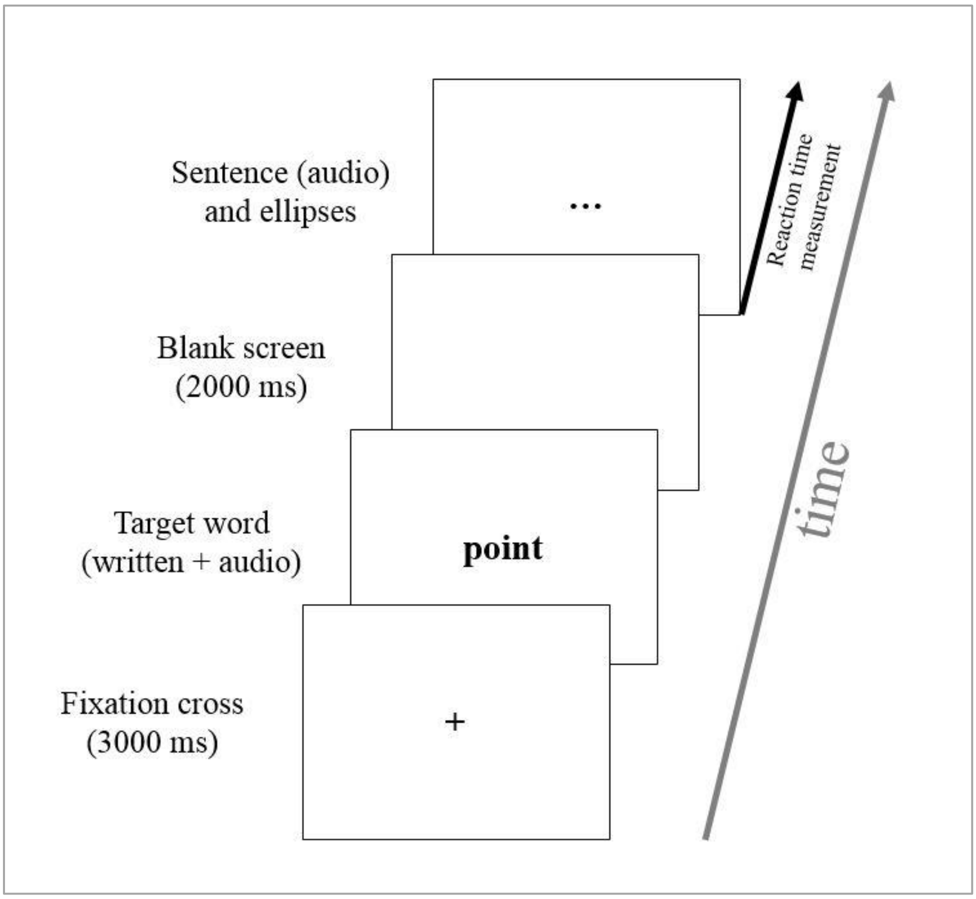


Figure 4.4. Illustration of a WMT trial.

4.3.3.2. Filler tasks

The pilot study cloze task was used in the present experiment as a filler task between WMT list 1 and 2. For the aphasic group, the playback speed of the sentences was decreased to 0.92, after the second PWA commented that the sentences were too rapid. The scoring procedure was revised from that of the pilot study (Section 4.3.6.6).

An additional filler task was included between the two WMT lists. This was a subtest from the Action for Dysphasic Adults (A.D.A.) Auditory Comprehension Battery (Franklin, Turner, & Ellis, 1992), an auditory processing assessment. In this subtest (P3) called ‘real word minimal pairs’, the participant listens to pairs of one-syllable words and decides whether these are the same or different. In addition to the verbal instructions, a visual response sheet was designed to support PWA (Figure 4.5).

For every correct response, a score of 1 was given. The performance in this task (i.e., number of correct judgements out of 40) was used as an indicator of auditory processing ability.

Figure 4.5. Visual aid used in A.D.A. minimal pairs subtest

4.3.3.3. Background assessments

PWA and older adults were profiled on a range of language and cognitive assessments. These will be outlined in the following sub-sections. Picture-based narrative samples were also collected, but they are not analysed in the current report.

4.3.3.3.1. Lexical retrieval

The BNT (E. Kaplan et al., 2001) is an indicator of a participant's lexical retrieval ability. The final score represents the number of correct responses (out of 60). The starting point for older neurotypical adults and PWA was the first picture.[11] The test was otherwise administered via standard instructions (if an answer reflected a

[11] Although the BNT manual suggests to start with picture item 30 for neurotypical adults and to continue forward (unless the participant misnames a picture before item 38), it was decided to use the same procedure across groups.

misperception of a picture, the researcher could give a semantic cue; testing was discontinued after 8 consecutive failures).

4.3.3.3.2. Phonological working memory

Subtest 13 from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Coltheart, & Lesser, 1992) is a digit span task administered by recognition, testing phonological working memory span. Participants were asked to judge whether two lists of digits were presented in the same order or not (i.e., yes / no decision). The number of digits ranged from two to seven. For the present study, PWA could choose to use a visual response sheet to point at 'same' or 'different' instead of giving a verbal response (Figure 4.6). A stepwise scoring procedure was applied. The final digit span score reflected the number of digits for which the participant made the highest number of correct yes-no-choices. More detailed scoring procedures, including an example score sheet of a control participant, is shown in Appendix B.7.

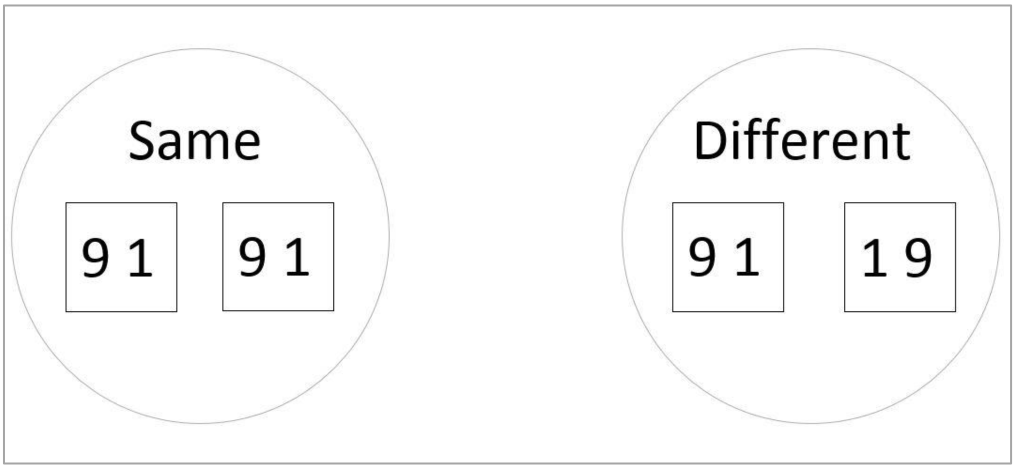


Figure 4.6. Visual aid designed for PALPA-13.

4.3.3.3.3. Comprehension of spoken words

A single word-to-picture matching subtest from the Comprehensive Aphasia Test (CAT; Swinburn, Porter, & Howard, 2004) was chosen to assess comprehension of spoken words. The maximum score was 30. The standard scoring procedure was applied. This test was only administered for participants in the aphasic group.

4.3.3.3.4. Comprehension of spoken sentences

The TROG-2 (Bishop, 2003), a receptive language test to identify the degree of language comprehension difficulties at the sentence level, was also part of the background assessment battery. Twenty different blocks (each consisting of four items) of grammatical constructions were assessed. Examples include ‘Reversible in and on’ (“*The cup is in the box*”), ‘Not only X but also Y’ (“*The pencil is not only long but also red*”), and ‘X but not Y’ (“*The cup but not the fork is red*”). In this study, the raw number of correct blocks constituted the final score.

4.3.3.3.5. Nonverbal cognitive abilities

The Matrix Reasoning (MR) subscale from the Wechsler Abbreviated Scale of Intelligence – Second Edition (WASI-II; Wechsler, 2011) was chosen as a nonverbal assessment of cognitive abilities. This subtest, consisting of 30 items, requires the participant to choose an image from a set of options that completes an incomplete matrix. For control participants, the two sample items were administered first, followed by item 4. For PWA, item 1 was administered after the sample items. The raw score (sum of correct items) can be transformed to age-corrected T-scores for control participants. However, since this procedure was not appropriate for PWA, raw scores were used for both groups.

4.3.4. Procedures

Participants gave informed consent to taking part in the project (Appendices B.8-B.11). Aphasia-friendly versions of information sheets and consent forms (Appendices B.12-B.13) were created by using a freely available resource with images from the NIHR website.[12]

The session protocol for each group is presented in Table 4.8. For some of the older control participants and PWA, some or all background profile scores (e.g., BNT, TROG-2) were available from a university research register. Therefore, testing took one single session.

[12] http://www.nihr.ac.uk/nihr-in-your-area/stroke/aphasia.htm (last accessed on 31/05/2018)

Table 4.8: Session protocol

Group	Session 1 – WMT Session (~60 minutes)	Session 2 – Profiling (~60-70 minutes)
Younger controls	**1)** Information sheet, consent form, questionnaire **2)** WMT list 1 **3)** A.D.A. (P3) (Franklin et al., 1992) **4)** Cloze task **5)** WMT list 2 **6)** 'Dinner Party' or 'Jogging' Narrative (Fletcher & Birt, 1983)	**N/A**
Older controls	**1)** Information sheet, consent form, questionnaire	**1)** Boston Naming Test (BNT) (E. Kaplan et al., 2001)
	2) WMT list 1	**2)** 'Dinner Party' Narrative
	3) A.D.A. (P3)	**3)** Test for Reception of Grammar (TROG-2) (Bishop, 2003)
	4) Cloze task	**4)** Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) 13 (digit span by recognition) (Kay et al., 1992)
	5) WMT list 2	**5)** Wechsler Abbreviated Scale of Intelligence – Second Edition (WASI-II), Matrix Reasoning (MR) (Wechsler, 2011)
PWA	**1)** Information sheet, consent form, questionnaire	**1)** BNT
	2) WMT list 1	**2)** Comprehensive Aphasia Test (CAT), comprehension of spoken words (Swinburn et al., 2004)
	3) A.D.A. (P3)	**3)** 'Dinner Party' Narrative
	4) Cloze task	**4)** PALPA 13 (recognition)
	5) WMT list 2	**5)** TROG-2
		6) WASI-II MR

4.3.5. Data analysis

Data were pre-processed in Microsoft Excel according to the procedures outlined in Section 4.2.5. Sensitivity to collocation strength was examined after monitoring latencies were transformed into normalized RT (z-score) differences per trigram pair. Facilitatory effects were determined by subtracting normalized RT to the weaker collocation by normalized RT to the stronger collocation.

Average z-score differences by participant and condition were the basis for subsequent statistical analyses completed with the program JASP (versions 0.7.5.6, 0.8.0.0 and 0.8.1.2), as well as SPSS (Section 4.2.5). Group comparisons are reported based on two 2x2 mixed factorial ANOVAs (repeated measures cells = first-word- and second-word manipulations; between-subject factors = group), using omega squared (ω^2) as a measure of effect size (conventions: .01 = small; .06 = medium; .14 = large effect; Kirk, 1996, as cited in Field, 2009, p. 390), as it provides an unbiased effect size measure for samples smaller than n = 30 (Goss-Sampson, 2018). The same analyses were also performed with a Bayesian model comparison approach, a Bayesian Repeated Measures ANOVA, to investigate how likely data were to fit to H_1 (alternative hypothesis) compared to H_0 (null hypothesis), as indicated by BFs (BF_{10}). To determine whether the distractor condition (binomials) elicited an effect, independent samples t-tests were performed, with Cohen's d as a measure of effect size (for conventions of small / medium / large effects, see section 4.2.5). For correlational analyses, (frequentist) Pearson correlations as well as Bayesian Correlation Pairs are reported.

4.3.6. Results

This section presents an analysis of missing WMT trials, outliers and average RTs (4.3.6.1), followed by the results of two group comparisons (2x2 ANOVAs and Bayesian Repeated Measures ANOVA equivalents, sub-sections 4.3.6.2 and 4.3.6.4). For older adults, the relationship between educational level and sensitivity to collocation strength will be explored in 4.3.6.3. An analysis of the relationships between facilitatory effects and t-score differences will be presented (4.3.6.5) as well as performance in the cloze task (4.3.6.6). Finally, 4.3.6.7 and 4.3.6.8 are concerned with relationships between sensitivity to collocation strength and performance in the cloze task and other off-line tasks for the aphasic group only.

4.3.6.1. WMT missing trials, outliers, and average RTs

Missing trials were classified as early responses (i.e., pressing the button before target onset) or timeouts (i.e., not pressing the button). The raw number of missing trials by condition and group is shown in Table 4.9. The highest number of timeouts was observed in the aphasic group. In younger adults, the number of timeout responses (sum = 10) was lower compared to early responses (sum = 49). Younger adults presented the highest number of early responses, followed by PWA (sum = 30) and older adults (sum = 14).

Table 4.9: Types of missing trials by group and condition

Condition	Type of missing trials	Younger adults	Older adults	PWA
First-word manipulations	Timeout	1	0	32
	Early response	13	5	8
Second-word manipulations	Timeout	7	0	54
	Early response	16	4	19
Binomials	Timeout	2	0	8
	Early response	20	5	3

A further 254 trials were lost across the three groups through removing outliers (i.e., more than 2 SDs below or above the individual mean RT difference). An overview of outliers by condition is shown in Table 4.10.

Table 4.10: Number of outlier trials removed

Condition	Younger adults	Older controls	PWA
First-word manipulations	19	23	26
Second-word manipulations	45	46	46
Binomials	14	18	17
Sum	***78***	***87***	***89***

One PWA (A22) was classified with word meaning deafness according to a Speech and Language Therapist working in the UCL Department of Language & Cognition. He had a large number of timeouts (N = 20 trials) and his overall rate of 'usable' pairs was lower than 2 SDs from the group mean. This participant was excluded from subsequent analyses.[13] After this exclusion, subsequent analyses were based on 21 participants in the aphasic group. The rates of complete pairs by participant groups were 83% in the aphasic group, 91% in the older control group, and 88% in the younger adults.

The younger control group presented the fastest average RT of 308.8 ms after target onset (SD = 57.92), followed by the average RT in the older control group which was 417.9 ms after target onset (SD = 76.13). The aphasic group showed the slowest average RT of 632.2 ms (SD = 196.02), with two individuals with an average RT above 2 SD of the group average (A4 and A24 with average RTs of 1108 ms and 1074 ms, respectively).

4.3.6.2. Group comparison I: younger vs older adults

Sensitivity to collocation strength was similarly affected by condition in younger and older adults (Figure 4.7). Appendices B.14-B.15 present average z-score differences by participant and condition for the two groups. Out of 22 participants in

[13] A22's proportion of usable pairs was .54 across the three conditions. The group mean was .82 with a SD of .10. Hence, the cut-off was .62.

the younger control group, 19 showed positive average z-score differences (i.e., facilitatory effect) in both experimental conditions. For three participants, negative z-score differences were recorded in first-word manipulations. In the older group, facilitation was observed in 20 of 23 participants (three negative z-score differences were recorded in first-word manipulations, and one in second-word manipulations).

In both groups, z-score differences were greater for second-word-manipulations compared to first-word manipulations, indicating more sensitivity to stronger collocations in which the modified word immediately preceded the target. In second-word manipulations, the average z-score difference was higher in the older control group (M = .59, SD = .34) compared to the younger control group (M = .52, SD = .28). In first-word manipulations, the younger control group showed an average z-score difference of .29 (SD = .28) which was higher compared to the older control group (M = .24, SD = .22).

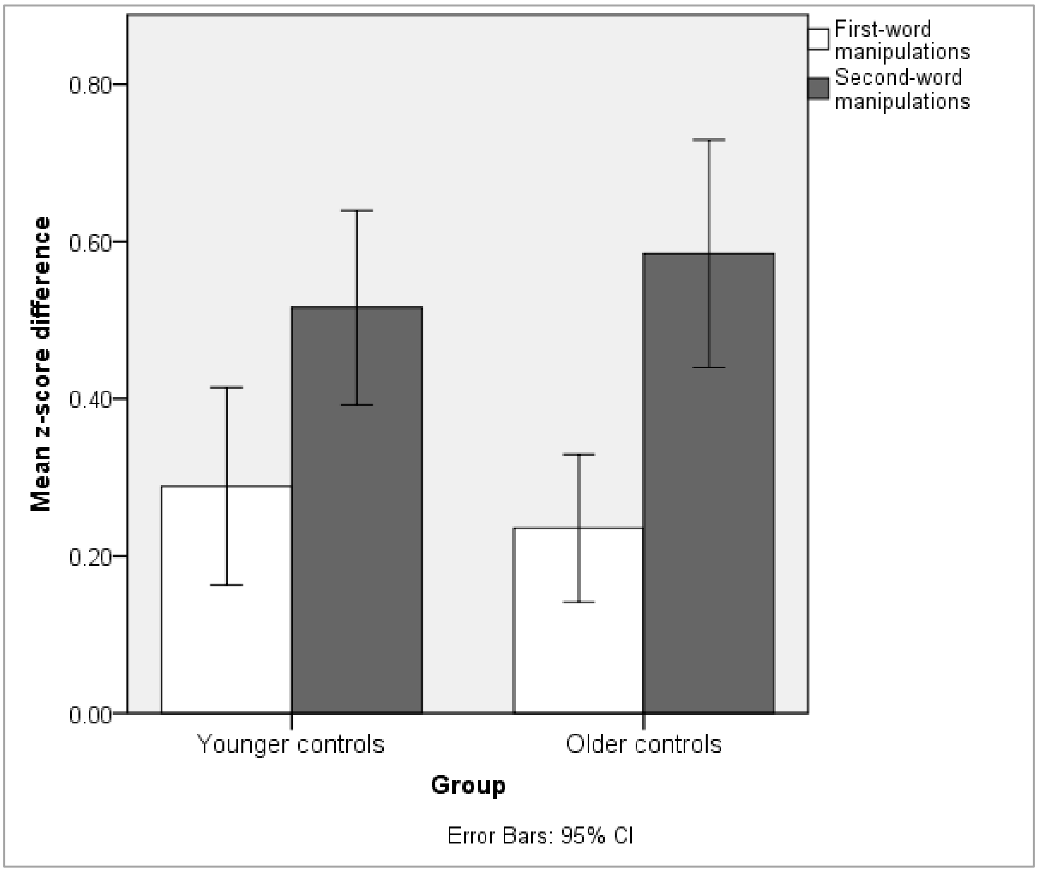


Figure 4.7. Facilitation (positive z-score differences) of younger and older adults in the two experimental conditions.

To statistically investigate whether performance in the two experimental conditions was dependent on condition and group, a mixed factorial ANOVA was carried out. The Shapiro-Wilk test indicated that the assumption of normality had been violated for second-word manipulations in the older control group ($W(23) = .909$, $p = .04$). However, visual inspection of the histogram showed that this was most likely caused by one outlying data point. It was therefore decided to perform the ANOVA despite this data point.[14] There was a significant main effect of condition ($F(1, 43) = 23.62$, $p < .001$, $\omega^2 = .329$, large effect), with second-word manipulations eliciting greater facilitatory effects compared to first-word manipulations (Figure 4.7). There was a non-significant main effect of group on z-score differences ($F(1, 43) = .017$, $p = .896$, *ns*).

The same comparisons were investigated from a Bayesian perspective. A Bayesian Repeated Measures ANOVA, using the default prior (r scale fixed effects = .5), was performed based on average z-score differences in first- and second-word manipulations. The results, shown in Table 4.11, revealed that there was extreme evidence for a model including 'Condition' ($BF_{10} = 18288.91$; compared to a null model), but no evidence for an improvement when adding 'Group' ($BF_{10} = .251$). In order to evaluate the evidence for an interaction, the BF of the interaction was divided by the BF of the model including the main effects only, resulting in .49, indicating that there was no evidence for an interaction, and anecdotal evidence against an interaction. This was also reflected in the inclusion BFs which were weak when 'Group' was involved ($BF_{Inclusion} = .262$ and .410 when 'Group' was involved, versus $BF_{Inclusion} = 13570.998$ for a model with only 'Condition').

[14] After removing this participant (OC22) as an outlier, the data were normally distributed ($W(22) = .982$, $p = .943$). The ANOVA with only 22 older controls revealed the same result: there was a main effect of condition ($F(1, 42) = 27.77$, $p < .001$, $\omega^2 = .369$), and a non-significant main effect of group ($F(1, 42) = .581$, $p = .450$, *ns*).

Table 4.11: Model comparison: Bayesian Repeated Measures ANOVA compared to null model (younger vs older adults)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model (incl. subject)	0.200	3.927e -5	1.571e -4	1.000	
Condition	0.200	0.718	10.191	18288.914	1.192
Group	0.200	9.856e -6	3.942e -5	0.251	0.935
Condition + Group	0.200	0.189	0.932	4810.825	1.922
Condition + Group + Condition * Group	0.200	0.093	0.410	2366.339	2.610

Note. All models include subject.

As described in section 4.2.3.1, there was a significant difference between the log word frequencies of critical words in first-word manipulation pairs. To investigate the impact of frequency on the facilitatory effect found in first-word manipulations, correlational analyses were performed. When correlating item-based average z-score differences with critical word frequency differences, there was no significant relationship between these two variables in older controls (r = -.312, p = .207, *ns*). In younger controls, the correlation revealed a significant negative relationship between these two variables (r = -.51, p = .03), pointing to a counterintuitive relationship for critical word frequency: when the critical word frequencies were more similar (i.e., smaller difference between the stronger and the weaker collocation of a pair), the average z-score difference decreased (i.e., less facilitation). However, a Bayesian correlational analysis indicated a BF_{10} of 2.58, corresponding to anecdotal evidence for H_1.

Facilitatory effects in the distractor condition (binomials) were investigated separately from the two experimental conditions. The younger control group showed a significantly lower average z-score difference (M =.53, SD = .30) than the older control group (M = .84, SD = .40), t(40.34) = -2.970, p = .005, d = .88 (large effect). This points to an influence of ageing on the sensitivity to binomials, with greater sensitivity in older, relative to younger adults.

4.3.6.3. Education-related differences in task performance (older adults only)

The relationship between task performance and level of education was explored for the older control group. Figure 4.8 shows the linear relationships between education and average z-score differences in the two experimental conditions (first- versus second-word manipulations). There was a significant and positive correlation for first-word manipulations, $r = .433$, $p = .020$ (one-tailed, assuming that there would be a positive relationship), but a non-significant relationship between number of years of education and second-word manipulations, $r = -.014$, $p = .525$ (one-tailed, assuming that there would be a positive relationship), suggesting that more educated participants were more sensitive to first-word manipulations. On the other hand, sensitivity to second-word manipulations was not associated with education.

The finding was supported by a Bayesian correlation with a BF_{10} of 3.754 in first-word manipulations. This BF suggest that it was almost four times more likely that the data were true given the alternative hypothesis (= moderate evidence for H_1). Compared to this, the BF_{10} for second-word manipulations was .247, indicating moderate evidence in favour of H_0 (no correlation).

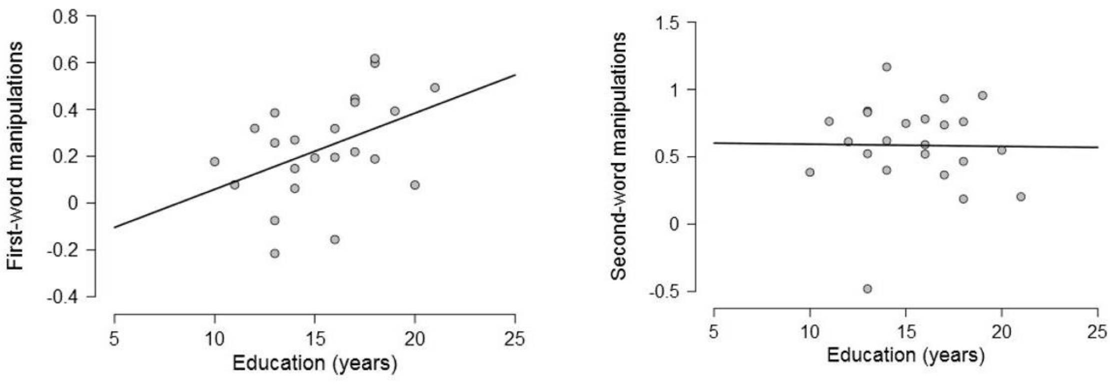


Figure 4.8. Relationships between education and sensitivity to collocation strength in first- and second-word manipulations

4.3.6.4. Group comparison II: PWA vs older controls

Figure 4.9 shows the mean z-score differences by experimental condition and participant group. Appendix B.16 presents average z-score differences by participant and condition for the aphasic group. Out of 21 PWA, there were 14 with positive z-score differences in both experimental conditions, while seven participants presented with negative z-score differences in one or both conditions (4 negative data points in first-word manipulations, 4 data points in second-word manipulations). Similar to the older control group, the z-score differences in the aphasic group were greater for second-word-manipulations ($M = .29$, $SD = .35$) compared to first-word manipulations ($M = .19$, $SD = .24$). Again, a mixed factorial ANOVA was performed. There was a significant main effect of condition ($F(1, 42) = 16.757$, $p < .001$, $\omega^2 = .242$, large effect), with second-word manipulations eliciting greater facilitatory effects compared to first-word manipulations. The effect of 'Group' on z-score differences was significant ($F(1, 42) = 6.181$, $p = .017$, $\omega^2 = .105$, medium effect), as was the interaction between condition and group, $F(1,42) = 5.242$, $p = .027$, $\omega^2 = .065$, medium effect). This indicates that PWA were less sensitive to second-word manipulations compared to older controls.[15]

[15] The same analysis was performed after removing OC22. The ANOVA with 22 older controls revealed the same pattern: There was a significant main effect of condition ($F(1, 41) = 20.806$, $p < .001$, $\omega^2 = .283$), a significant main effect of group ($F(1, 41) = 10.80$, $p = .002$, $\omega^2 = .186$), and a significant interaction effect ($F(1,41) = 7.259$, $p = .010$, $\omega^2 = .089$).

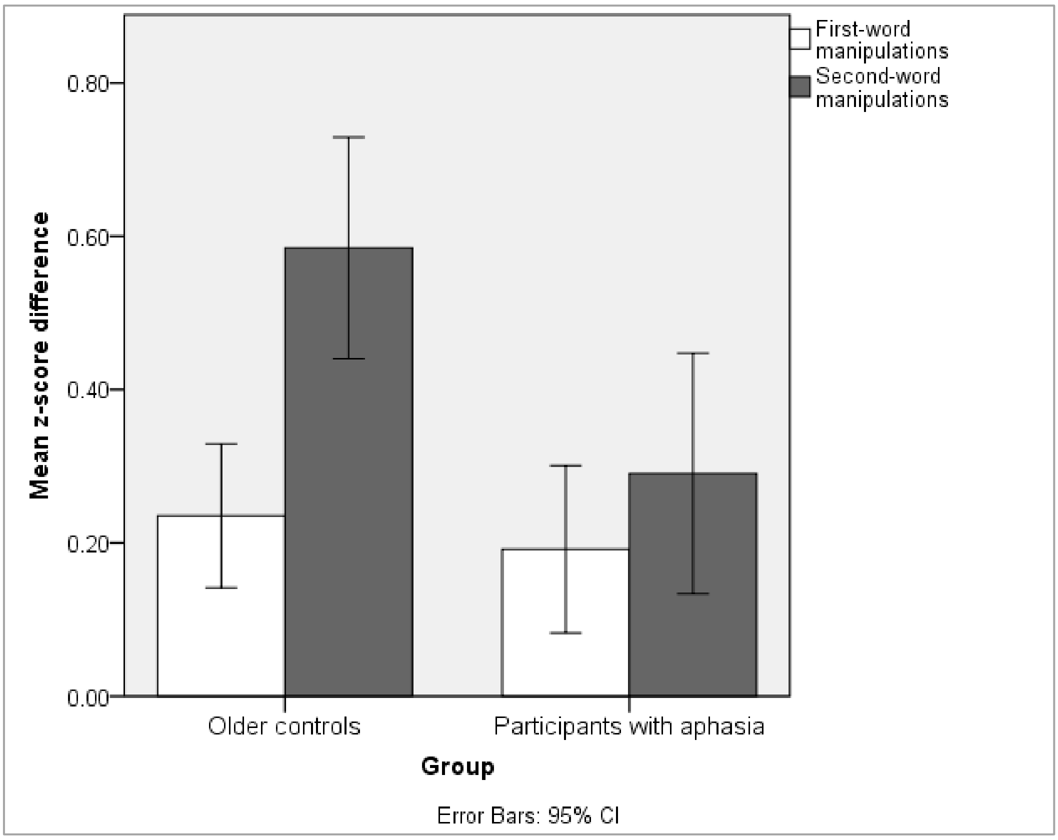


Figure 4.9. Facilitation (z-score differences) of PWA and older controls in the two experimental conditions

A Bayesian Repeated Measures ANOVA supported these conclusions. Using the default prior (r scale fixed effects = .5), results revealed that there was extreme evidence for a model including ‘Condition’ (BF_{10} = 225.585; compared to a null model), and there was improvement when adding ‘Group’ (BF_{10} = 2.026; Table 4.12). In order to evaluate the evidence for an interaction, the BF of the interaction was divided by the BF of the model including the main effects only, resulting in 2.67, indicating that there was evidence for an interaction. This was also supported by the inclusion BFs which were strong when ‘Group’ was involved ($BF_{Inclusion}$ = 6.328 when ‘Group’ was involved and $BF_{Inclusion}$ = 7.686 for the ‘Condition by Group’ interaction).

Table 4.12: Model comparison: Bayesian Repeated Measures ANOVA compared to null model (PWA vs older controls)

Models	**P(M)**	**P(M|data)**	**BF_M**	**BF_{10}**	**error %**
Null model (incl. subject)	0.200	4.206e -4	0.002	1.000	
Condition	0.200	0.095	0.419	225.585	1.365
Group	0.200	8.522e -4	0.003	2.026	1.214
Condition + Group	0.200	0.246	1.306	585.155	4.565
Condition + Group + Condition * Group	0.200	0.658	7.686	1563.710	1.859

Note. All models include subject.

The influence of critical word frequency on the z-score differences in first-word manipulations was investigated. For the aphasic group, there was no significant relationship between mean item z-score difference (across the 21 participants) and the difference in the log frequency of critical words (r = -.34, p = .17, *ns*).

An independent samples t-test for binomials indicated that there was a significant difference between average z-score differences in PWA (M = .52, SD = .52) and older controls (M = .84, SD = .40), t(42) = 2.307, p = .026, d = .696 (medium effect). This suggests that while PWA were sensitive to more common noun pairings, the average z-score differences were smaller compared to an age-matched control group.

4.3.6.5. Relationships between facilitation and t-score differences

Next, the relationships between z-score- and t-score differences were analysed for all three participant groups. The prediction was that with a higher t-score difference between the trigram pairs, the average z-score difference (i.e., the effect of the experimental manipulation) should increase. JASP was used to compile traditional Pearson correlations as well as Bayesian Correlation Pairs, using ranked trigram t-score differences. As the underlying effect of collocation strength was investigated, both experimental conditions, i.e., 36 data points, were included (average z-score differences per item).

Figure 4.10 shows the linear relationship between the two variables in the younger control group. There was a significant and positive correlation (r = .409, p = .007, one-tailed, assuming that the correlation would be positive), indicating that items with greater t-score differences elicited greater z-score differences. Bayesian Correlation pairs revealed a Bayes Factor of $BF_{10} = 7.908$ in favour of the alternative hypothesis, i.e., that there was a positive relationship between these two variables. This indicates that it is nearly 8 times more likely that there was a positive relationship compared to no correlation (H_0). A Bayesian robustness check, presented in Figure 4.11, indicates that a Beta prior width varying between around 0.6 and 2 is in accordance with moderate evidence in favour of H_1.

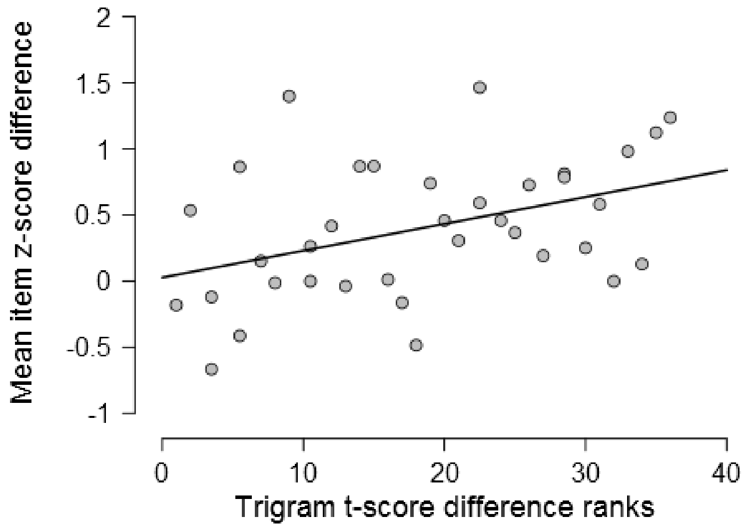


Figure 4.10. Scatterplot of ranked item-based trigram t-score differences and average z-score differences in younger adults.

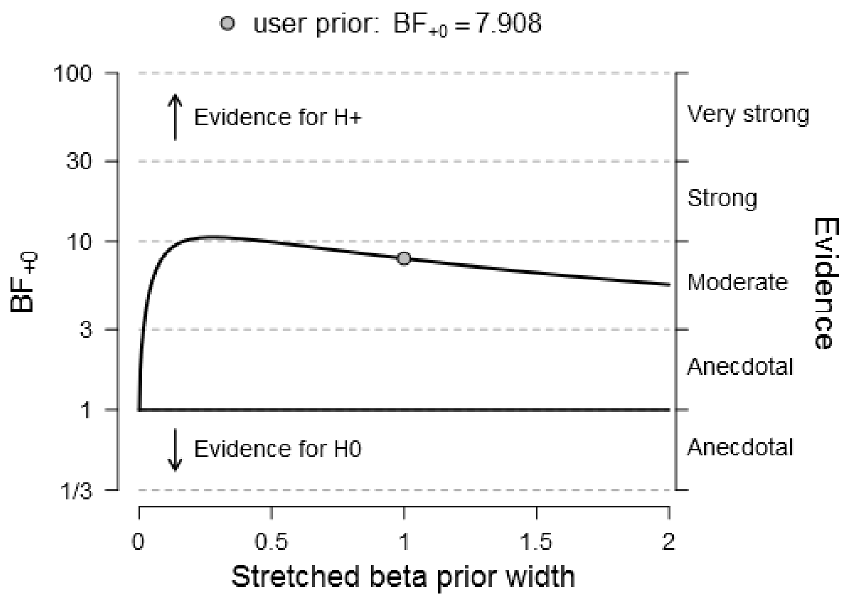


Figure 4.11. Bayes Factor Robustness Check for correlational analysis of ranked item-based trigram t-score differences and average z-score differences in younger adults.

Figure 4.12 shows a scatterplot of the data in the older control group. There was a positive but non-significant relationship between t-score- and z-score difference across the 36 trigram pairs (r = .067, p = .350, *ns*, one-tailed, assuming that the correlation would be positive), showing that items with larger t-score differences did not elicit significantly larger z-score differences. A Bayesian correlation analysis indicated no evidence in favour of H_1 ($BF_{10} = 0.46$), but moderate evidence in favour of H_0, i.e., that there was no correlation ($BF_{01} = 2.18$).

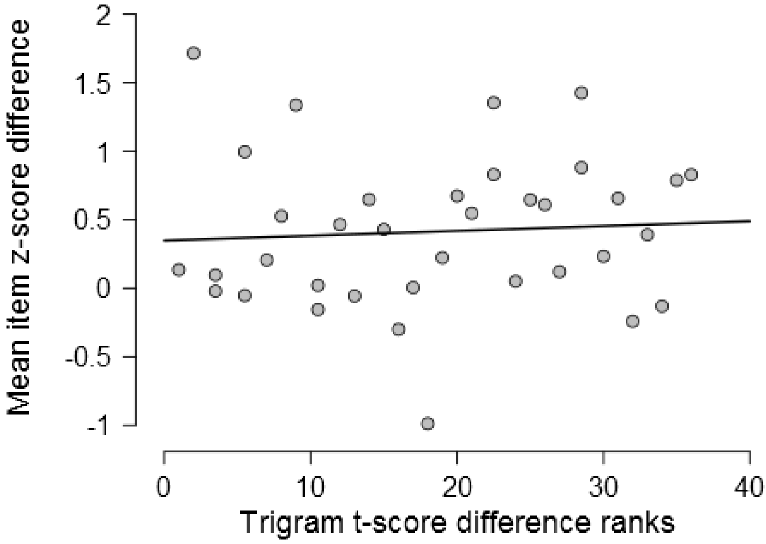


Figure 4.12. Scatterplot of ranked item-based trigram t-score differences and item-based average z-score differences in the older control group.

The relationship between ranked item-based t-score differences and average z-score differences in the aphasic group is presented in Figure 4.13. The pattern was similar to the older control group: there was a positive but non-significant relationship between the two variables ($r = .087$, $p = .307$, *ns*, one-tailed, assuming that the correlation would be positive). A Bayesian correlation analysis showed no evidence in favour of H_1 ($BF_{10} = 0.32$), but moderate evidence in favour of H_0, i.e., that there was no correlation ($BF_{01} = 3.10$).

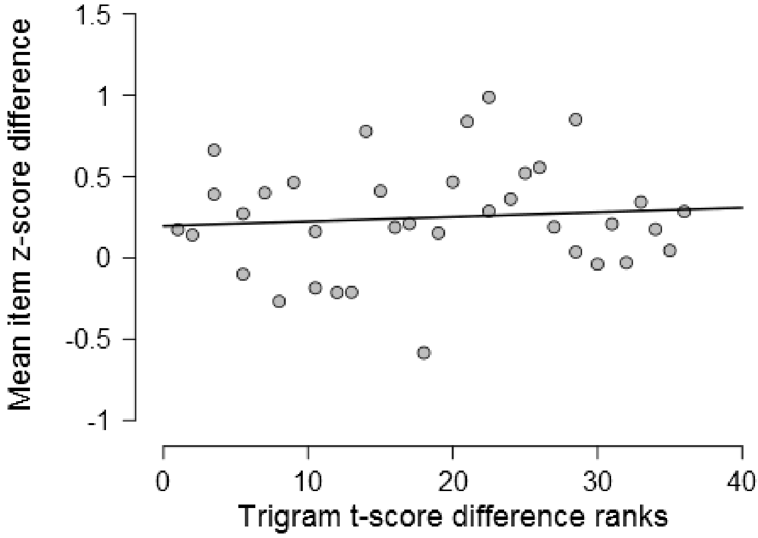


Figure 4.13. Scatterplot of ranked item-based trigram t-score differences and item-based average z-score differences in the aphasic group.

4.3.6.6. Cloze task

Eight out of 20 items elicited a cloze probability of 1 in both the younger and the older control group. The two items with the lowest cloze probabilities of only .41 in the younger and .52 in the older control group were "*get out of here*" and "*that's alright*". The most common alternative answers for these ("*get out of there*" and "*that's okay*", respectively) as well as three other items with a relatively low response accuracy, can be found in Appendix B.17. These patterns are in line with the findings from the pilot study (Section 4.2.6).

There was no significant response accuracy difference across the 20 items between the younger and the older control group, W = 198.0, p = .966, *ns*, r_B = -.01 (small effect). Therefore, joint response accuracies were used in order to further investigate normative patterns (Table 4.13). For three items, joint response accuracy was lower than .8. Therefore, common answers which were synonyms of the expected completions were accepted as correct responses when scoring the aphasic group. The formula "*let's go*" was removed from the scoring for PWA, as the most frequently occurring answer “*let’s hurry*“ was not regarded as a synonym formula for the original phrase “*let’s go*“.

Table 4.13: Joint response accuracies for cloze task items across 45 younger and older controls

Items 1-10	*aboard*	*in*	*worry*	*here*	*night*	*yourself*	*one*	*you*	*know*	*sure*
Joint response accuracy	.87	.87	.87	.47	1.00	.82	.93	1.00	1.00	1.00
Items 11-20	*sorry*	*go*	*you*	*all*	*you*	*alright*	*way*	*about*	*happened*	*mind*
Joint response accuracy	1.00	.76	1.00	.89	1.00	.47	.98	1.00	.93	.98

4.3.6.7. Relationship between facilitation and cloze task performance (PWA only)

After refining the cloze task scoring procedure, the relationship between sensitivity to collocation strength and cloze task performance was investigated in the aphasic group. Cloze scores (number of expected answers out of 19 items) were used as an indicator of off-line processing of familiar collocations. Figure 4.14 shows the distribution of average cloze scores in the aphasic group (range: .16 to .95).

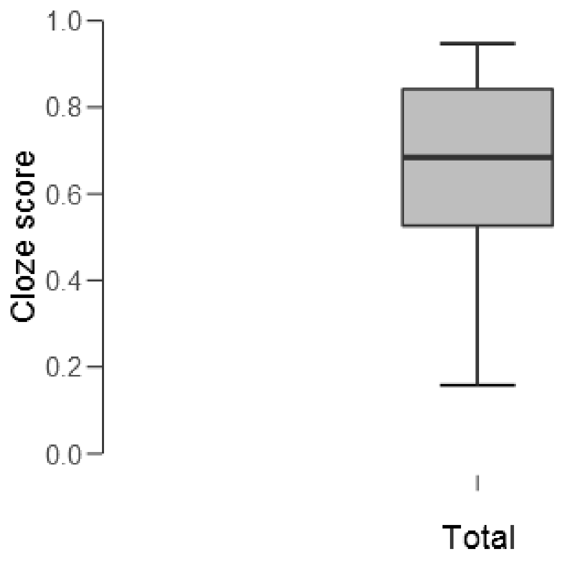


Figure 4.14. Distribution of cloze scores in the aphasic group.

The scatterplots in Figure 4.15 show the relationships between cloze task performance and z-score differences in first- and second-word manipulations. Since the Shapiro-Wilk test indicated deviation from normality in the cloze task scores ($W(21) = .903$, $p = .04$), the nonparametric correlation coefficient Kendall's tau (τ) was used. Cloze scores were positively related to sensitivity to collocation strength in first-word manipulations ($\tau = .49$, $p = .001$, $BF_{10} = 55.71$, one-tailed, assuming a positive relationship), and second-word manipulations ($\tau = .28$, $p = .044$, $BF_{10} = 2.27$, one-tailed, assuming a positive relationship). Bayesian Kendall correlation indicates very strong evidence for a positive relationship between cloze scores and first-word manipulations ($BF_{10} = 55.709$), but only anecdotal evidence for the same relationship in second-word manipulations ($BF_{10} = 2.273$).

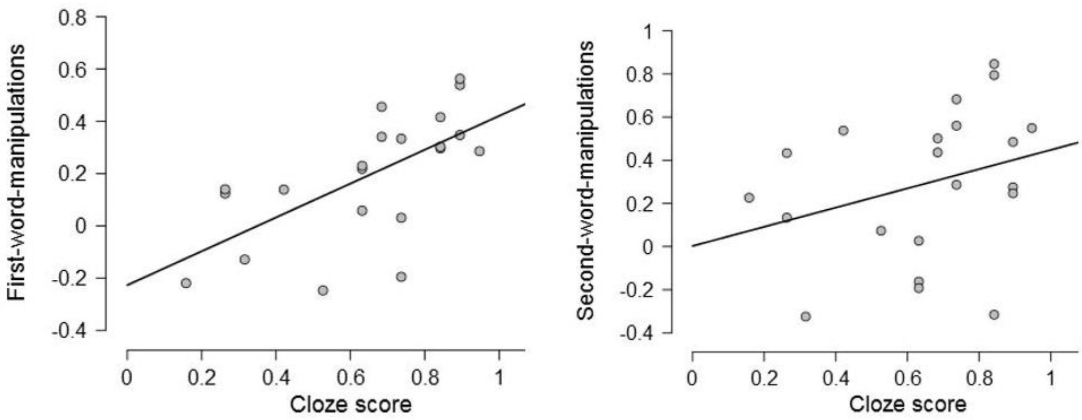


Figure 4.15. Average cloze probability and average z-score difference in first- and second-word manipulations across 21 PWA.

4.3.6.8. Relationship between facilitation and performance in off-line tasks (PWA only)

A range of language and cognitive assessments were administered with PWA and older control participants. An overview of background profiles in the aphasic group is shown in Table 4.14. Older control participants performed at ceiling in the PALPA-13, TROG-2 and the BNT. WASI-II MR raw scores, assessing nonverbal cognitive abilities, ranged between 12 and 22 (M = 17.87, SD = 2.5) in the older control group. The A.D.A. auditory minimal pair test indicated unimpaired auditory-lexical perception in all 21 PWA and 23 older control participants. The average score (out of 40) was 39.5 in the older control group (SD = 2.1; with one outlier participant, OC7, with a score of 30), and the average score in the aphasic group was 38.7 (SD = 2.3, range: 30-40).

Table 4.14: Background profiles of PWA and z-scores in the three WMT conditions

ID	**Gender**	**Age**	**TPO** (months)	**CAT subtest** (max. 30)	**PALPA-13** (max. 7)	**TROG-2 (blocks)** (max. 20)	**BNT** (max. 60)	**WASI-II MR** (max. 30)	**First-word manipulations**	**Second-word manipulations**	**Binomials**
A1	m	51	48	24	*n/a*	15	47	20	0.42	0.85	-0.20
A2	m	54	45	28	4	3	25	21	-0.25	0.07	1.03
A3	m	76	104	24	4	6	11	19	0.14	0.54	0.44
A4	m	63	76	27	7	2	17	9	-0.22	0.23	0.29
A5	m	57	79	30	7	19	59	21	0.54	0.28	0.53
A6	m	37	43	29	7	18	39	20	0.30	0.79	0.14
A7	m	51	87	28	7	7	32	14	0.46	0.50	0.63
A8	m	64	126	29	7	15	52	16	0.29	0.55	1.01
A10	f	48	38	26	7	11	27	18	0.03	0.29	1.80
A11	m	61	16	24	3	9	22	8	0.12	0.43	0.66
A13	m	67	54	27	7	11	52	19	0.56	0.25	0.75
A14	m	55	49	26	5	18	52	22	0.33	0.56	1.26
A15	m	84	156	30	4	5	14	7	0.14	0.13	0.58
A16	m	67	78	24	4	6	41	14	-0.13	-0.32	0.44
A17	f	58	98	28	7	12	33	11	-0.20	0.68	0.09
A18	m	68	75	28	5	17	52	21	0.34	0.44	0.52
A19	m	59	120	24	4	18	35	18	0.06	-0.16	-0.11
A20	f	71	162	29	7	13	48	12	0.22	0.03	0.76
A21	m	57	57	26	5	14	48	17	0.30	-0.32	0.59
A23	m	55	23	29	7	16	56	11	0.35	0.49	0.27
A24	m	67	62	14	*n/a*	2	9	16	0.23	-0.19	-0.53
AVERAGE		***60.5***	***76***	***26.4***	***5.7*** *6.8*	***11.3*** *19.2*	***36.7*** *56.3*	***15.9*** *17.9*	*.19*	*.29*	*.52*
SD		***10.3***	***40.1***	***3.5***	***1.5*** *0.5*	***5.7*** *1.3*	***15.8*** *2.8*	***4.7*** *2.5*	*.24*	*.34*	*.52*

Note. *Average and SD values of the older control group (N = 23 participants) are highlighted in grey.*

Next, relationships in the aphasic group between participants' sensitivity to collocation strength in first- and second-word manipulations, and TROG-2, BNT and WASI-II MR scores were investigated. Plotting the PALPA-13 and CAT scores did not reveal a linear relationship.

TROG-2- as well as BNT performance was positively related to facilitation in first- but not second-word manipulations. Table 4.15 presents the matrix for the correlation coefficients. The correlation coefficients between first-word manipulations and TROG-2 and BNT-scores were $r = .54$ ($p = .012$, $BF_{10} = 5.2$) and $r = .56$ ($p = .008$, $BF_{10} = 7.5$), respectively, showing significant relationships between these variables. However, the coefficients were lower and did not reach significance for second-word manipulations ($r = .35$, $p = .124$, *ns*, $BF_{10} = .8$ for TROG-2 scores and $r = .17$, $p = .47$, *ns*, $BF_{10} = .4$ for BNT scores). These results suggest that with better lexical retrieval and sentence comprehension abilities, the sensitivity to collocation strength in first-word manipulations was stronger. On the other hand, sensitivity to collocation strength in second-word manipulations did not seem to be related to lexical retrieval and sentence comprehension abilities. Nonverbal cognitive abilities as measured by the WASI-II MR subtest were not related to facilitation in stronger collocations (Table 4.15).

Table 4.15: Correlations between TROG-2-, BNT- and WASI-II MR scores and z-score differences in first- and second-word manipulations in the aphasic group.

	TROG-2	**BNT**	**WASI-II MR (raw score)**	**First-word manipulations**	**Second-word manipulations**
TROG-2	--	r = 0.802 (p < .001) *******	r = .435, p = .049 *****	r = .537, p = .012 *****	r = .346, p = .124
BNT		--	r = .383, p = .087	r = .564, p = .008 ******	r = .166, p = .472
WASI-II MR (raw score)			--	r = .356, p = .113	r = .129, p = .577
First-word manipulations				--	r = .271, p = .235
Second-word manipulations					--

*Note. * p < .05, ** p < .01, *** p < .001*

Finally, the relationships between proportion of ‘error’ responses (i.e., the sum of timeout and early responses divided by the number of trials by participant) and BNT-, TROG-2- and WASI-II MR performance were explored. Table 4.16 presents the correlation matrix. Results revealed that there was a significant negative relationship between the proportion of 'error' responses in first-word manipulations and both BNT- and TROG-2 performance. Moreover, the proportion of ‘error’ responses in second-word manipulations was also significantly correlated with both BNT- and TROG-2 performance. Hence, the higher the proportion of ‘error’ responses in a word monitoring experiment, the lower the performance in off-line tasks such as BNT and TROG-2. WASI-II MR performance, on the other hand, was not correlated with the proportion of ‘error’ trials in the WMT.

Table 4.16: Correlations between TROG-2-, BNT- and WASI-II MR scores and proportions of 'error' responses first- and second-word manipulations in the aphasic group.

	Proportion of 'error' responses (first-word manipulations)	**Proportion of 'error' responses (second-word manipulations)**	**BNT**	**TROG-2**	**WASI-II MR (raw score)**
Proportion of 'error' responses (first-word manipulations)	--	r = .614 (p = .003) **	r = -.719, p < .001 ***	r = -.609, p = .003 **	r = -.375, p = .094
Proportion of 'error' responses (second-word manipulations)		--	r = -.561, p = .008 **	r = -.457, p = .037 *	r = -.235, p = .305
BNT			--	r = .802, p < .001 ***	r = .383, p = .087
TROG-2				--	r = .435, p = .049 *
WASI-II MR (raw score)					--

*Note. * p < .05, ** p < .01, *** p < .001*

4.3.7. Discussion

The word monitoring paradigm is a method to study word recognition within real-time processing. Participants react to target words that appear in spoken sentences. These reflect the speed with which listeners are confronted with words and sentences in typical discourse. In the current study, stimuli were drawn from a large collection of everyday speech, the spoken BNC. Unlike previous word monitoring experiments in aphasia, which often addressed word recognition based on different syntactic cues, or the disruption of language processing within ungrammatical sentence contexts (e.g., Friederici, 1983; Tyler, 1985), the current experiment examined facilitatory processing of stronger versus weaker trigram collocations.

The data from the current study demonstrate sensitivity to collocation strength in all three participant groups (younger adults, older adults, PWA), with more rapid responses to final words of stronger compared to weaker trigram collocations. This confirms previous findings on phrase frequency effects in neurotypical adults (e.g., Arnon & Snider, 2010; Jiang & Nekrasova, 2007). Furthermore, these results provide novel evidence for a processing speed advantage of strongly collocated word combinations in aphasia, although it should be noted that effects in the aphasic group were weaker relative to the control group, and there was a relatively low number of participants with severe aphasia.

Sensitivity to collocation strength was modulated in all participant groups by manipulation of the first versus second element of a trigram, with a larger effect of collocation strength in second-word manipulations, i.e., when monitoring for a target word that directly followed the critical word. This suggests that larger effects can be elicited in the latter manipulation, compared to a condition where the critical word and the word to be monitored occur at a distance. This is also commonly reported in the literature (Marinis, 2010).

4.3.7.1. Facilitation in binomials

Binomials (e.g., "*bride and groom"* versus "*horse and groom"*) were added as a distractor condition. Any effects were explored separately from the experimental conditions. As expected, all three groups showed a processing speed advantage of binomials, relative to less common noun pairings, indicating anticipatory processes (Millar, 2010) based on the first two words of a binomial. Moreover, PWA displayed the lowest number of missing trials for binomials compared to the two experimental conditions, which may be attributed to the salience of noun pairings in a sentence.

Older adults showed a higher degree of facilitation of binomials, relative to younger adults. When comparing PWA to older controls, results indicate less sensitivity to manipulation in this condition in the aphasic group. Together with the significant difference between younger and older adults, these patterns might be indicative of hyper-facilitation in older adults. However, it should be noted that younger adults frequently anticipated target words in this condition (i.e., pressed the button before the target onset point) which may suggest an equal sensitivity in both neurotypical control groups, but since younger adults generally reacted faster to targets than older adults, younger adults' sensitivity might have been driven by anticipation.

This processing advantage for binomials over less common noun pairings is in line with a previous eye-tracking study (Siyanova-Chanturia, Conklin, & van Heuven, 2011), and a more recent experiment by Siyanova-Chanturia et al. (2017) who recorded event-related potentials (ERPs) and found electrophysiological evidence for binomial facilitatory effects. In the current study, the critical word in the binomials condition was a noun. Thus, facilitation in binomials might resemble a semantic priming effect (e.g., "*bride*" primes "*groom*") due to their semantic association, boosted by the additional presence of collocation (i.e., the presence of the noun with the conjunction "*and*"). The following paragraphs discuss the findings of the two experimental conditions (first- versus second-word manipulations) in the two neurotypical groups (younger versus older adults) and in the aphasic group as compared to an age-matched control group.

4.3.7.2. Age effects in the sensitivity to collocation strength

An analysis of missing trials and average RTs revealed that younger adults reacted to the target on average 100 ms faster than older controls. They also anticipated more and presented with a slightly lower overall ‘accuracy’ relative to older adults which points to a speed/accuracy trade-off in this group.

As predicted, findings from a mixed-factorial ANOVA suggest that facilitation in both younger and older adults was equally affected by experimental condition which was supported by a Bayesian Repeated Measures ANOVA. For both younger and older neurotypical adults, faster monitoring latencies for words embedded in stronger as compared to weaker collocations were driven by collocation strength of the trigram as well as the distance between the critical word and the target. There was little evidence for an influence of critical word frequency on sensitivity to collocation strength. While a significant correlation between these two variables in first-word manipulations was found for younger adults, a Bayesian correlational analysis revealed that there was only anecdotal evidence for a correlation. Thus, Bayesian procedures are a useful way of exploring how much evidence there is for a null- versus an alternative hypothesis.

Higher average z-score differences in second- compared to first-word manipulations are indicative of a distance effect, i.e., that the distance between critical word and target might lead to a more rapid decay of co-activation. In comparison, if the critical word immediately precedes the target, then co-activation of the critical word and target may still be high. To some extent, this result is compatible with the finding in Vilkaite & Schmitt (2017), where facilitatory effects on collocation reading in non-native speakers of English did not reach significance when there were intervening words between the first and last word of a collocation (e.g., “*provide information*” versus “*provide* <u>*some of the*</u> *information*”). The intervening words in Vilkaite & Schmitt (2017) resulted in a much greater distance between the beginning of the collocation and the target word as compared to the stimuli in the current study. Moreover, the stimuli in Vilkaite & Schmitt (2017) consistently represented verb-noun collocations, whereas the present study used different underlying phrase structures. The majority of critical words in first-word manipulations were prepositions, while most of the critical words in second-word manipulations represented determiners. This leaves open the question of whether the word class of the critical words (i.e., that mostly function words were manipulated) may have interacted with the magnitude of facilitation in the two experimental conditions. For instance, function words in first-word manipulations may not be salient enough to activate the lexical items which complete the collocation, whereas function words immediately preceding the target may lead to high enough co-activation for relatively large facilitatory effects.

For older adults, results suggest education-related differences in task performance in the WMT, a finding that is partly in line with the prediction that there would be a positive correlation between educational level and facilitatory effects. There was a link between higher education and greater sensitivity only in first-word manipulations, whereas no such relationship was found for second-word manipulations, i.e., facilitation in the latter condition was unrelated to educational level. This highlights the role of linguistic experience in sensitivity to usage-based variables such as collocation strength within higher task demands: Higher educated participants who likely have more experience with language, especially with written texts (Dąbrowska, 2012), showed relatively strong collocational effects even when there was a distance between the critical word and the target. This result is supported by Dąbrowska (2012) who found education-related effects in a series of experiments testing comprehension of a variety of syntactic structures. It is also supportive of another study by Dąbrowska (2014b), where strength of collocational knowledge was related to measures of linguistic experience such as education. This result can be explained by usage-based, constructional models of language, where more language experience leads to more collocational knowledge (Ellis, 2008a). The present findings suggest that collocational knowledge may not only increase, but may also become more robust with more linguistic experience.

For both experimental conditions (i.e., all 36 trigram pairs), there was a significant relationship between z-score- and t-score differences in younger adults (supported by an additional Bayesian analysis), but not in the older control group. This suggests that degree of facilitation in younger, but not older adults is influenced by the strength of the association between components of a trigram as compared to its counterpart. In light of the prediction that there would be a positive correlation between these two variables, independent of age, this was an unexpected finding. This age effect could stem from the slower RTs in older adults which may have ‘masked’ a subtle effect in this group. The same explanation may hold for the aphasic group, where there was no correlation between z-score- and t-score differences.

4.3.7.3. Sensitivity to collocation strength in PWA

A comparison between facilitatory effects in PWA and an age-matched control group revealed that PWA showed sensitivity to collocation strength, displaying faster RTs to the final words of strongly collocated as compared to weaker collocated trigrams. However, the interaction of condition and group on sensitivity to collocation strength indicates that PWA, as compared to neurotypical control participants, showed less facilitation when the word immediately preceding the target was manipulated. Despite this finding, these results provide more support for the role of frequency-based variables in aphasic language processing, which is in line with previous aphasia research on structural frequency effects (e.g., Knilans & DeDe, 2015). Strongly collocated word combinations facilitate word recognition, however, to a weaker degree compared to age-matched control participants, indicating weaker anticipatory activation in PWA relative to control participants. In comparison, Huck et al. (2017) found effects of word frequency and contextual predictability on silent reading, with a word frequency effect similar to the control group, and a greater effect of contextual predictability in PWA as compared to controls. While Huck et al. (2017) report similar or greater effects of usage-based variables in the aphasic group, relative to controls, the current results indicate diminished effects in the aphasic group. The difference in the magnitude of these effects may have been influenced by differences in participant groups (mainly mild aphasia in Huck et al. versus mainly mild to moderate aphasia in the present study) and the nature of the task: In Huck et al., participants did not have to focus on an additional task while reading sentences, whereas in the present study, they had to remember the target word and press a button under time pressure. Furthermore, in Huck et al. (2017), better sentence comprehension scores were associated with a greater effect of contextual predictability. The present findings are compatible with this finding, in that better sentence comprehension correlated with facilitation in first-word manipulations, which might indicate that the ability of using the preceding sentential context improves sentence comprehension.

The presence of aphasia affected RT to the target words. The average RT in the aphasic group was lower compared to controls, and PWA had more timeout responses, which may be related to aphasic sentence processing difficulties. In the aphasic group, 17 of the 21 participants presented with TROG-2 scores below the neurotypical range, indicating auditory sentence comprehension difficulties which may partly explain the relatively high number of timeout responses and overall slower RT compared to the age-matched control group. Marinis (2010, p. 141) notes that “participants with low working memory or problems with lexical access may show a lower accuracy rate (…) and/or slower reaction times compared to adults with typical language abilities”. These are observations which the current data partly support. Task demands, i.e., remembering the word while listening to spoken sentences and pressing a button may have influenced the relatively high number of timeouts in the aphasic group. This is related to an issue outlined in Kilborn & Moss (1996), who state that remembering the target and task instruction, while attending to the carrier sentence(s), may make the task unsuitable for some participants.

While the results of the current study suggest that word monitoring experiments represent a suitable paradigm to investigate on-line spoken word recognition within sentences in aphasia, three participants had difficulties in performing the task. Possible reasons may have been the severity of underlying language and cognitive difficulties. In one case, word meaning deafness seems to have caused problems with reacting to target words in aurally presented sentences, which resulted in a large number of unusable items. Because of the limited interpretability of this participant’s data, this participant was removed from subsequent analyses. In a second case, the individual with severe aphasia showed signs of frustration with the technology. A third participant with more marked aphasia who completed the task was unable to respond to targets before the whole sentence finished and seemed frustrated and overwhelmed by the experiment.

To reduce task demands, future WMTs for studying aphasic sentence processing could be refined by adding a picture of the word to be monitored (Tyler, 1985). When using such a format, the words to be monitored would have to be highly imageable.

Trigram collocations such as "*answer* / *repair the phone*" or "*catch* / *paint the train*" (used in a second language learning study by Matsuno, 2017) would be suitable for such a format. These stimuli have high functional relevance and are similar to binomials, where both the critical word and the target are content words, and it would be likely that such stimuli, varying in degree of collocation strength, would elicit low 'error' rates and strong facilitatory effects. Another opportunity for future research into real-time language processing in aphasia could be to explore differences in brain activation between stronger and weaker collocations by using ERPs, as ERP studies may reduce the processing load by eliminating behavioural responses (e.g., no button press needed).

4.3.7.4. The relationship between sensitivity to collocation strength in PWA and performance in off-line tasks

The pilot results in the single-response cloze task were replicated in a larger sample of neurotypical speakers, and the scoring procedure for PWA was informed by the findings from the normative sample. Cloze task performance in PWA was positively related to facilitation in first- and second-word manipulations. However, there was more robust evidence for a correlation between cloze task score and sensitivity to collocation strength in first-word manipulations. This result needs to be interpreted in the context of the potential confound of accessing and producing single words in a sentence completion task. In 15 of the 21 PWA, lexical retrieval of nouns, as measured by the BNT, was below the neurotypical range. In addition, the aurally presented stories of the cloze task need to be comprehended in order to achieve a high score, i.e., sentence comprehension needs to be relatively well preserved which was likely impaired in many of the participants in the aphasic group. Nevertheless, despite these confounds, a relationship between the degree of sensitivity to collocation strength and the ability to complete formulas within the off-line cloze task was still detected. This tentatively suggests that sensitivity to collocation strength may be related to a participant's ability to complete (and potentially to produce) formulas, indicative of co-activation of the first elements and the remaining collocation. However, formulas in the cloze task were often 'primed' with content words, and varied in length and mostly represented conversational formulas such as "*good* *night*", "*I'm terribly* *sorry*" and "*pleased to meet* *you*"; thus, they are not directly comparable with the stimuli employed in the WMT.

Sensitivity to collocation strength did not correlate with nonverbal IQ scores, confirming that the present experiment did not measure general ability (Dąbrowska, 2014b). TROG-2- and BNT scores in the aphasic group correlated with facilitation in first-word manipulations, showing that PWA with more impaired lexical retrieval and sentence comprehension difficulties were less likely to show facilitation in that condition. In contrast, sensitivity to collocation strength within second-word manipulations was found regardless of degree of language impairment. This mirrors the finding that education level in older controls influenced sensitivity to collocation strength in first-word manipulations only, and that there was stronger evidence for a relationship between cloze task performance and first-word manipulations. Hence, an on-line task where there is distance between the target and the critical word (e.g., first-word manipulations) may be indicative of aphasia severity. Future word monitoring experiments could examine the relationship between overall aphasia severity and sensitivity to collocation strength more closely.

Finally, an error analysis in addition to differences in monitoring latencies can be useful in exploring which language profiles may be most suitable for such an experiment. 'Error' and facilitation correlations suggest that participants with better language comprehension and lexical retrieval abilities performed better in a WMT compared to participants with more severe language impairment. Although this relationship should be kept in mind when designing future word monitoring experiments, facilitatory effects were still observed in many participants with more severe aphasia (see Table 4.14).

4.3.8. Conclusion

This study investigated word recognition processes in different frequency contexts in PWA and neurotypical controls. A word monitoring experiment was devised to examine usage-based effects (facilitatory processing of familiar collocations) in real-time processing. The findings extend previous research by demonstrating that PWA process familiar, collocated word combinations more easily than less familiar, weaker collocations, although effects were stronger for neurotypical control participants. This may be explained by weakened co-activation processes in aphasia.

Stronger facilitatory effects were observed when the critical word immediately preceded the target, or when the target was primed by a noun and the conjunction "*and*", as it was the case in the distractor condition (binomials). This suggests that facilitation was greater when there was a rich semantic association on top of existence of a collocation. Sentences where there was a distance between the critical and the target words seemed to be more sensitive to education-related differences in older neurotypical participants, as well as off-line processing in PWA. Thus, while sentences with no distance between the critical word and target are a more robust condition to detect collocational facilitatory effects, both experimental manipulations have advantages, depending on the specific research question or application. Future work could explore the relationship between the word class of the critical word, its position in the collocation and the size of the facilitatory effect. In summary, the word monitoring paradigm represents a useful instrument for assessing on-line collocational and sentence processing in aphasia.

5. Study 3: A case series report of a novel intervention targeting flexible use of familiar phrases in NFA

5.1. Introduction

Aphasia results in difficulties in processing words, multi-word utterances and sentences, linguistic structures that are important features of everyday conversations. Some word combinations, many of which can be described as familiar collocations, remain available despite aphasia (study 1; Zimmerer et al., 2018). Investigating the conversational functions of the familiar collocation IDK, study 1 found that individuals with Broca's aphasia used the phrase predominantly as a fixed construction, often with a turn-constructional function. This finding is supported by previous literature (Beeke, 2003; Simmons-Mackie & Damico, 1997). In contrast, neurotypical CPs produced more flexible syntactic variations of IDK (for example with a complement phrase), and a higher number of IDK tokens were used in a multifunctional way. With regard to recognition of familiar collocations, study 2 indicated that listeners with aphasia show sensitivity to collocation strength, although to a weaker degree compared to an age-matched control group. This study investigates whether individuals with NFA can learn to use residual, familiar collocations such as IDK more flexibly (e.g., as an abstract construction "*I don't know* [X]" to express a lack of knowledge).

As outlined in Chapter 2, some elements of existing speech and language therapy interventions for NFA are in accordance with usage-based assumptions, where linguistic form is paired with semantic-pragmatic meaning. These include ILAT (Stahl et al., 2016), REST (Ruiter et al., 2010; Schlenck et al., 1995; Springer et al., 2000), Carragher's hybrid therapy (Carragher et al., 2015), SPPA (Helm-Estabrooks & Nicholas, 2000), script training (Bilda, 2011; Cherney et al., 2008; S. Goldberg et al., 2012; Kaye & Cherney, 2016), and SentenceShaper® (Linebarger, 2015; Linebarger et al., 2008). In a usage-based framework, the basic linguistic unit can be a single word, a familiar collocation ("*thank you*", targeted in MIT and also employed in ILAT), a semi-fixed frame with open slots ("*I'd like X*", an example from script training), or an abstract constructional schema (e.g., the elliptical [DOING/DONE] + [WHAT] construction used in REST).

As yet, however, no intervention program for aphasia has explicitly tested application of usage-based principles such as high-frequency phrases, constructional grounding (Israel et al., 2001; Riches, 2013) or superimposition (Dąbrowska, 2014a), processes that help create novel utterances (Chapter 2). With regard to constructional grounding, source constructions are often lexically specific and more frequent than the target constructions (Israel et al., 2001). In this way, a familiar collocation such as "*I like it*" may act as a source construction, representing an instance of the more abstract "[REFERENT] *like*-TENSE [THING]" constructional schema. Superimposition (Dąbrowska, 2014a) is a process where lexical items or chunks are inserted into an open slot of a semi-fixed construction (e.g., "*<u>all of us</u> like <u>Anna's lemon drizzle cake</u>*"). Learning such mechanisms can help speakers with communication difficulties to produce longer and more flexible utterances, and usage-based theory allows targeting words, word combinations or sentences, depending on a speaker's resources. There is a need for developing usage-based intervention programs for aphasia and evaluating their impact on language production at the connected speech level.

Many of the intervention studies that touch on usage-based assumptions used outcome measures of connected speech, involving sentence production tasks (e.g., subtests of standardised assessment batteries such as the AAT, used by Stahl et al., 2016), spontaneous speech and discourse tasks (e.g., Ruiter et al., 2010; Springer et al., 2000) and everyday conversational data (Carragher et al., 2015). Whether connected speech is stable over time, and thus represents a valid format to assess baseline performance in individual participants, has begun to be examined by Beales, Whitworth, Cartwright, Panegyres, & Kane (2018) and Whitworth et al. (2018). Beales et al. (2018) demonstrated that the connected speech of participants with primary progressive aphasia and AD was stable across measures and different genres (e.g., everyday monologic speech versus picture description). Whitworth et al. (2018) extended these findings with data from participants with post-stroke aphasia, and found stability for some genres, which points to the utility of such outcome measures. They emphasize that procedural monologues were less detailed when repeating the assessment after one week, as compared to a 3-week interval. Thus, the sampling interval (e.g., testing 1 week versus 3 weeks apart) may be crucial for some genres.

Economic and reliable measures that effectively evaluate outcomes at the level of multiword utterances need to be developed for connected speech interventions. While some investigators included analysis of multiword production such as the proportion of verb phrases (Carragher et al., 2015) and length of ellipses produced (Ruiter et al., 2010), most measures of connected speech evaluated features of single words, for instance the number of content words, or number of words produced per minute (e.g., Cherney et al., 2008). Zimmerer et al.'s (2018) frequency-based methods offer one avenue for exploring multiword effects for bi- and trigrams (see also study 1). However, whether these measures are suitable to reliably assess therapeutic change in aphasia is unclear.

The present study tested a novel behavioural intervention for NFA aimed at increasing the productivity of multi-word expressions, as measured by frequency-based variables. Flexible use of multi-word utterances was approached from a usage-based perspective, loosening open slots in familiar semi-fixed constructions such as "[REFERENT] *like*-TENSE [THING]" by superimposing lexical items or chunks. The intervention incorporated psycholinguistic and neuroscientific learning principles: structural priming, errorless learning and enhancing dose through home practice. Moreover, this study considered social-motivational factors: constructions were practiced by making use of a participant's own voice: target phrases are recorded and modified to create fluent versions ('self-voice').[16] The following paragraphs explain each of these principles and their rationale in more detail.

The defining characteristic of NFA is disfluent language production which can extend to disrupted repetition of sentences and reading aloud (Papathanasiou, Coppens, & Potagas, 2013). In a study motivated by the finding that listeners often have negative perceptions of speakers with NFA, Harmon, Jacks, Haley, & Faldowski (2016) modified non-fluent language output from nine participants with NFA by using computer software. Listeners rated non-fluent recordings and the simulated fluent versions of the same recordings with regard to speech output and their thoughts and

[16] One motivation to include 'self-voice' in the current intervention was Dr Carolyn Bruce's ongoing work exploring use of self-voice in aphasia therapy.

feelings about the speaker. Findings showed that digitally altered recordings improved listeners' perceptions of speakers with NFA. The current study applied a procedure to enhance fluency via waveform editing, similar to the procedures reported in Harmon et al. (2016). These recordings of target constructions were incorporated into the intervention. This principle will be referred to as 'self-voice'.

In this study, the self-voice component acted as a social-motivational learning element (Varley, 2011). The underlying mechanism of self-voice is the production effect, or enhanced memory for words that are read aloud (Forrin & MacLeod, 2017). Forrin & MacLeod (2017) found that participants (undergraduate students) who heard recordings of themselves reading words aloud showed better memory for these words as compared to a condition where recordings were those of a different speaker, highlighting the importance of both a motor speech component and a self-referential auditory component in learning and remembering. Brain imaging data suggest that the RH is responsible for processing self-voice (J. T. Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008).

Another principle applied in the current study is structural priming, a phenomenon where a speaker is more likely to re-use a structure they have previously encountered (Kaschak, Kutta, & Coyle, 2014). In the structural priming paradigm, a participant is primed with a specific syntactic structure or a schema (e.g., "*he gave me a book*"), for instance in a sentence repetition task or by reading aloud. The second part typically uses a picture description task where the participant tends to describe the event employing the syntactic structure they were primed with (e.g., "*she sent him an email*" as opposed to "*she sent an email to him*"; examples taken from Blumenthal-Dramé, 2016).

Individuals with aphasia can produce transitive and dative sentences when primed with these structures, as shown in Hartsuiker & Kolk (1998) who examined structural priming effects in 12 participants with Broca's aphasia, and in Saffran & Martin (1997) who tested five participants with varying types of aphasia (fluent and NFA). Saffran & Martin (1997) targeted transitive (actives and passives) and dative (prepositional and double object datives) structures in their experiment, where sentence repetition (e.g., repeating a passive sentence) alternated with picture description tasks (e.g., describing a picture of a transitive event). Priming was found for transitive but not for dative structures. In addition, they used a pre-post design, where participants completed elicitation tasks (picture descriptions) for passives and datives before and after the actual structural priming experiment. The interval between the pre- and post-test of both sentence types ranged from 1 hour to 1 week. While there was no significant difference between passives produced before versus after the experiment, longer-term structural priming effects were found for datives, with significantly more dative structures produced at post-test probes. This is in line with long-term cumulative priming effects found in neurotypical adults (Kaschak et al., 2014), where structural priming accumulates over many sentences, and effects can be observed even when the 'priming phase' (i.e., the picture description session) is delayed by one week. Such long-term effects occur if the same task is used in the bias phase (where participants are primed with a specific structure) and the priming phase (Kaschak et al., 2014). If the bias- and priming phases employ different tasks, however, effects are weaker. Lee & Man (2017) used implicit structural priming training with an individual with Broca's aphasia and found improved sentence production of prepositional dative structures (e.g., "*the boy is giving a guitar to the singer*"). This was found both directly after the intervention and at 4 weeks maintenance.

Another learning principle that has been found effective in aphasia intervention studies is errorless learning, a method where "the possibility of the learner making errors is reduced or eliminated entirely" (Varley, 2011, p. 16). Errorless learning methods employed in aphasia studies include decreasing cues (Conroy & Scowcroft, 2012) or, as in a 'pure' errorless learning method, direct repetition of a target structure. In the 'Sheffield WORD – Structured speech therapy' program (SWORD; Varley et al., 2016; Whiteside et al., 2012), designed for individuals with acquired apraxia of speech (AoS), fluent and flexible use of constructions (words in isolation and embedded in phrases such as "*Where is* ___?") is practiced in a stepwise way, using errorless methods. In the first step, the participant with AoS is presented with a video clip of a neurotypical speaker producing a target word. This is followed by a second step, where the participant imagined saying the word. In a third step, the participant attempts production (repetition) of the word. This procedure – which was implemented in the current intervention – seeks to facilitate fluent language production by minimising errors.

Finally, the intervention devised in the current study was computerised, to allow opportunities for self-managed home practice. This increases the frequency with which participants engage with intervention activities (Nobis-Bosch, Springer, Radermacher, & Huber, 2011). High intensity aphasia therapy (8.8 hours per week for 11 weeks) has been found to be more effective compared to lower intensity therapy, administered over a longer time span (Bhogal, Teasell, & Speechley, 2003). The neuroscientific principle of increasing therapeutic dose (total number, length and frequency of sessions) through home practice has become a common component of aphasia interventions. Varley et al. (2016), for instance, found a positive relationship between dose (the time spent on a self-administered computerised intervention) and treatment outcome (correctly named words post-intervention) in a large group of participants with AoS.

The current study set out to evaluate the intervention's acceptability as well as the outcomes in a case series, to test the potential of this computerised, usage-based intervention for improving expressive abilities in NFA. These are elements typically explored in feasibility studies (Orsmond & Cohn, 2015). The following main research questions will guide the analysis:

1. Is there evidence that after intervention participants with NFA demonstrate
   a. enhanced connected speech, as measured by a higher proportion of multiword utterances in narratives?
   b. increased use of trained constructions?
2. Which is the most appropriate outcome measure?
3. How acceptable is the intervention to participants with NFA and their CPs?

5.2. Method

5.2.1. Participants

Following ethical approval from the UCL Language & Cognition Departmental Ethics Committee (Project ID: LCRD.2017.01), five PWA and their regular CPs were recruited via non-NHS routes including a university research register. Participants were initially contacted by the researcher via email or telephone to explain the purpose, procedures and timeline of the study. If a dyad was interested in the project, written information sheets (one version for PWA, one for CPs, see Appendices C.1-C.2) were supplied and each dyad was given time to talk about the project with family members and friends before making a decision. When a dyad decided to volunteer to take part in the project, written consent was obtained both from the PWA and their CP (Appendices C.3-C.4).

Three participants were female and two were male. The mean age was 60 (range: 48-68). All participants presented with chronic, post-stroke aphasia (average 91 months post stroke onset, range 24-165), characterized by grammatically impoverished, non-fluent and effortful speech output in spontaneous speech and narrative production. Table 5.1 presents an overview of demographic characteristics and performance on background/profiling assessments. All participants:

- were adults, with a lower age limit of 18 years (no upper age limit);
- had a stroke at least 6 months prior to recruitment;
- presented with NFA, as identified by Speech and Language Therapists working in the UCL Department of Language & Cognition;
- had no neurodegenerative illness;
- had a spouse, adult family member or friend who was willing to take part in conversation recordings prior to and after the intervention;
- used English as their main language;
- had sufficient auditory and visual acuity to interact with a laptop computer;
- volunteered to join a university research register.

Table 5.1: Demographic characteristics and background assessments of participants with NFA

Participant ID	**Age**	**Gender**	**Months post onset**	**Highest level of education**	**Naming objects (BNT)** [1] (max. 60)	**Digit Span by recognition (PALPA-13)** [2] (max. 7)	**Spoken word comprehension (CAT subtest)** [3] (max. 30)	**Nonverbal IQ (WASI-II MR)** [4] (max. 30)
P1	48	f	46	Secondary school	30	6	28	17
P2	63	m	84	Secondary school	25	3	28	13
P3	60	m	165	University degree	10	4	24	12
P4	68	f	24	College	7	3	24	10
P5	60	f	137	Postgraduate degree	55	6	28	15
Average	*59.8*	-	*91.2*	-	*25.4*	*4.4*	*26.4*	*13.4*
SD	*7.4*	-	*59.5*	-	*19.2*	*1.5*	*2.2*	*2.7*

[1]*Boston Naming Test (BNT; E. Kaplan et al., 2001);* [2]*Subtest 13 from the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay et al., 1992);* [3]*Subtest from the Comprehensive Aphasia Test (CAT; Swinburn et al., 2004);* [4]*Matrix Reasoning subscale from the Wechsler Abbreviated Scale of Intelligence – Second Edition (Wechsler, 2011)*

As shown in Table 5.1, all participants showed relatively preserved spoken word comprehension (range: 24-28 out of 30), as determined by the CAT subtest (Swinburn et al., 2004). However, the group was heterogeneous with regard to object naming-, digit span- and nonverbal IQ performance, as measured by the BNT (E. Kaplan et al., 2001), subtest 13 from the PALPA battery (Kay et al., 1992) and the WASI-II subtest (Wechsler, 2011), respectively. The background profiles revealed variability with respect to aphasia severity: P4 presented with the most severe aphasia, followed by P3 and P2 (Table 5.1). P1's and P5's background profiles indicated less severe aphasia compared to the rest of the group, although P1's object naming was considerably more affected by her aphasia than P5's naming performance.

5.2.2. Study design

Employing a case series design, each participant was involved in a 16-week study, consisting of three phases: baseline (weeks 1-3), intervention (weeks 4-9), and post-intervention probes (weeks 10-16). Section 5.2.3 details the intervention and section 5.2.4 describes the measures used in the baseline and post-intervention phases.

5.2.3. Intervention

During the 6-week intervention phase (weeks 4-9), individual sessions in the UCL Department of Language & Cognition alternated with home visits. These visits were provided to ensure that participants could complete self-managed activities on a laptop supplied by the research team. The main aim of the intervention was to train flexible use of 12 constructions (see Chapter 2 for a definition of this term). Each represented a familiar collocation with high usage frequency (a source construction such as "*I like it*") which mapped onto a semi-fixed frame with open slots (e.g., "[REFERENT] *like* [THING / PERSON / LOCATION / PROCESS]", as in "*you like cake*"). Table 5.2 shows the full list of target constructions, as well as their semantic-pragmatic functions. These corresponded to 'topics', overarching interactional functions relevant to everyday talk about experiences, opinions and exchanging information. The topics listed in Table 5.2 reflect the order in which constructions were practiced.

Table 5.2: Target constructions (source constructions and frames) of the intervention

Source construction and frame	Semantic-pragmatic function	'Topic'
I made it [REFERENT] **made** [THING]	accomplishing or producing something	1.1: Giving information
I want that [REFERENT] **want** [THING]	to wish for a particular thing or plan of action	
I said it [REFERENT] **said** [UTTERANCE]	referring to something that has been reported before / that someone said	
I went to Spain [REFERENT] **went to** [LOCATION / EVENT]	reporting about having travelled or moved to a place	1.2: Giving information
I had it [REFERENT] **had** [THING]	to own; experience something; eat or drink something	
I want to swim [REFERENT] **want to** [PROCESS]	to wish for a particular plan of action	
It's alright [REFERENT / THING]**'s** [PROPERTY]	assessing a situation / a person / etc.	2: Giving an opinion
I like it [REFERENT] **like** [THING / PERSON / LOCATION / PROCESS]	enjoying or approving of something or someone / an activity	
I hate it [REFERENT] **hate** [THING / PERSON / LOCATION / PROCESS]	disliking something / someone / an activity very much	
Where is it? **Where is** [THING / PERSON / EVENT]?	seeking information with regard to place or location	3: Asking a question / expressing lack of knowledge
When is it? **When is** [THING / PERSON / EVENT]?	seeking information with regard to time	
I don't know [REFERENT] **don't know** [SOMETHING]	declaring insufficient knowledge	

The 12 source constructions were derived from the spoken BNC (average raw frequency = 1049 occurrences; range: 35-9034). Relative frequencies and t-scores of the source constructions can be found in Appendix C.5. First, the most frequent "[pronoun] + [verb] + *it*" and "*it's* + [adjective]" constructions were searched in the spoken BNC. Phrases that were regarded as representative of everyday language use were selected and categorized into the broader interactional functions of 'giving an opinion' or 'giving information'. The aim was to identify three constructions to represent each interactional topic (Table 5.2). In order to achieve this aim, the "*want*" and "*went to*" constructions completed the existing topics 1.1, 1.2 and 2. To reduce lexical-semantic demands, the open slots in source constructions mostly consisted of pro-forms (e.g., "*that*" or pronouns such as "*I*", "*he*" etc) – items with high unigram frequency (e.g., "*I*", "*he*", and "*that*" have 24,602, 7,552, and 22,780 occurrences per million words in the spoken BNC, respectively) and versatile as to the conversational contexts in which they can be used. Finally, the phrase IDK was added as it represents the most frequent trigram in neurotypical talk and was found to be a relatively fixed, yet frequently used phrase in NFA (study 1). The prototypical 'lack of knowledge' usage of IDK (e.g., "*I don't know where*") was targeted. To complete a set of three constructions around the topic of 'expressing lack of knowledge and asking a question', two *wh*-questions with high usage frequency were added (Table 5.2).

The intervention was divided into three phases (Figure 5.1), and the following underlying techniques and principles were applied: self-voice, structural priming, superimposition, errorless learning and increasing dose through home practice. Phase 1 (week 4) elicited fluent as possible versions of a participant's own productions of the target phrases (self-voice). In Phase 2 (week 5), participants were exposed to variations of the target constructions within a gamified RT experiment using the word monitoring paradigm (structural priming), and Phase 3 (weeks 6-9) served as a production training phase (practicing flexible use of constructions via superimposition, self-voice, errorless learning and increasing dose through home practice).

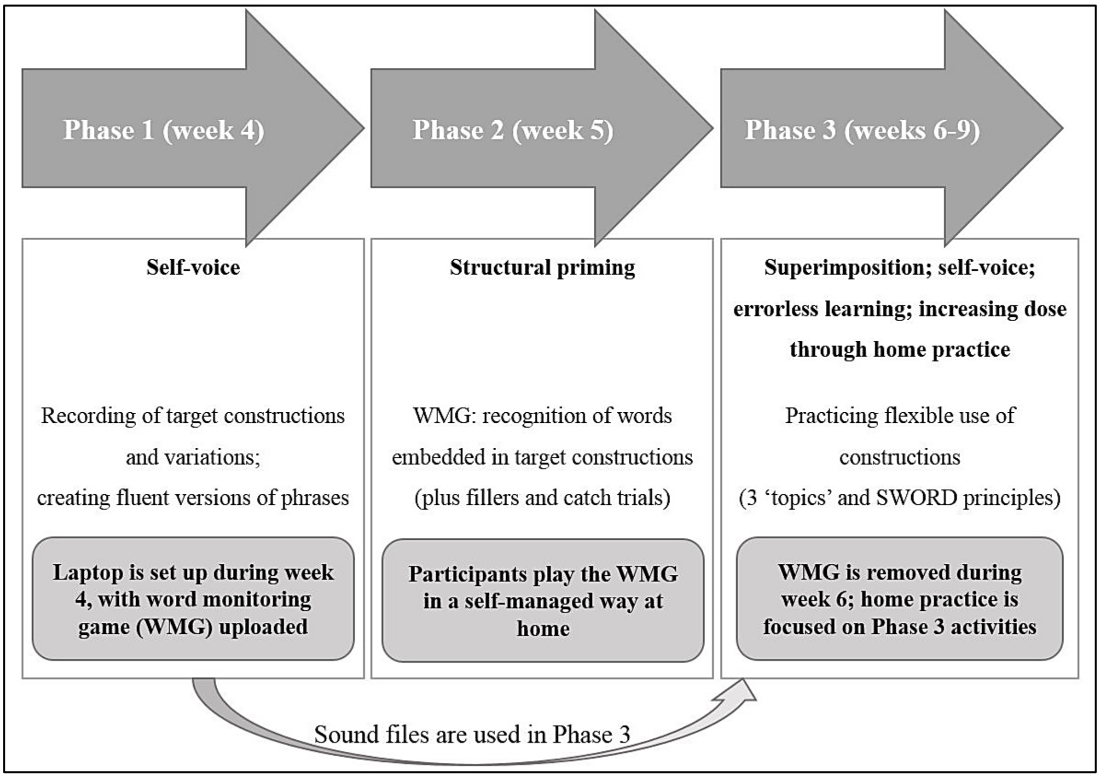


Figure 5.1. Three phases of the intervention.

Phase 1 consisted of one recording session of target constructions (e.g., "*I had it*", "*I made it*") and variations of these (e.g., "*she had it*", "*he made cake*"). Participants repeated phrases or read them aloud. Each participant recorded the same set of 59 phrases (names were individualised, e.g., "*[NAME] had dinner*"). If parts of a recording were effortful and non-fluent, the phrase was modified using the sound-editing software Audacity (version 2.1.2; available at https://www.audacityteam.org/). Modifications through waveform editing included deleting fillers and reducing long vowel durations or long pauses between words, similar to the procedures described in Harmon et al. (2016). Figure 5.2 gives an example of such a modification. If a recording included errorful elements (e.g., apraxic errors), more than one attempt was recorded and correctly produced words were 'copied and pasted' into the final sound file. In this way, fluent versions of all 59 phrases were created for each participant. These recordings were part of the computerised practice for Phase 3 of the intervention.

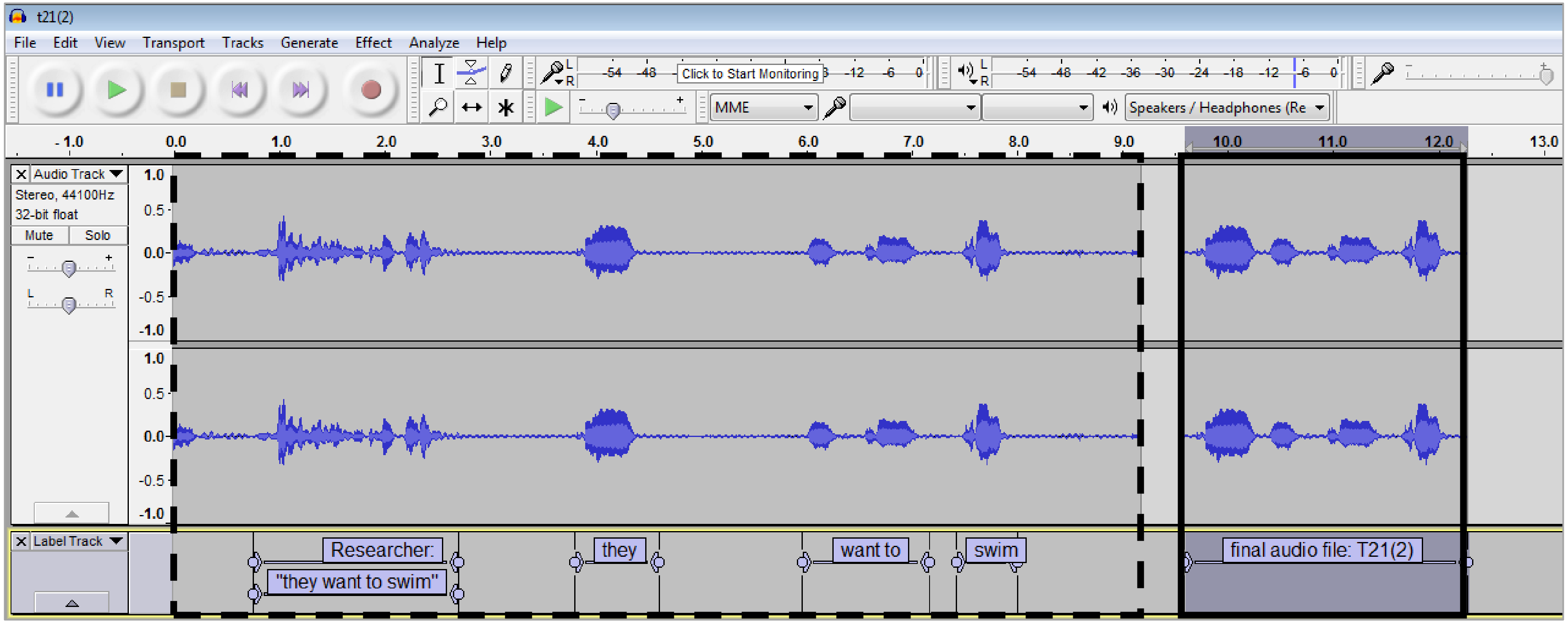


Figure 5.2. Example: Researcher models "they want to swim", participant (P2) repeats the phrase (highlighted with a dashed line). The final phrase (edited) is shown on the right hand side (highlighted with a solid line): pauses have been shortened compared to the original recording.

During Phase 2 (week 5), participants were exposed to variations of constructions via a 'word monitoring game' (WMG), a gamified, computerised WMT. In the WMG, the participant reacts as quickly as possible (via button press) to a pre-specified target word (presented in written and auditory form) once it is encountered in an auditory sentential context. A total of 320 trials were created, and divided into three categories (see Figure 5.3 for an example of each): a) typical trials (TT), where participants reacted to words embedded in target constructions (structural priming), b) filler trials (FT) where target words were not part of target constructions, but appeared at a different place in the sentence, and c) catch trials (CT), where the target word did not appear in the sentence, i.e., the participant was not supposed to press the button. FT and CT were included to ensure participants were not aware of the purpose of the task.

Table 5.3: Examples of WMG conditions

Condition	**Target word**	**Sentential context**
Typical trial	made	This bread recipe shows you how I **made** it.
Filler trial	game	If you make a mistake the **game** plays a sound.
Catch trial	market	All those clothes are made by hand.

200 trials represented TT (62.5%), 50 represented FT (15.6%), and the remaining 70 were CT (21.9%). The program randomly generated fourteen unique sets of 20 items (consisting of 13-14 TT, 3 F and 3-4 CT) to form sessions of four minutes. After each trial, participants received feedback on accuracy, using smiley / neutral faces and short messages such as "*too early*" (if pressing the button before the actual target word occurred in the sentence) or "*it wasn't there this time*" (if pressing the button in a CT). After each four-minute session, the program displayed two feedback graphs, providing an overview of RTs and accuracy of the previous 10 sessions (see Figure 5.3 for an example of a RT feedback graph).

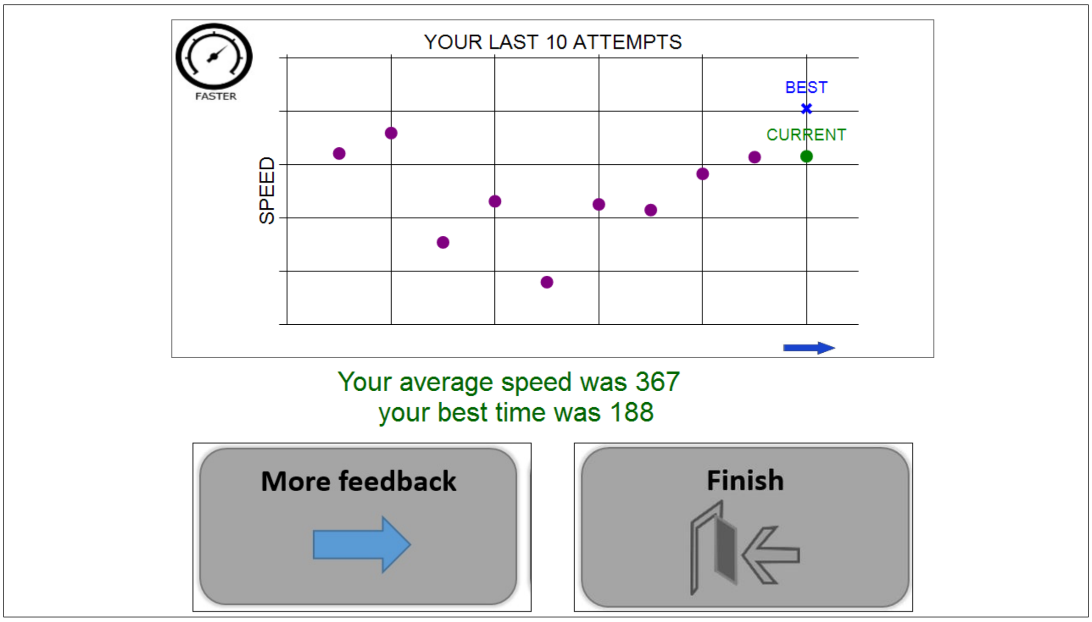


Figure 5.3. Screenshot of a WMG feedback graph (displaying RT).

The target words and sentential contexts were recorded in a sound treated room at UCL, using the program ProRec (version 2.2, Huckvale, 2016; downloaded in June 2016). A female native speaker of English read out the target words and the sentential contexts with a relatively slow pace and natural prosody. The WMG was self-managed and the program automatically recorded when and for how long the WMG was practiced. Each participant was loaned a laptop and was provided with aphasia-friendly instructions of how to start / shut down the laptop and play the WMG. The researcher set up the laptop in each participant's home during week 4 and provided initial training including practicing using the laptop and playing the WMG.

The intended learning mechanism of Phase 2 was structural priming. In the current study, the WMG was comparable with a bias phase. By reacting to target words that were part of trained constructions, participants were paying attention to constructional schemas. The aim here was to activate constructional schemas by repeated exposure in week 5, which may prime and thus facilitate production of these constructions in the following weeks (Phase 3). The WMG was removed from a participant's laptop during week 6, when the focus shifted to Phase 3 activities.

In Phase 3, the sound files from Phase 1 (recording session) were incorporated into a computerised program designed to practice more flexible use of constructions. Phase 3 consisted of individual sessions, home visits and additional self-managed home practice, using the same laptop that had already been used for the WMG. In each week during weeks 6-9, three different constructional schemas were practiced. Four slide shows were created using Microsoft PowerPoint, where constructions were presented in written form, accompanied by the participant's audio recordings. Each slide show (referred to as 'topic', Table 5.2) included three constructions that related to one overarching interactional function such as 'giving an opinion'.

The main underlying usage-based principle in Phase 3 was superimposition (Dąbrowska, 2014a), where lexical items (e.g., "*Claire*") were superimposed over an open slot (e.g., "*Where is* ____*?*" → "*Where is Claire*").[17] Moreover, a lexically complex construction (in this study, for example "*they like flowers*") was grounded in a relatively simple source construction (e.g., "*I like it*"). Open slots in source constructions were often filled with pro-forms (e.g., pronouns and deictic expressions such as "*that*") as well as high-frequency verbs and nouns (e.g., "*dinner*", "*swim*") to reduce lexical demands. Semantically more rich items were used in following steps.

During Phase 3, the three steps used in SWORD (Whiteside et al., 2012) were applied, motivated by errorless learning strategies. Figure 5.4 demonstrates these steps based on the "[REFERENT] *like* [THING / PERSON / LOCATION / PROCESS]" constructional schema which started with the source construction "*I like it*", followed by two variations of the pre-core word slot (e.g., "*you like it*") and finally two variations where both open slots were replaced ("*they like flowers*"). The first step of practicing each phrase was listening to a pre-recorded self-voice model (i.e., errorless modelling). Importantly, in this step, participants could listen to a recording several times by pressing a button (Figure 5.4). This was followed by a step where the participant imagined saying the phrase aloud, and only in a third step the target phrase was actively produced. Thus, these three steps reflect a combination of self-voice / social learning factors and errorless learning techniques.

[17] All names of people or places referred to in the data are pseudonyms.

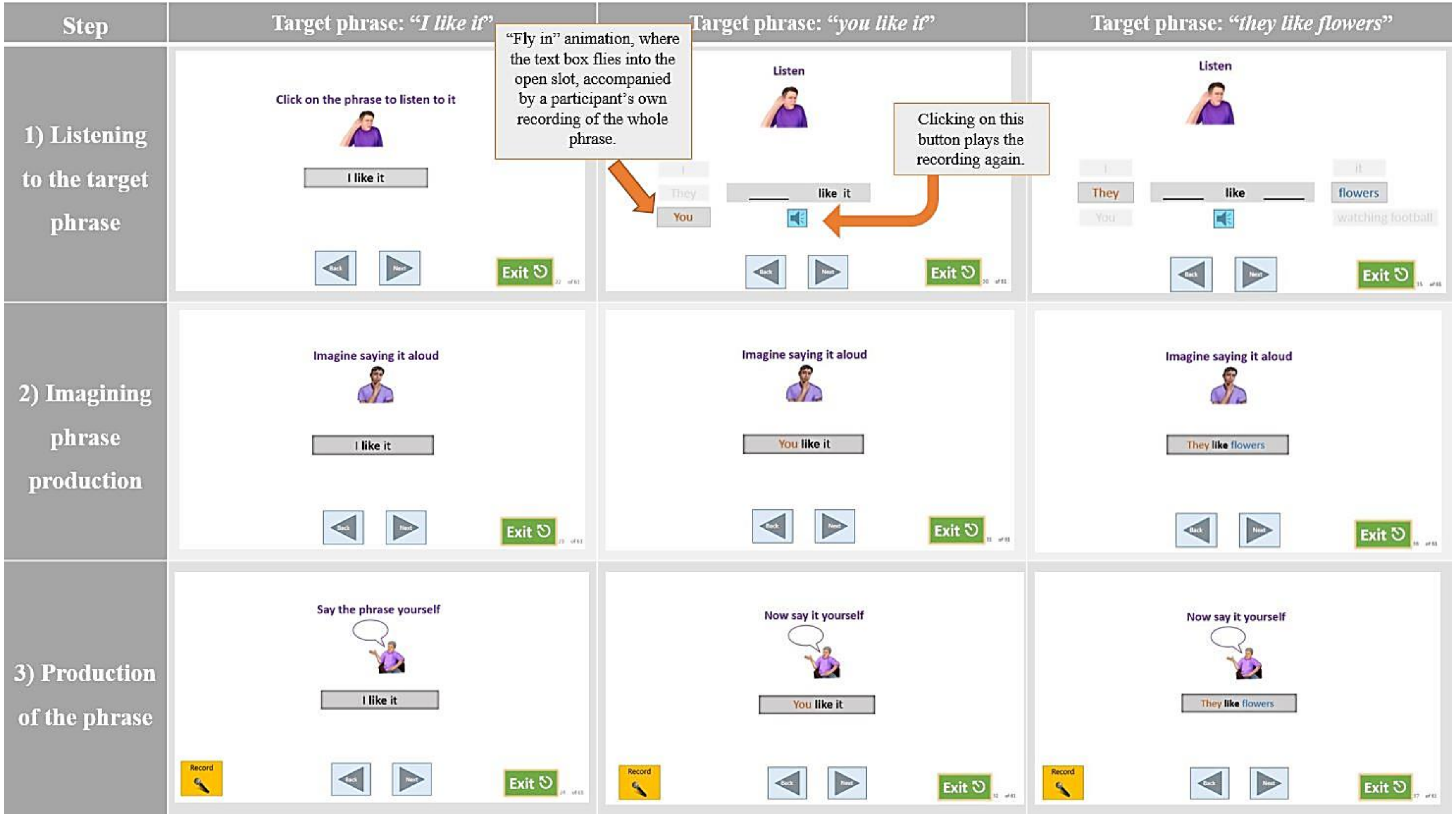


Figure 5.4. Demonstration of practicing variations of the '[REFERENT] like [THING]' construction.

The element of superimposition was further highlighted during the 'listening' step, where variations of a core construction were animated (Figure 5.4). For instance, in the "*you like it*"-'listen'-slide, the animation started with the text box "______ *like it*", accompanied by a sound file where the open slot was filled by a beep ("*beep* *like it*", created with Audacity). Next, the pronoun "*you*" option was flown into the open slot, accompanied by the sound file of the whole phrase "*you like it*". The same procedure was applied to all 'listen' slides where one or both open slots were replaced with lexical items (Figure 5.4).

Each week of Phase 3 consisted of two one-to-one sessions that focused on a single 'topic'. The first session was typically spent on the 'standard form' of a topic (i.e., the pre-recorded phrases from Phase 1), followed by a second session where personally relevant variations of constructions were practiced. For example, one participant decided to practice the following variations of the "[REFERENT] *like* [THING / PERSON / LOCATION / PROCESS]" construction: "*We like red wine / caravans / cruise*" and "*I like CDs / Jake Bugg / Elton John*". The researcher assisted in identifying these individualized lexical items and modelled the whole phrase for the participant to repeat. Again, fluent versions of each of these individualized phrases were created using Audacity, and sound files were incorporated into the PowerPoint program. The individualized phrases could be practiced on the participant's laptop in a self-managed way throughout Phase 3. Laptops were withdrawn in week 10, when reassessment was carried out.

5.2.4. Outcome measures

The following sections describe primary and secondary outcome measures, the control measure and questionnaires to assess acceptability of the intervention. Table 5.4 provides an overview of when each measure was administered.

Table 5.4: Overview of outcome and control measures

	Baseline			Intervention	Post-intervention probes					Variable(s) of interest
Week	**1**	**2**	**3**	**4-9**	**10**	**11**	**12**	**13 - 15**	**16**	
Primary outcome measures										
Dinner Party [1]	✓				✓				✓	Combination ratio; bi- and trigram t-scores
Jogging [1]			✓		✓				✓	
Spontaneous speech [2]	✓		✓		✓				✓	
Secondary outcome measures										
Story Completion Test [3]	✓		✓		✓				✓	'Expected answer' score; Number of well-formed utterances
TROG-2 [4]		✓			✓					Number of items correct
AIQ-21 [5]			✓						✓	Participant ratings
Control measure										
Synonym matching task [6]	✓		✓		✓				✓	Number of items correct
Other measures										
Acceptability questionnaire								✓		Participant ratings and comments
Conversation videos	✓	✓	✓		✓	✓	✓		✓	Not analysed as part of the current report

[1] (Fletcher & Birt, 1983) [2] questions to elicit personal narratives; [3] adapted from Goodglass et al., 1972; [4] Test for Reception of Grammar – Version 2 (TROG-2, Bishop, 2003); [5] Aphasia Impact Questionnaire (AIQ-21, available from https://www.aiq-21.net/); [6] shortened form of a written Synonym Matching Task, taken from the Action for Dysphasic Adults Auditory Comprehension Battery (A.D.A., Franklin et al., 1992).

5.2.4.1. Primary outcome measures

Personal and picture-based narratives served as primary outcome measures to assess participants' language production at the multiword level. Narratives were administered at multiple baselines (weeks 1 and 3), once after the intervention (week 10) and after a 6-week no treatment period (week 16) to assess the stability of any communicative change. All narratives were audio recorded.

The Dinner Party narrative (Fletcher & Birt, 1983), carried out in weeks 1, 10 and 16, is a black and white cartoon about a dinner party where the main meal, a fish, disappears, because it was eaten by the pet cat. It consists of eight separate scenes. In addition, a second cartoon from the same resource (Fletcher & Birt, 1983) was used, called 'Jogging', depicting a series of events where an overweight man goes running to lose weight, but comes back home soaked and injured because he was bitten by a dog and it started to rain. It is comparable with the Dinner Party narrative in that there are eight scenes and the style of the black and white drawings is similar. The Jogging narrative was administered in weeks 3, 10 and 16. The cartoons were shown to each participant and the instruction was: "*Have a look at these pictures. Together, these pictures constitute a story. Tell me in your own words everything you see going on (whenever you are ready).*"

Personal narratives were elicited using four different questions: a) "*Can you tell me about the last time you went on holiday, or the last trip you took*" (week 1, baseline 1); b) "*Thinking back, can you tell me a story about the most frightening experience you've had? It could be from any time from when you were a kid or more recently*" (week 3, baseline 2); c) "*Can you tell me a story about the most interesting person you've met*" (week 10, post-intervention); d) "*Can you tell me about the best / most recent film you've seen"* (week 16, maintenance). These questions were designed to elicit extended monologues. The researcher was allowed to give general encouragement for more speech (e.g., "*Can you tell me more about X*?").

Narrative outcome variables were frequency-based measures from the FLAT Version 2 (Zimmerer et al., 2018; Zimmerer & Wibrow, 2015). The main outcome variable was combination ratio which quantifies the amount of a participant's connected speech by dividing the number of trigram tokens by number of words.

Higher combination ratio indicates a higher proportion of trigrams in a speaker's connected speech. Secondary variables were bi- and trigram t-scores. These measures quantify the degree of association strength of word combinations and point to the degree of creativity of connected speech. Higher t-scores suggest combinations which may represent (parts of) familiar collocations, lower t-scores point to more weakly collocated combinations, i.e., more flexible and creative language production. For instance, the trigram "*I said it*" has a t-score of 15, while the more creative variation "*you said yes*" has a t-score of 3. For t-scores, type-based values were used, as they point to the inventory of constructions rather than the usage. By contrast, token-based values are biased by repeated utterances (constructions such as "*I'm sorry*" that may be overused by a participant). Since several samples were collected at each time point (e.g., one picture-based and one personal narrative at baseline 2, two picture-based and one personal narrative at post-intervention assessment), average combination ratio and t-score averages per assessment point were calculated across each participant's samples.

5.2.4.2. Secondary outcome measures

Secondary outcome measures were an adaptation of the Story Completion Test (Goodglass et al., 1972), the TROG-2 (Bishop, 2003) and the Aphasia Impact Questionnaire 21 (AIQ-21; Swinburn et al., 2018; https://www.aiq-21.net/).

The Story Completion Test was originally developed by Goodglass et al. (1972) to study the availability of abstract syntactic structures in NFA. In the current study, it was adapted to probe a participant's ability to produce the constructions targeted in the intervention. It was administered at multiple baselines (weeks 1 and 3) and twice during the post-intervention phase (weeks 10 and 16), and was audio recorded. The task with scoring examples is shown in Table 5.5. As in Goodglass et al. (1972), the researcher presented a short story orally, where the final sentence or phrase was missing, as in the following example (item 4): "*I'm fond of ball games. My friend asks me: 'What are your thoughts on football?', so I say...*?". Seven out of 12 story situations ended with "*..., so I say…?*" / "*…, so he says…?*", and participants attempted to complete each story with a single sentence.

Table 5.5: Study-specific Story Completion Test including scoring examples

	Scenarios [target phrase]	Answer	Score (out of)		G
P1	My friend comes in. I want him **sit down**, so I say…? [Sit down]	*Sit down! (laughing)*	✓	*(34)*	
P2	My cousin is at the door. I want him to **come in**. I open the door and say…? [Come in]	*Hello, come in! Sit down, chill!*	✓	*(88)*	
1	My friend looks for his keys. He asks me if **I know** where they are, **but I don't**, so I say…? [I don't know]	*Sorry, I don't know.*	39	***(39)***	*1*
2	I broke my arm last year, and **it has healed well**. The doctor asks me, "**How is your arm?"**, so I say…? [It's alright]	*Brilliant. It's really good now.*	*100*	***(100)***	*1*
3	**I baked a cake**. My friend asks: "Did you **buy** this cake?", so I say…? [I made it]	*er…I-no? I make it myself.*	85	***(85)***	*1*
4	**I'm fond of ball games**. My friend asks me: **"What are your thoughts on football?"**, so I say...? [I like it]	*Yeah it's really good. But I ('ll) no play it though.*	*0*	***(71)***	*1*
5	Tom **swore** in class. The teacher **heard** and asked Marc "Marc, did **you** say this?", so Tom **confessed** and admitted…? [I said it]	*No, sorry, it was me.*	*0*	***(7)***	*1*
6	**I detest spicy food**. My friend asks **how I find spicy food**, so I say…? [I hate it]	*It's it's too hot. I can't take it.*	*0*	***(49)***	*1*
7	My husband looks for the **last piece of the cake**, but **I ate it** for lunch. He asks "**where's the cake?**", so I say…? [I had it]	*Oh, I don't know, it's it's erm it's my mice coming (and) eat it up. (laughing)*	*0*	***(76)***	*1*

8	Sarah was **looking for a holiday destination** and saw a **lovely picture of a beach** in the brochure. She had a **smile** on her face and thought to herself…? [I want to go (there)]	*I'm bookin' (it) now. (sighs)*	*0*	***(27)***	*1*
9	The woman heard that her **flight** was **delayed** and was **keen** to **find out more** about the delay. She thought to herself…? [I want to know (what's going on)]	*I'm going to the…the: erm ts (again), oh my god there is er really weird is I, I (sighs) erm. d'y know, I phone (laughing) to er and why is er is the plane, what's going on? (sighs)*	*0*	***(7)***	*1*
10	Jane **can't find her coat**. Her mother has just cleaned the flat. She knows her mother put it somewhere. Jane wants to know **where the coat is**, so she **asks** her mum…? [Where is my coat]	*Where's the coat, mum?*	*71*	***(71)***	*1*
11	John can't remember the **day** of the football match, so he **asks** his friend…? [When is the match]	*What day is the football match?*	*100*	***(0)***	*1*
12	Paul lives in Bristol. He **spent the weekend in London**. He wants to tell his friend **what he did on the weekend**, so he says…? [I went to London]	*Gosh er I went football, I want to the… the play, is erm…erm Covent Garden, everything is... really good.*	*76*	***(76)***	*1*
		Total:	*395*	*/ 708*	*12/12*

Scoring examples are taken from P1's maintenance probe; ***P1 / P2*** *= Practice items;* ***Answer:*** *P1's answer;* ***Score:*** *for every expected construction;* ***G:*** *assign 1 for every grammatically well-formed utterance;* "*I don't know*": target phrase for item 1, did not count as a grammatical utterance unless it was produced for item 1; "*(I'm) sorry*" did not count as a grammatical utterance unless it was produced for items 5 and 7, depending on intonation; "*No way*" was accepted as a grammatically well-formed utterance for items 3 and 6, where it was judged as an appropriate utterance given the semantic context.

There were two categories for scoring participants' answers, an 'expected answer' category and a 'grammatically well-formed utterance' category (Table 5.5). As an example, for item 4 (see above), where the expected construction was "*I like it*", the answer "*it's fun*" would be scored with a 0 in the 'expected answer' category, while a 1 would be noted in the 'grammatically well-formed utterance' category. The scoring procedures are described in more detail in the following paragraphs.

Although story situations were designed to elicit trained constructions within highly predictable sentential contexts, a norming sample of 41 native speakers of English revealed that the probabilities with which the target constructions were produced varied considerably across items (range: 7-100%). The 'expected answer' scores were therefore weighted according to the normative cloze probability. For instance, if an item elicited a normative cloze probability of 39% (i.e., 39% of the 41 neurotypical speakers answered with the target construction), the score a participant would receive for producing an expected construction for that item was 39, while a normative cloze probability of 100% would mean that a score of 100 would be assigned to that item if the participant answered with the target construction. The score (e.g., 39 for item 1; 100 for item 11) was given for the target construction or a construction with the same grammatical structure (e.g., item 4: "*I like it*" / "*I enjoy football*"), and a score of 0 was assigned if the answer reflected another structure (e.g., "*it's fun*" instead of "*I like it*").

Similarly to Goodglass et al. (1972), in addition to the 'expected answer' score, described above, the number of grammatically well-formed utterances was recorded in the following way. For every answer reflecting a grammatically well-formed utterance (disregarding semantics), a score of 1 was given. Ellipsis was allowed (e.g., "*fantastic*" instead of "*I like it*"; "*fixed*" instead of "*it's alright*"). Formulas such as "*oh dear*", "*dear me*", "*oh gosh*" and "*come on*" were not scored as grammatically well-formed utterances. Other formulas ("*I don't know*", *"(I'm) sorry*", "*no way*") however, counted as a grammatically well-formed utterance for certain items (Table 5.5).

The TROG-2 (Bishop, 2003) was administered once before and after the intervention (weeks 2 and 10), to investigate whether there was change a participant's spoken sentence comprehension. Additionally, the AIQ-21 (Swinburn et al., 2018;

https://www.aiq-21.net/) was used as a participant-reported measure of communication abilities, emotional state / well-being, and participation. It was administered once during the baseline- and the post-intervention phase (weeks 3 and 16).

5.2.4.3. Control measure

A shortened form of a written Synonym Matching Task, taken from the A.D.A (Franklin et al., 1992), was created as a control measure. This was administered over multiple baselines (week 1 and 3), and twice during the post-intervention phase (weeks 10 and 16). In this task, the participant judges whether pairs of written words (e.g., "*grin*" and "*smile*") have a similar meaning or not, by ticking 'yes' or 'no' for each pair. The original task includes a set of 40 high- and 40 low-imageability, as well as 40 high- and 40 low-frequency word pairs (160 pairs in total). The revised version consisted of 40 out of the 160 items, with 50% representing 'same' and 50% 'different' judgements. For the subset, 16 low- and 4 high imageability word pairs (e.g., "*clumsy – frail*" versus "*mail - post*"), plus 16 low- and 4 high frequency (e.g., "*couch – sofa*" versus "*lunch – river*") word pairs were chosen (Appendix C.6). This test was devised to assess written word processing which was not directly targeted in the intervention, and it was expected that a higher proportion of low frequency/imageability word pairs would elicit a performance above chance and below ceiling, i.e., would make the task more suitable as a control measure.

5.2.5. Acceptability of the intervention

Acceptability of the intervention to both participants with NFA and their CPs was investigated through post-intervention study-specific questionnaires (Appendices C.7-C.8). These were designed to capture the views of participants regarding different aspects of the intervention. Questions for participants with NFA focused on a) helpful and unhelpful elements of the intervention; b) the overall usefulness of the intervention; c) the level of difficulty of practicing the WMG and activities at home; d) the frequency with which they use the trained constructions, and e) self-reported change in communicative abilities. The questions for CPs were similar, focusing on perceived helpful and unhelpful elements of the intervention, its overall usefulness for their family member or friend with aphasia, and any observed changes in the person with aphasia's communicative abilities after the intervention. Open-ended questions allowed both participants with NFA and their CPs to add further comments on the intervention in general. Participants filled in the questionnaires in their own time without the researcher present, and participants with NFA were encouraged to fill in the questionnaire together with their CP. To reduce participant bias, questionnaires were posted back to the researcher's primary supervisor.

5.2.6. Conversation videos

During the baseline- and post-intervention phases, participants videotaped weekly 10- to 20-minute conversation samples (weeks 1, 2, 3, 10, 11, 12, 16) with a regular CP (Table 5.4). All videos were recorded in the participant's home, with no persons other than the participant and their CP present. The instructions were to video record 20 minutes of conversation in total each week, and participants could choose topics themselves. The project supplied digital video cameras for this purpose.

5.2.7. Data analysis

Audio-recorded connected speech samples (personal and picture-based narratives) were orthographically transcribed by a student research assistant blinded as to the sample collection point. Pre-, post-intervention and maintenance data were analysed individually for each participant, using the frequency-based analysis tool FLAT, Version 2 (Zimmerer et al., 2018; Zimmerer & Wibrow, 2015). Before performing the frequency-based analysis, the researcher checked all transcriptions for accuracy and formatted them in line with FLAT transcription conventions (study 1).

Since the group represented a relatively small, heterogeneous sample of individuals with NFA, all results were evaluated at the individual rather than the group level. However, group means were plotted for the main outcome variables and the control measure, to identify patterns across individuals. Individual recurrent phrases (e.g., "*oh dear me*" or "*the work*", frequently produced by P4) were included in analysis of spontaneous speech samples as each participant acted as their own control. For primary outcome measures, the mean combination ratio and type-based t-scores were compared within participants across the four assessment points (baseline 1, baseline 2, post-intervention and maintenance) using descriptive statistics. Because of the novelty of these FLAT measures, reference values from a normative sample were reported for each variable, taken from Zimmerer, Coleman, Hinzen, & Varley (in prep). Zimmerer et al. (in prep) analysed personal and picture-based narratives ('Last

Holiday' spontaneous speech samples and 'Dinner Party' narratives) from of 30 neurotypical adults with the FLAT Version 2.

Performance on secondary outcome measures was analysed by comparing pre- and post-intervention scores for each participant. For the adapted Story Completion Test, McNemar tests for related samples were conducted individually to investigate whether baseline performance was stable (baseline 1 versus 2). Where baseline performance was found to be stable, Cochran's Q tests were used to examine each participant's performance prior to and following the intervention (baseline 2 versus post-intervention versus maintenance). For these comparisons, the final baseline measurement was used as the pre-intervention measure as participants were more familiar with the tasks and the researcher at the second probe compared to baseline 1. The same analysis was applied to the control measure. Conversation videos were not analysed as part of this thesis.

5.3. Results

All five participants completed the intervention study. Results are reported measure by measure, evaluating individual change patterns after intervention.

5.3.1. Control measure: Synonym matching task

The results of the synonym matching task, administered at multiple baselines (weeks 1 and 3) and twice during post-intervention probes (weeks 10 and 16), detected relatively stable performance across participants (Figure 5.5). Baseline scores were stable (as indicated by McNemar tests, P1: $p = .375$; P2: $p = .625$; P3: $p = .625$; P4: $p = .754$; P5: $p = .500$). When comparing baseline 2, post-intervention and maintenance using Cochran's Q tests, there was no significant change for any of the participants' scores (P1: $Q = 1.2$, *ns*, $p = .549$; P2: $Q = 1.0$, *ns*, $p = .607$; P3: $Q = 2.0$; *ns*, $p = .368$; P4: $Q = 2.8$, *ns*, $p = .247$; P5: $Q = 3.5$, *ns*, $p = .174$).

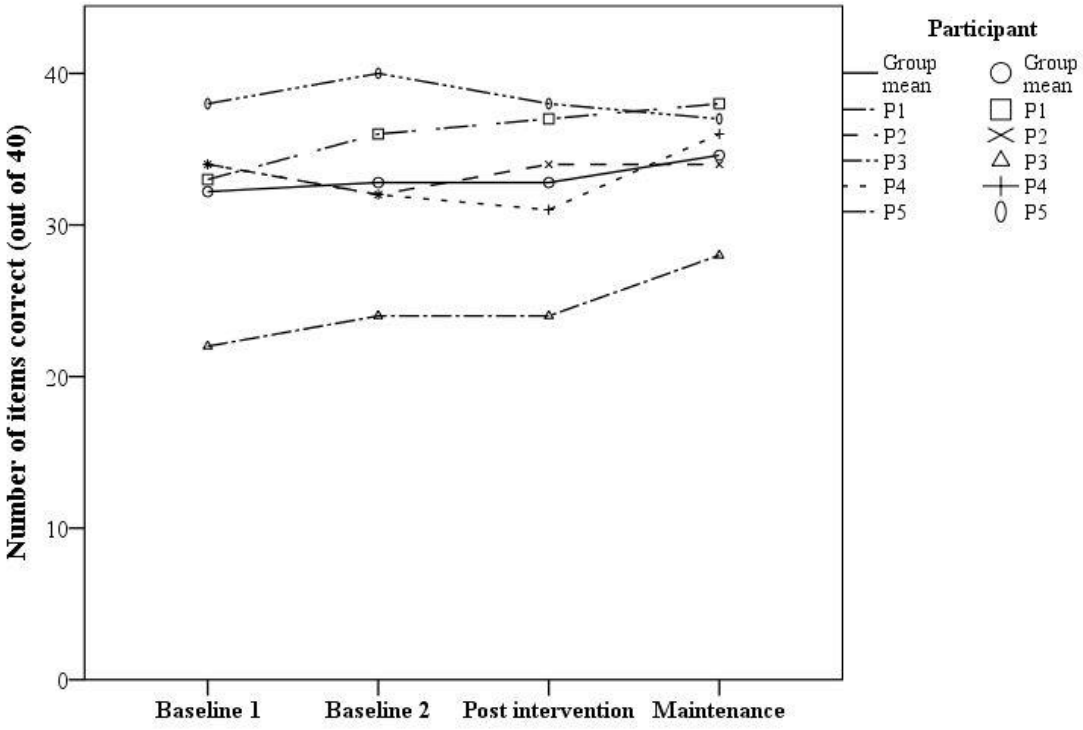


Figure 5.5. Synonym matching task – number of correct items by participant over time.

P3's pattern, however, stood out in that he performed just above chance level at baseline 1, baseline 2 and post-intervention (22, 22 and 24 items correct, respectively), as he had difficulty with single-word reading. Whenever he could not read a word in a pair, he ticked 'no' (indicating that the words of a pair have a different meaning), whereas he only ticked 'yes' if he could read both words of a pair and judged them to mean something similar. His performance at maintenance, with 28 correct items, seems to have improved, compared to previous attempts. Further analysis of correct 'yes' (same) judgements revealed that these judgements increased from 4 at baseline 2 to 6 post-intervention and 13 at maintenance. Thus, his relatively good performance at maintenance could be explained by the increased number of correct 'same' judgements which needs to be interpreted in light of his strategy, outlined above (that he only ticked 'yes' if both words could be read and were judged to mean something similar). This may point to improved single word reading in P3, following intervention.

5.3.2. Narratives

For the primary outcome variable, combination ratio, it was expected that the intervention would lead to increased connected speech, i.e., higher values during post-intervention probes. While within-participant comparisons indicate an increase in combination ratios between baseline 2 and the first post-intervention probe for four participants (P2, P3, P4, P5), these differences did not exceed the baseline variation in three of these participants. For one participant (P5), however, the increase between baseline 2 and post-intervention values was higher than the baseline variation and remained stable across both post-intervention probes. Figure 5.6 displays participants' scores over time as well as a group mean for each assessment point.

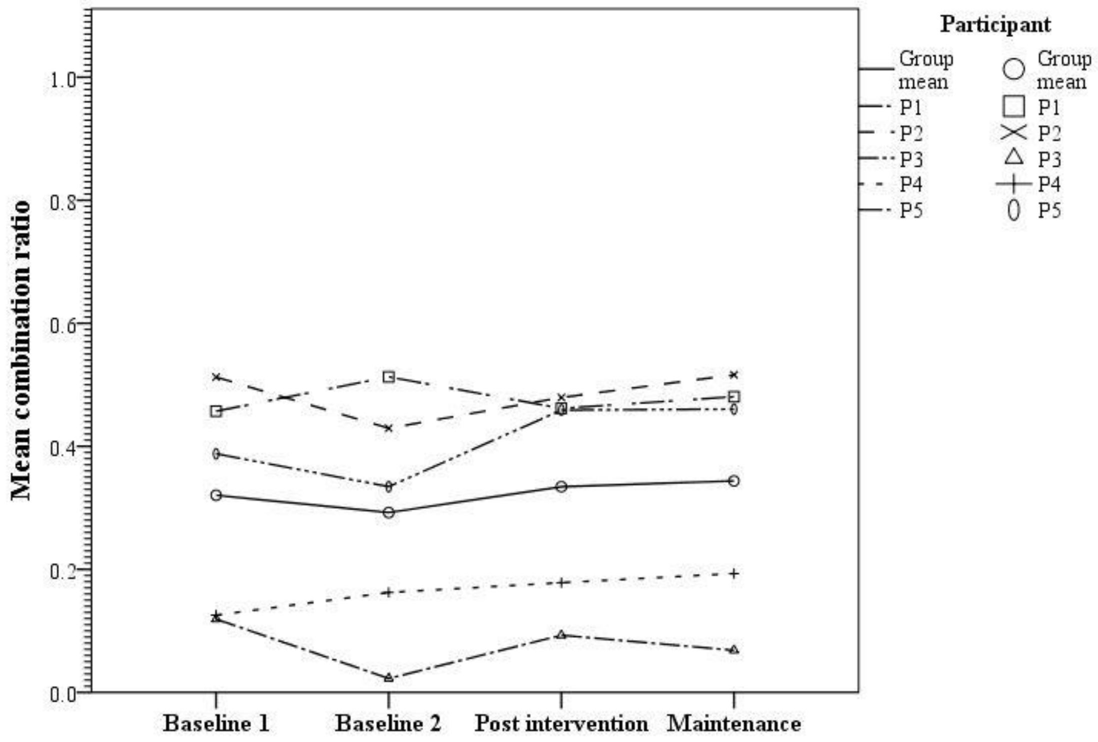


Figure 5.6. Mean combination ratio by participant over time.

P1's combination ratio was .46 and .51 at baseline 1 and 2, respectively, dropping back to .46 after intervention, with a value of .48 at maintenance. Thus, over time, P1's amount of connected language appeared relatively stable. P2 showed more variability before the intervention, with .51 at baseline 1 and .43 at baseline 2. Post-intervention, the combination ratio was .48, and the highest value of .52 was observed at maintenance, indicating a slight increase in connected speech across post-intervention assessment points. However, the increase of .08 between post-intervention and maintenance was not greater than the variation observed between baseline 1 and 2. P3's combination ratio varied considerably before intervention, with .12 at baseline 1 and .02 at baseline 2. After intervention, values were .09 and .07 (maintenance), which, however, did not exceed the average combination ratio at baseline 1. For P4 and P5, results provided some evidence for positive treatment effects: P4's average combination ratio showed an increase from baseline (.13 and .16 at baseline 1 and 2, respectively) to post-intervention (.18) and maintenance (.19). However, the increase from baseline 2 to maintenance (.03) was not greater than the pre-intervention variation. The pattern observed for P5 shows an increased average combination ratio after the intervention which exceeded baseline variation, and this increased combination ratio remained stable across both post-intervention probes: compared to .39 at baseline 1 and .33 at baseline 2, values increased to .46 at post-intervention and maintenance.

As a comparison, normative combination ratio values, taken from Zimmerer et al. (in prep) were .73 (SD = .05) for 'Last Holiday'- and .79 (SD = .05) for 'Dinner Party' narratives. As shown in Figure 5.6, all five participants presented with combination ratios below these values, indicating less connected speech compared to neurotypical speakers. While these results confirm the limited ability of participants to create word combinations, with lower combination ratios as compared to normative controls, they point to enhanced connected speech as a result of intervention for P5.

Turning to secondary outcome variables for connected speech, lower bi- and trigram t-scores were expected following intervention as compared to baseline performance, since lower t-scores point to more flexible and creative combinations. Thus, lower scores in Figure 5.7 and Figure 5.8 denote progress. The following paragraphs report type-based values, indicative of a participant's inventory of constructions.

An overview of average bigram t-scores by assessment point and participant is shown in Figure 5.7. Overall, a decrease from baseline 2 to post-intervention values was observed for P1, P2, P4 and P5. However, none of the five participants showed the expected pattern which was a relatively stable baseline, a clear decrease of average t-scores between baseline 2 and post-intervention probe, and a relatively stable performance during the post-intervention phase.

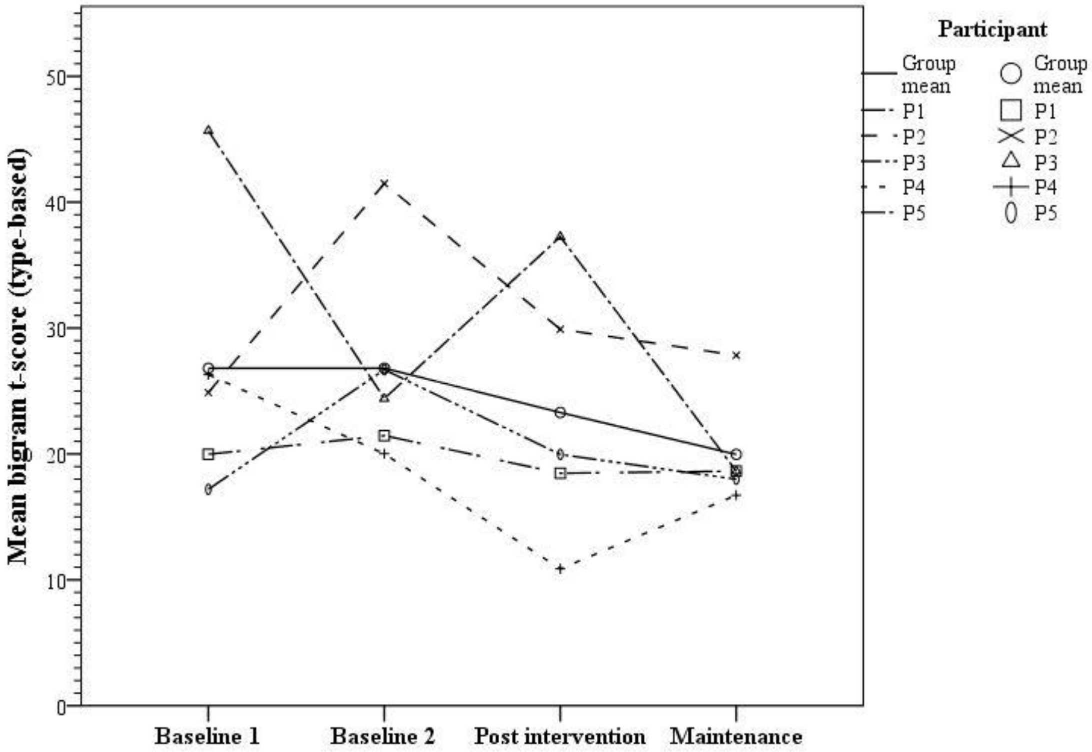


Figure 5.7. Mean type-based bigram t-scores by participant over time.

Normative values for bigram t-scores are 22.3 (SD = 4.57) for ‘Last Holiday’- and 19.29 (SD = 3.17) for ‘Dinner Party’ samples (Zimmerer et al., in prep). For P1, the values appeared relatively stable over time and all values were within the normative range (baseline 1 = 19.97, baseline 2 = 21.45, post-intervention = 18.47, maintenance = 18.63). For P5, all four average bigram t-scores were close to normative values, and while there was considerable variation in baseline t-scores (baseline 1 =

17.19, baseline 2 = 26.69), P5's post-intervention values were similar to the first baseline probe (post-intervention = 19.96, maintenance = 17.99).

P2 and P4 displayed a variable baseline, but showed the greatest decrease in bigram t-scores when comparing baseline 2 and post-intervention probes (P2: baseline 1 = 24.87, baseline 2 = 41.48, post-intervention = 29.90; P4: baseline 1 = 26.30, baseline 2 = 20.01, post-intervention = 10.87). In terms of stability across post-intervention probes, P2's bigram t-scores decreased further by 2.07 (from 29.90 to 27.83), while P4's scores increased by 5.85 (from 10.87 to 16.72).

P3's average bigram t-scores varied most dramatically over time, with 45.68 (baseline 1), 24.40 (baseline 2), 37.25 (post-intervention) and 18.63 (maintenance). This was the only participant where there was an increase in bigram t-scores between baseline 2 and post-intervention. This may have been related to the relatively low overall number of bigram types produced by this participant. As a consequence, P3's t-scores may have been biased towards a small number of combinations with very high t-scores (e.g., "*it's*" with a t-score of 243, "*don't*" with a t-score of 199) or very low t-scores (e.g., P3's picture-based narrative at baseline 2 consisted of three bigram types ranging between t-scores of -15 and 6).

For trigram t-scores, control values, taken from Zimmerer et al. (in prep), range between an average of 8.3 (SD = 1.61) for 'Last Holiday'- and 7.56 (SD = 1.57) for 'Dinner Party' samples. As shown in Figure 5.8, scores decreased from baseline 2 to post-intervention assessment for all participants. Overall, P2's values most closely resembled the expected pattern of more creative trigram combinations following the intervention and stability across post-intervention samples.

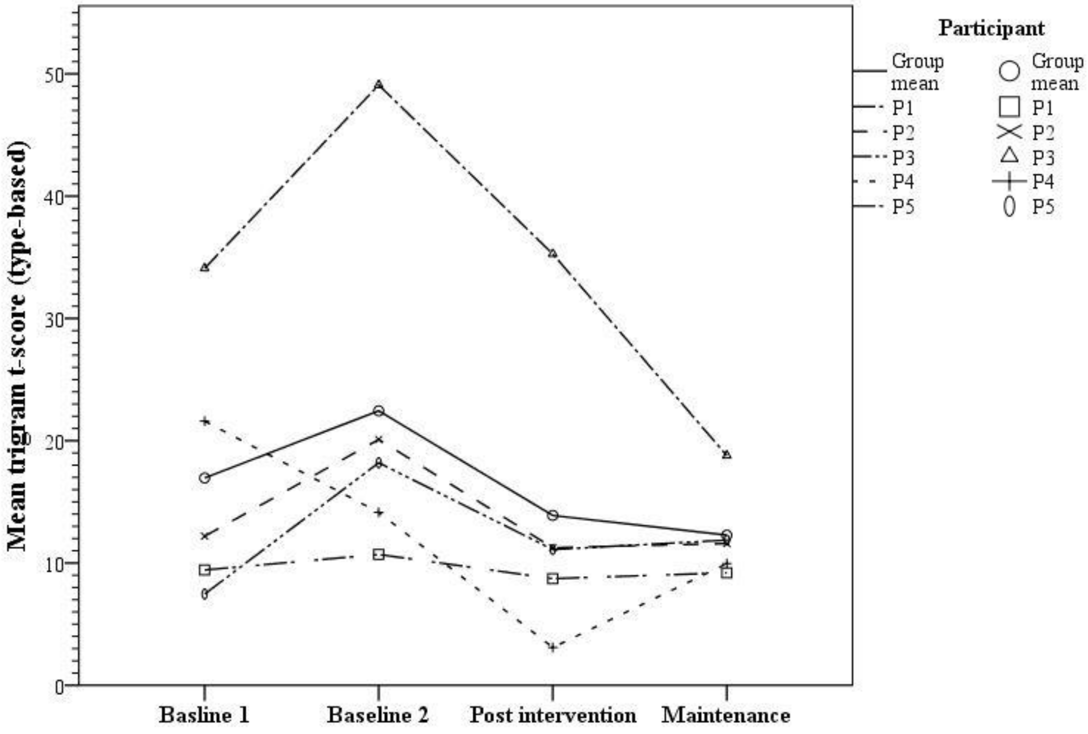


Figure 5.8. Mean type-based trigram t-scores by participant over time.

P1 presented trigram t-scores within the normative range (baseline 1 = 9.43, baseline 2 = 10.70, post-intervention = 8.72, maintenance = 9.20). While there was a slight decrease in t-scores following intervention as compared to baseline 2, P1's values appeared relatively stable over time, mirroring her performance in other frequency-based measures (see combination ratio and bigram t-scores). For P2, P3, P4 and P5, there was a more marked decrease in trigram t-scores between baseline 2 and the post-intervention samples. However, there was variability in all four participants' baseline values (Figure 5.8). Baseline 1 and 2 varied between 12.19 and 20.10 (P2), 34.09 and 49.09 (P3), 21.60 and 14.14 (P4) and 7.47 and 18.20 (P5). Only P2's and P4's decrease in average t-scores between baseline 2 and post-intervention probes exceeded their baseline variation (P2: baseline variation of 7.91, decrease baseline 2 versus post-intervention of 8.85; P4: baseline variation of 7.46, decrease baseline 2 versus post-intervention of 11.05). For P2 and P5, the decrease between baseline 2 and post-intervention was stable across post-intervention assessment points: post-intervention and maintenance scores increased only minimally for these two participants (+.35 and +.77, respectively). For P4, on the other hand, the decrease from baseline 2 to post-intervention did not appear stable, with a higher maintenance average t-score of 9.95, as compared to 3.09 at post-intervention assessment.

Again, P3's pattern represented a special case. Compared to others, his language production consisted of more strongly associated trigrams, as indicated by t-scores, at all assessment points. However, he was the only participant showing further decrease of t-scores between post-intervention (35.27) and maintenance (18.78) which may point to consolidation.

5.3.3. Story Completion Test

Performance on the adapted Story Completion Test was analysed separately for scores in the 'expected answers' category and for number of grammatically well-formed utterances. Figure 5.9 shows each participant's overall scores in the four assessment probes. P3 was unable to retrieve any constructions at baseline ("*dunno*" answers for all items, resulting in scores of 0 at baseline 1 and 2), but did so following intervention. As noted for the narratives, participants' raw scores at baseline varied considerably, and only P5's performance after the intervention showed a significant increase compared to pre-intervention scores.

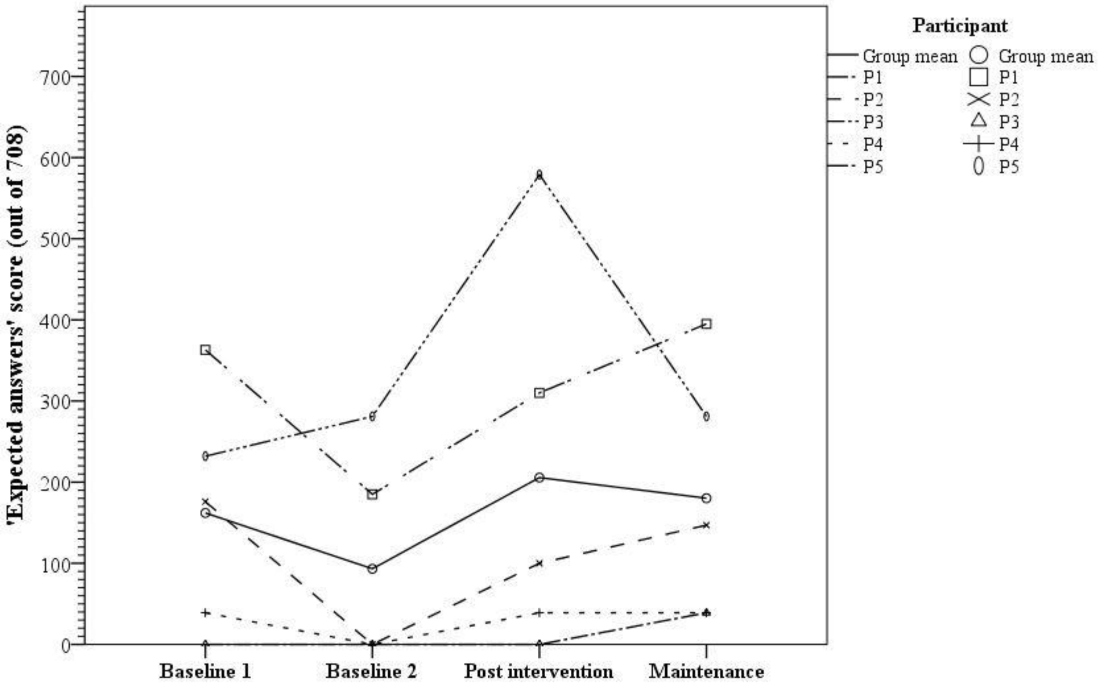


Figure 5.9. Story Completion Test – overall score based on expected answers by participant over time.

To statistically investigate change before versus after intervention, raw scores in the ‘expected answer’ category were converted into a binary scale (1 = expected construction, 0 = unexpected answer) to perform McNemar- and Cochran’s Q tests. McNemar tests revealed no significant difference between baselines 1 and 2 for P1, P2, P4 and P5 (P1: p = .25; P2: p = .50; P3: stable 0 baseline; P4: p = 1.00; P5: p = 1.00). There was a significant difference between P5’s baseline 2 versus post-intervention and maintenance scores, as indicated by Cochran’s Q test, Q = 8.0, p = .018. This result points to an increase in overall score immediately after intervention which dropped back to baseline level at maintenance, as shown in Figure 5.9. For the remaining participants, there were no significant differences (P1: Q = 3.50, *ns*, p = .174; P2: Q = 2.00, *ns*, p = .368; P3: Q = 2.00, *ns*, p = .368; P4: Q = 2.00, *ns*, p = .368). P3 was included here since both baselines resulted in a score of zero, which was regarded as stable.

In terms of grammatically well-formed utterances, participants’ patterns over time are shown in Figure 5.10. P1’s and P2’s proportions of grammatically well-formed utterances increased significantly after the intervention. Although P3 was able to give differentiated responses after the intervention and produced more grammatically well-formed utterances across post-intervention probes, this difference was statistically not significant.

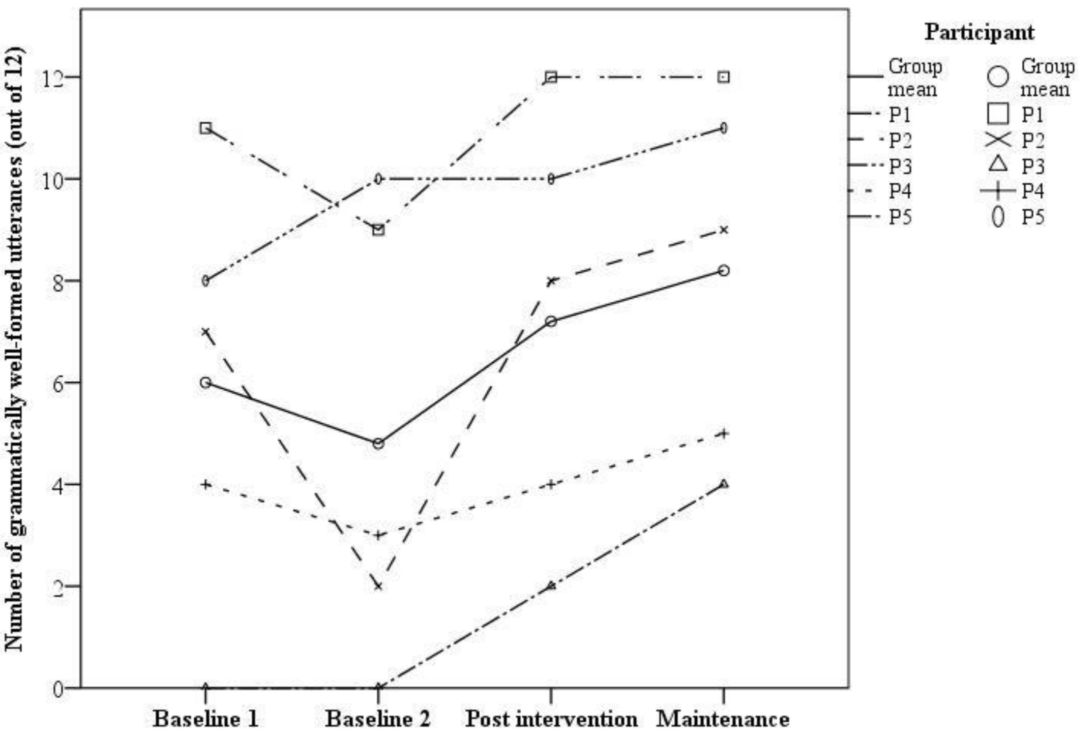


Figure 5.10. Story Completion Test – number of grammatically well-formed utterances by participant over time.

Again, stability at baseline for P1, P2, P4 and P5 was examined using McNemar tests, where non-significant results were found for P1 (p = .625), P4 (p = 1.00) and P5 (p = .500) and a trend toward significance for P2 (p = .063). This indicates that performance at baseline 1 versus baseline 2 was not significantly different for these participants. Note that P3's scores were 0 prior to the intervention (explained above). Change in the proportion of grammatically well-formed utterances following the intervention was analysed using Cochran's Q tests (where baseline 2 was used as the pre-intervention score). While there was no significant change in performance over time for P3, P4 and P5 (P3: Q = 4.8, *ns*, p = .091; P4: Q = 1.2, *ns*, p = .549; P5: Q =

.500, *ns*, p = .779), P1 and P2's proportion of grammatically well-formed utterances changed significantly (P1: Q = 6.0, p = .05; P2: Q = 9.6, p = .008). It should be noted, however, that for both P1 and P2, post-intervention scores exceeded baseline 1 probes by only 1 point. As shown in Figure 5.10, P1 and P2's increase from baseline 2 to post-intervention remained stable at maintenance, or improved further by 1 point (P2).

5.3.4. TROG-2

A comparison of TROG-2 scores before and after intervention (weeks 2 and 10), shown in Table 5.6, revealed a relatively stable performance for P2 (18/28 and 20/28 items correct, respectively). While P4's and P1's sentence comprehension scores decreased by 1 block (-14 and -3 items, respectively), P3's score increased by 2 blocks (+23 items), and P5's score by 3 blocks (+1 item). This tentatively suggests that some participants with NFA may show improved sentence comprehension after this novel intervention.

Table 5.6: TROG-2 scores before and after intervention

Comprehension of spoken sentences (TROG-2)					
Participant	**Pre (week 2)**		**Post (week 10)**		***Change pre- vs post-intervention (number of blocks)***
	Blocks	**Items** (out of)	**Blocks**	**Items** (out of)	
P1	16	75 (80)	15	72 (80)	***-1***
P2	2	18 (28)	2	20 (28)	***0***
P3	6	39 (52)	8	62 (80)	***+2***
P4	5	46 (72)	4	32 (52)	***-1***
P5	12	71 (80)	15	72 (80)	***+3***

5.3.5. AIQ-21

An overview of AIQ-21 ratings (sampled at weeks 3 and 16) is presented in Table 5.7. Findings suggested an increase in perceived well-being and emotional state following intervention. However, with regard to communication, the ratings indicated that for some, there was increased awareness of communication difficulties after intervention.

Table 5.7: AIQ-21 ratings before intervention (week 3) and at maintenance (week 16)

		P1		P2		P3		P4		P5	
		Before	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**	**Before**	**After**
Average ratings	Communication	1.83	1.83	1.83	1.50	2.17	2.33	0.50	0.83	1.67	1.83
	Participation	0.00	0.00	0.75	1.00	1.25	1.25	0.25	0.25	1.50	1.25
	Emotional state / wellbeing	1.36	1.00	2.36	2.36	1.27	0.73	0.55	0.36	1.64	1.55
Difference score	Communication	0.00		-.33		.17		.33		.17	
	Participation	0.00		0.25		0.00		0.00		-.25	
	Emotional state / wellbeing	-.36		0.00		-.55		-.18		-.09	

The AIQ-21 uses a Likert type scale from 0 (most positive rating) to 4 (most negative rating). Difference scores were expected to be 0 or negative, e.g. 1.00 – 1.36 = -.36.

Four out of five participants (P1, P3, P4, P5) showed improved ratings in emotional state / well-being, while there was no change in this area for P2. In the area of participation, no change was detected in three participants (P1, P3, P4). P5, however, showed an improved average rating, and P2's average rating decreased. In terms of communication, three participants (P3, P4, P5) rated their communicative skills more negatively after the intervention, and for P1, there was no change. P2, on the other hand, showed improved ratings.

5.3.6. Difference scores: narratives and Story Completion Test

Despite the mixed findings in the evaluation of outcomes, promising changes were found for some participants after this pilot intervention. Table 5.8 presents an overview of outcomes for frequency-based measures and the main secondary outcome measure, the adapted Story Completion Test. Difference scores were calculated by participant to evaluate communicative change between baseline average and the post-intervention probe, as well as between post-intervention and maintenance. An average baseline score was used for calculating difference scores, as some descriptive measures (e.g., bigram t-scores) pointed to an inconsistent performance across these two assessment points. With respect to a successful response to the intervention, some difference scores were expected to be negative (bi- and trigram t-scores), while others were expected to be positive (combination ratio, Story Completion Test). In Table 5.8, scores that are highlighted in green reflect the expected direction. Since a difference of zero (i.e., no change) between post-intervention and maintenance probes indicated stability, these zeros were highlighted in green.

Table 5.8: Difference scores: baseline average versus post-intervention (B – P) and post-intervention versus maintenance (P – M)

	Narratives						**Story Completion Test**				**Total number of difference scores reflecting the expected direction**
	Combination ratio		**Bigram t-scores**		**Trigram t-scores**		**Expected answers**		**Grammatically well-formed utterances**		
	B - P	**P - M**	**B - P**	**P - M**	**B - P**	**P - M**	**B - P**	**P - M**	**B - P**	**P - M**	
Expected difference score	+	+ / **0**	-	- / **0**	-	- / **0**	+	+ / **0**	+	+ / **0**	
P1	-0.02	0.02	-2.25	0.17	-1.34	0.47	36	85	2	0	7
P2	0.01	0.04	-3.27	-2.07	-4.90	0.35	12	47	3.5	1	9
P3	0.02	-0.02	2.22	-18.62	-6.32	-16.49	0	39	2	2	7
P4	0.03	0.01	-12.29	5.85	-14.78	6.86	19.5	0	0.5	1	8
P5	0.10	0.00	-1.98	-1.97	-1.73	0.77	322.5	-298	1	1	8

B – P = Change baseline average versus post-intervention; P – M = Change post-intervention versus maintenance;

+ *= positive difference score was expected;* **+ / *0*** *= positive difference score or zero was expected;*

- *= negative difference score was expected;* **- / *0*** *= negative difference score or zero was expected.*

As Table 5.8 shows, for P2, most of the comparisons yielded the expected change patterns (nine out of 10 difference scores). P2's difference scores in connected speech variables mostly reflected the expected patterns. One exception was trigram t-scores which did not remain stable after the intervention. The progress in P2's performance on the Story Completion Test which was observed between baseline and post-intervention probes, improved even more from post-intervention to maintenance. Another participant that seemed to have benefitted from the intervention was P5, with eight out of 10 difference scores representing the expected direction or zero. For P5, the increased combination ratio between the baseline and post-intervention probe should be highlighted, as well as the improvement in grammatically well-formed utterances and expected answers in the Story Completion Test. However, these positive changes did not consistently remain stable across post-intervention probes (see Story Completion Test).

P4's language output, throughout the assessment probes, consisted of a large number of recurrent phrases (e.g., "*oh dear me*"). Despite more impaired language output compared to other participants, P4 presented with eight out of 10 expected difference scores. Her connected speech variables showed positive changes following intervention, although some of these changes were not stable across post-intervention assessment points. P4's performance in the adapted Story Completion Test revealed little evidence of availability of target constructions after the intervention, but a slight increase in the number of grammatically well-formed utterances.

P3 and P1 showed the smallest number of expected difference scores (seven out of 10). P3's expressive language was characterized by relatively few bi- and trigrams at all four assessment points. This restricted inventory of word combinations might have resulted in the great variation of connected speech variables across assessment points. On the other hand, P3 showed signs of consolidation in trigram t-scores during the post-intervention period. With regard to the adapted Story Completion Test, P3 could give differentiated responses after the intervention, where the number of grammatically well-formed utterances increased across post-intervention probes.

For P1, difference scores in frequency-based measures did not consistently reflect the expected direction, despite some evidence of expected change in trigram t-scores. This could be explained by frequency-based values (t-scores) within the neurotypical range prior to intervention, which may have limited further progress in these frequency-based measures. P1's outcomes in the adapted Story Completion Test were more positive, with a greater number of grammatically well-formed utterances after the intervention, and some evidence of more expected answers, i.e., better availability of target constructions following intervention.

Finally, Table 5.9 presents participants' average number of uni-, bi- and trigram types and tokens across personal and picture-based narrative samples. It displays the sizes of participants' inventories of words and word combinations for each assessment point. After intervention, there was no clear increase in bi- and trigram use for P1, P3, P4 or P5. However, P2 showed improvement in producing word combinations. He produced an average of 93 and 103 bigram types per sample after the intervention, as compared to 88 and 63 before. In terms of trigram types, he produced an average of 86 and 91 after the intervention compared to 79 and 52 before. This points to an extended inventory of P2's word combinations as a result of the intervention (Table 5.9), mirroring the high number of expected difference scores (Table 5.8).

Table 5.9: Narratives: average bi- and trigram types and tokens produced by each participant

		Unigram tokens	Unigram types	Bigram tokens	Bigram types	Trigram tokens	Trigram types
P1	Baseline 1	215	80	146	114	98	89
	Baseline 2	166	66	120	91	88	80
	Post intervention	210	77	140	106	92	86
	Maintenance	219	81	148	110	104	93
P2	Baseline 1	201	61	142	88	99	79
	Baseline 2	146	48	99	63	65	52
	Post intervention	218	59	152	93	103	86
	Maintenance	231	66	165	103	117	91
P3	Baseline 1	50	26	13	10	6	5
	Baseline 2	45	21	9	6	2	1
	Post intervention	48	26	12	10	5	5
	Maintenance	66	29	14	11	4	4
P4	Baseline 1	104	25	44	22	13	8
	Baseline 2	106	24	45	22	17	14
	Post intervention	100	24	47	23	18	12
	Maintenance	141	24	66	24	27	15
P5	Baseline 1	139	68	77	67	52	51
	Baseline 2	153	73	80	70	51	49
	Post intervention	113	58	73	62	50	46
	Maintenance	119	61	77	68	54	54

5.3.7. Self-managed home practice: WMG

Phase 2 of the intervention, where participants practiced with the WMG, lasted between 9 and 12 days across participants, depending on when the laptop was set up at a participant's home and the date of the next home visit during week 6 (i.e., when the WMG was removed and replaced with Phase 3 activities). During these 9 to 12 days, participants engaged in the WMG as often as they liked, but the recommendation was to practice 'little and often'. As shown in Table 5.10, the frequency with which participants played the WMG varied between 14 and 32 times. Exposure to test trials (where target words were embedded in constructions targeted in the intervention) varied between 149 and 434 trials. P2, P3 and P4 engaged more frequently in these home activities than P1 and P5. It should be noted, however, that P5 practiced on most days, but played few sessions per day which may have been related to technical difficulties experienced by P5 during some trials (as reported by the participant and visible from results files recorded by the program).

Table 5.10: WMG practice overview

	P1	P2	P3	P4	P5
WMG phase duration (in days)	11	12	9	12	11
Number of days practiced	2	8	6	7	7
Number of times practiced	18	31	32	27	14*
Test trials (n)	245	420	434	364	149

** In three of these sessions, there were technical difficulties.*

5.3.8. An exploration of the relationship between intervention dose and outcome

An exploration of intervention dose and outcome was performed. A participant's intervention dose consisted of the number of one-to-one sessions (which was held constant across participants) plus additional self-managed home practice with the WMG as well as Phase 3 activities between weeks 6 and 9. While the WMG automatically recorded how often and how long a participant practiced, the time a participant spent on their home practice in weeks 6-9 could not be recorded by the computer program. To attempt to quantify engagement, the researcher made field notes each week after asking participants whether they practiced, on how many days, and whether they had any problems using the laptop / program.

In order to quantify intervention dose, the number of days each participant practiced with the WMG was regarded as most systematic. This figure was used as an estimate for dose (the number of trials was judged to be less representative due to technical difficulties experienced by P5) and was correlated with pre-post-intervention difference scores of the main outcome variable (combination ratio). To calculate difference scores, the baseline average was used and the difference between this average and the post-intervention probe was determined. A positive difference score suggests more connected speech following intervention.

It was expected that there would be a positive relationship between the number of days that the WMG was practiced and the combination ratio difference score. Figure 5.11 shows the dose-outcome scatterplot, suggesting that there was a positive relationship between these two variables. Due to the small sample size, it is not possible to identify whether this resembles a linear relationship.

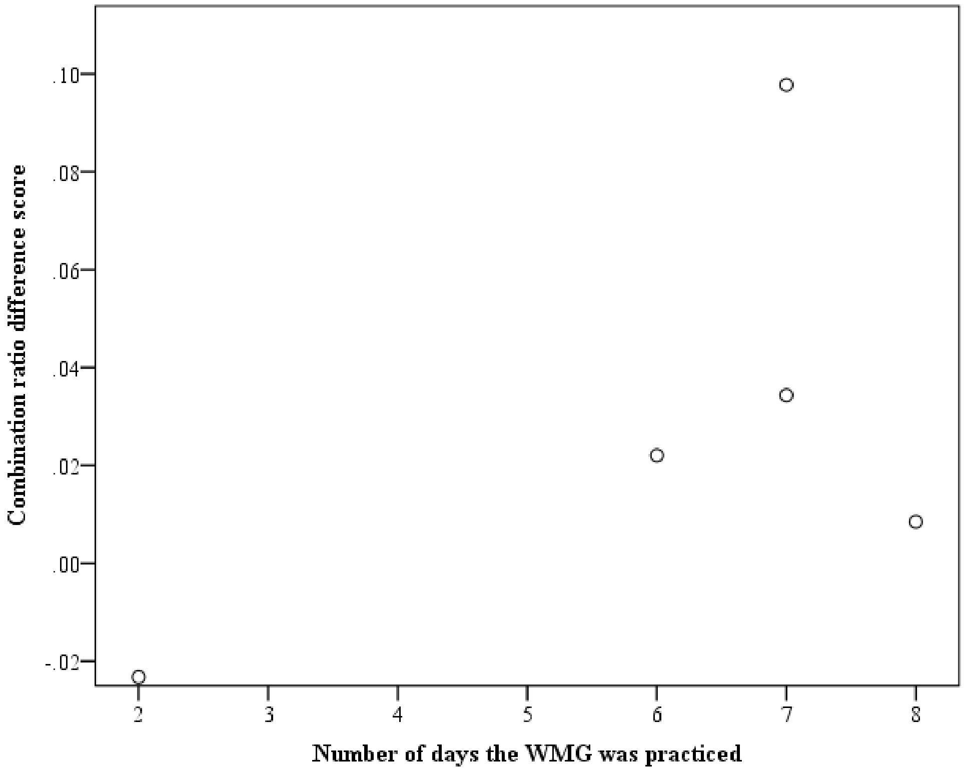


Figure 5.11. Exploration of the relationship between intervention dose and outcome.

5.3.9. Acceptability

5.3.9.1. Feedback from participants with aphasia

The feedback forms for participants with NFA included a list of intervention elements. Participants were asked to tick which of these they found particularly helpful (Appendices C.7-C.8). All five participants selected 'one-to-one sessions to practice speaking', four participants selected 'practicing speaking at home'; 'making recordings of own voice and listening back to them'; and 'three steps to practice speaking'. One participant also gave critical feedback, evaluating on what was not helpful. This included comments such as "*too posh*" for 'making recordings of my own voice and listening back to them', and "*too vague*" as a comment for 'playing the game on the laptop', as well as constructive feedback for future refinement of the intervention: "*home therapy – a microphone to make your own recordings*" (which was not a feature of the current intervention).

The overall usefulness of the intervention was rated on a scale from 1 ("*not at all helpful*") to 5 ("*very helpful*"). The resulting average rating was 4.7. Ease of playing the WMG at home was rated from 1 ("*not at all easy*") to 4 ("*very easy*"). The average rating was 3.2, with two participants circling '4', two '3' and one participant '2'. The same scale was used for a question about doing home practice in weeks 6-9. Three participants indicated a '4', one a '3' and one did not provide a rating, resulting in an average of 3.75, indicating that the home practice during Phase 3 was perceived slightly easier than the WMG during Phase 2. When asked how often participants used the constructions practiced in the intervention, the answers included "*every day*" (one participant) and "*a few times a week*" (four participants). This is suggestive of the everyday relevance of the linguistic material practiced in the intervention. In terms of self-reported changes in communication, one participant circled the statement "*my speaking got **much** easier*", while the remaining four participants circled "*my speaking got **a bit** easier*". Some qualitative comments pointed to a short duration of the intervention: "*it's all too brief*"; "*more one to one sessions speaking*".

5.3.9.2. Conversation partner feedback

CPs were asked to rate the overall usefulness of the intervention for their friend / family member with aphasia. Answers were mainly positive, with two CPs selecting "*very helpful*" another two "*a little helpful*", and one "*not very helpful*". This resulted in an average of 4.0 which was lower than the average of ratings from participants with NFA (4.7, see above), suggesting that participants with NFA perceived the intervention as more helpful compared to their CPs. Qualitative comments about elements that CPs found particularly helpful included one-to-one sessions and practicing at the computer. One CP wrote "*taking time to reflect before answering was useful, but I'm not sure how much of that will be a permanent change*". CPs reported that, after intervention, the expressive abilities of their friend or family member with aphasia "*got a bit easier*" (two CPs) or "*no change*" (three CPs).

5.4. Discussion

The present investigation piloted a novel computerised intervention to enhance the connected speech of individuals with NFA. The intervention was motivated by usage-based principles, training semi-fixed constructions with high functional value by making use of superimposition. Moreover, the three intervention phases employed psycholinguistic and neuroscientific learning elements (structural priming, errorless learning, and increased dose through home practice) and self-voice, a social-motivational learning component. Findings from a case series of five individuals with NFA revealed promising changes in some participants which encourage further development of the intervention program and future testing of its effectiveness using a larger trial.

5.4.1. The current intervention in the context of previous approaches to NFA

The focus of the current intervention on practicing flexible use of functionally relevant constructions by inserting lexical units makes it similar to elements of existing approaches for NFA such as REST (Ruiter et al., 2010; Schlenck et al., 1995; Springer et al., 2000) or Carragher's hybrid intervention (Carragher et al., 2015). While Carragher et al.'s intervention used action verbs and a lexical insertion element, motivated by mapping therapy (where participants were encouraged to find similarities between the words that can be inserted into open slots such as the 'agent' slot), the present intervention used overall semantic-pragmatic meaning of constructions when inserting lexical items. REST was designed to train a compensatory elliptical style. The participants of the present study all had chronic aphasia, and so compensation is a valid principle to target. Practicing more flexible use of high frequency, residual utterances (e.g., "*I don't know*"), combined with techniques that facilitate building novel utterances (superimposition) can be viewed as a compensatory strategy, and the present intervention therefore overlaps – to a certain extent – with REST.

The core words of constructions often represented semantically light verbs, e.g., "*made*" or "*had*", instead of semantically more specified verbs such as "*baked*" or "*ate*". The advantage of using light verbs is that they are usually of high frequency, less constrained by semantic context (Carragher et al., 2013) and can therefore be applied to more conversational topics. All 12 target constructional schemas mapped onto a specific semantic-pragmatic meaning, a key tenet of usage-based grammar.

The present intervention made use of fluent versions of participants' production of phrases. Recording and playing back a participant's language production is a vital element of SentenceShaper® (Linebarger, 2015; Linebarger et al., 2008). SentenceShaper® is a program that was designed to overcome working memory limitations by allowing a user to listen back and re-order parts of sentences. In contrast, the present intervention incorporated fluent versions of target phrases both as a social-motivational component and as part of an errorless learning technique that has been found effective in a previous intervention study (Whiteside et al., 2012).

5.4.2. Outcomes and outcome measures

This study, for the first time, used FLAT variables to evaluate outcomes of aphasia intervention. To address the first research question, "Is there evidence that after intervention participants with NFA demonstrate enhanced connected speech, as measured by a higher proportion of multiword utterances in narratives", results were evaluated at the individual level (i.e., combination ratio was compared for each participant across the four assessment points). P5 showed remarkable change in combination ratio, with more connected speech produced after intervention, and performance stayed stable across post-intervention assessment points. P2 and P4 also showed change in the expected direction, but their pre-post intervention difference scores were smaller compared to P5. This is indicative of the intervention's potential to enhance the ability to combine words into well-formed utterances, at least for a subset of participants. In addition, group means showed a promising increase of combination ratio values across participants.

With regard to type-based t-scores, trigrams, in comparison to bigrams, seemed to reveal a clearer picture of intervention outcomes. For trigram t-scores, the expected decrease in post-intervention t-scores as compared to baseline 2 could be observed for all participants. Lower t-scores indicate more creative word combinations. Although P2's and P4's lower trigram t-scores after intervention stood out from the group, there was little evidence that this decrease was retained across post-intervention assessment points; for P4, the maintenance probe was more similar to baseline level. P3 was the only participant whose trigram t-scores showed further change in the expected direction during the post-intervention phase. This may be indicative of consolidation. Another finding was that, if a participant's word combinations in the baseline phase were characterised by t-scores within the normative range (P1), little or no change after the intervention was detected. Thus, in such cases, the usage-based intervention may not be beneficial to enhance connected speech, or t-scores may not be sensitive measures to capture change in expressive abilities.

FLAT variables do not answer the question of whether participants showed an increased use of trained constructions after intervention. To address this question, performance on the adapted Story Completion Test was analysed. While the results showed that availability of target constructions was enhanced in four participants at a descriptive level, these gains were only significant for one participant (P5). However, this result was not stable across post-intervention assessment points. There was more evidence for positive change with regard to the number of grammatically well-formed utterances, where two participants (P1 and P2) showed significant gains following intervention. The results of the Story Completion Test and its suitability to detect change after intervention should also be considered in light of P3's performance. This participant was unable to retrieve any answers other than "*dunno*" prior to the intervention, but did so after the intervention, where he showed promising gains, especially with regard to grammatically well-formed utterances.

The limited outcomes on the 'expected answers' category of the adapted Story Completion Test might be related to the variable cloze probabilities that the items elicited in the norming sample of 41 neurotypical adults. Despite adjusting the scoring procedure to account for this, 'expected answer' scores fluctuated considerably across the baseline and post-intervention probes. The findings suggest that the second layer of scoring, number of grammatically well-formed utterances, was more sensitive to communicative change following intervention. However, the 'grammatically well-formed utterance' score did not measure whether a participant produced a target construction, and since ellipsis was allowed, the score did not necessarily reflect the presence of word combinations either (e.g., "*fixed*" was permitted instead of "*it's alright*"). While this task was well-suited for the purposes of the studies by Goodglass et al. (1972) and Gleason et al. (1975), the adapted version was not ideal to evaluate the availability of the target constructions used in the present intervention. Moreover, the present Story Completion Test did not probe the "*I want* [THING]" construction, but included two items on the "*I want to* [PROCESS]" construction, which is a point for future refinement.

This leads to the next research question, "Which is the most appropriate outcome measure?". Findings suggest that the less functional outcome measures (e.g., number of grammatically well-formed utterances in the Story Completion Test) might be better at capturing change than purely frequency-based, FLAT variables, which were applied to connected speech measures. Combination ratio, however, is a useful variable to characterize an individual's ability to combine single words into well-formed word combinations, and the difference between bigram and trigram t-scores in the current data set suggests that the more sensitive measure might be at the trigram level. While frequency-based variables are well-suited to evaluate a participant's ability to create well-formed word combinations, and may be useful to detect more creative combinations, further investigations should explore how pre-post difference scores correspond to small / medium / large effects.

Whitworth et al. (2018) suggest that one baseline probe might be sufficient for some connected speech measures (for example, narratives). If sampling takes place twice within a baseline period, samples seem to be more stable when taken 3 weeks apart than when they are taken within a shorter interval (1 week). In contrast to

Whitworth et al. (2018), the current case series indicates considerable intra-individual variation in connected speech across two baseline assessment points (2 weeks apart). However, the present study reported descriptive statistics to compare connected speech samples across the two baseline probes, while Whitworth et al. (2018) applied inferential statistics to investigate stability of monologic discourse samples. Future evaluations of this intervention should further examine intra- and inter-individual stability of connected speech probes (potentially comparing personal and picture-based narratives), ideally using a baseline of at least 4 weeks to ensure sampling occurs with a minimum interval of 3 weeks.

With regard to research question 3 “How acceptable is the intervention to participants with NFA and their CPs?” results suggest that the intervention, with its combination of usage-based principles, underlying learning mechanisms and inclusion of self-voice is an acceptable format for this client group. Acknowledging this is a small case series, participants with NFA and their CPs rated the intervention as helpful, with slightly higher ratings from participants with NFA. Critical feedback related to the short duration of the intervention phase and the related suggestion of including more one-to-one sessions during the intervention phase.

There is some tentative evidence that the intervention may raise awareness of communicative abilities in some participants, as reflected in more negative scores on the AIQ-21 communication section. The results also suggest positive changes in participants’ perceived well-being and emotional state, a finding in line with the suggestion that practicing FEs might have positive non-linguistic effects such as improved well-being and quality of life (Stahl & Van Lancker Sidtis, 2015). This intervention might also lead to changes in sentence comprehension abilities, as measured by TROG-2 performance. These effects could be further explored in future investigations.

In intervention studies, a common issue is whether any of the change patterns identified could be attributed to a Hawthorne or ‘charm’ effect (e.g., Riches, 2013), where participants show change due to the fact that they take part in testing or therapy sessions. In the current study, the relatively stable performance in the control measure indicates that any effects are likely to be related to the intervention. However, the results of the acceptability questionnaires need to be interpreted with participant bias in mind.

There were two participants with more severe aphasia (P4 and P3) compared to the remaining participants. Although each participant acted as their own control, the outcome patterns found for these two participants did not consistently point to positive change. Together with the findings for P1, who had relatively slight aphasia, and whose performance in connected speech measures was relatively stable across baseline- and post-intervention probes, the question of who benefits most from this intervention could tentatively be answered by examining P2's and P5's language profiles more closely. Overall, due to the mixed outcomes and the multifaceted nature of this intervention, it is not possible at this stage to specify which elements may have been most beneficial to individual participants.

Finally, Bhogal et al. (2003) recommend at least 8.8 hours over 11 weeks for speech and language therapy to be effective. The present intervention has the potential to reach a high intensity in a relatively short amount of time (6 weeks intervention) via self-managed computerised home practice. In the current study, there were eight 60-minute intervention sessions over four weeks (intervention Phase 3), plus a structural priming element (WMG, intervention Phase 2) and one to two individual recording sessions (intervention Phase 1). Participants also undertook self-managed home practice during Phase 3 of the intervention. However, the exact amount of time that participants spent on these self-managed Phase 3 activities at home could not be evaluated. Therefore, dose was estimated by using the number of days each participant spent on the WMG and these figures were correlated with outcome (combination ratio difference score), indicating that there may have been a positive relationship between dose and outcome (which would be in line with previous findings, e.g., Varley et al., 2016). This would be one tentative explanation for P1's response to intervention: this participant spent the least amount of time on the WMG and made little change in primary outcome measures and the Story Completion Test, as compared to the remaining four participants. Further investigations using a larger sample size and a refined version of the current intervention should measure self-administered therapeutic dose more accurately, to explore whether there is a positive dose-outcome correlation.

5.4.3. Suggestions for future research

Initially, 10 individuals with aphasia were contacted via email or telephone. The recruitment rate therefore was 50%. Barriers to participation included the time-consuming nature of the project, that speech and language therapy sessions were already being received, and that the participant with NFA did not have a friend or family member with whom they could record conversation videos. This information should be taken into account when designing future trials using a refined version of the current intervention.

For most outcome measures, considerable baseline variability was observed. One suggestion might be to reduce the range of talk types, for instance by focusing on picture-based narratives only, to reduce variability and allow better comparison to existing literature (Beales et al., 2018; Whitworth et al., 2018). In order to detect meaningful gains following intervention, a quantitative, automated analysis could be combined with a qualitative analysis, examining the number of constructional schemas and their variations. In the present study, such a combined approach was attempted by evaluating frequency-based characteristics of word combinations produced within narratives, and availability of constructions by analysing performance in the Story Completion Test. However, since the Story Completion Test did not seem to be an ideal tool to capture the ability to produce the target constructions, analysis of combination ratio could be combined with a qualitative analysis of the types of constructions and their variations. An example of P1's narrative production is the presence of the "[REFERENT] *is*-TENSE [PROPERTY]" construction, with examples of variations including: "*he's really funny*" and "*it's okay*". The availability of such constructions and their variations could be compared before and after intervention and would thereby add a valuable layer to an evaluation of whether this type of intervention facilitates more productive use of target constructions. Although such a qualitative analysis would require raters to categorize each utterance, which can be time-consuming, it would provide insight into the availability and use of specific constructions. Such a procedure would be similar to the techniques applied in Dąbrowska & Lieven (2005) and Lieven et al. (2009) who analysed fixed phrases and frames with slots in the language of children (Chapter 2).

The results raise the question of whether the Story Completion Test, in its present form, can capture change for trained constructions, given that one participant was unable to retrieve any constructions prior to the intervention (floor effect) and also the mixed cloze probabilities found in the normative sample. One way of refining the Story Completion Test would be to include different levels of difficulty, as in the SPPA (Helm-Estabrooks & Nicholas, 2000), where a story completion format with two levels is applied. Level A reflects repetition of a sentence included in a probe such as: "*It is very hot in the kitchen, so Mary says to Paul, 'Open the window'. What does Mary say to Paul?*" Level B follows level A if the participant repeats a phrase correctly. It uses the principle of benefit of repetition (i.e., delayed repetition). An example for level B is: "*It is very hot in the kitchen, so what does Mary tell Paul to do?*" These two levels could be applied to a refined version of the Story Completion Test. By starting with level A, the test a) would resemble the format of the intervention more closely, and b) would be more suitable for participants with more severe aphasia. Level A could also be utilized to rate the fluency with which constructions are repeated before as compared to following the intervention. However, what would be measured by employing these two levels would not require participants to generate constructions on their own, which would be the ultimate aim of an outcome measure evaluating this type of intervention.

The original Story Completion Test was successfully administered with nine individuals with moderate to severe aphasia in a study by Gleason et al. (1975). Interestingly, Gleason et al. report that the most difficult items in their original task were structures that required three elements (e.g., WH-questions such as "*where did you put my shoes*", declarative transitive structures such as "*the dog chases the cat*"). Most of the constructions in the current study required three elements (e.g., SVO structures, WH-questions) for a full 'expected answer' score, which is another factor to consider when using a similar task in future studies.

The same 12 constructional schemas were practiced with every participant in the current study. Although the activities at the end of each topic allowed for personalized items, future studies could try to further individualize parts of the intervention, to ensure they build upon a speaker's existing inventory of constructions and have relevance to their everyday activities. For instance, a more detailed analysis of existing construction frames produced by a participant during the pre-intervention phase could inform selection of relevant intervention stimuli.

While these phrases can be applied to several scenarios, it is difficult to relate them to events that can be illustrated on pictures which could add to or replace written forms of the phrases. It would be ideal, however, to map some of the constructions practiced in the intervention to everyday events. For instance, some items from the SPPA (Helm-Estabrooks & Nicholas, 2000) consist of similar constructions such as "*Where is the pen*" which are embedded in short stories, and are linked to illustrations – this could be one aim for a refined version of the present intervention. In this context, another idea for refining this intervention could be to practice the structures within dialogic settings, as is the case in ILAT. Possible formats could be to structure sessions according to themes such as talking about opinions or asking for information (similarly to the 'topics' used in the present intervention). The researcher or clinician could support production of phrases by embedding these into dialogues.

For home activities, it would be ideal if the computerised program automatically recorded participants' attempts at producing constructions. This refinement was also suggested by one participant with NFA as a comment in the post-intervention feedback questionnaire. Moreover, the WMG could be implemented as an ongoing module throughout the entire intervention phase (weeks 5-9). For instance, each home practice activity could begin with a 5- to 10-minute WMG practice, before moving on to production of phrases. In this way, structural priming could be embedded in a more systematic and consistent way in the weekly practice, and the duration of Phase 3 could be extended by one week.

A study by J. T. Kaplan et al. (2008) suggests that the RH has a major role in the processing of self-referential stimuli such as self-voice, and the RH has also been associated with the processing of FEs (Van Lancker Sidtis & Yang, 2017). It would be interesting to investigate further in what way the RH might be involved in this type of intervention, where usage frequency and self-voice are core elements. This might be achieved by carrying out an fMRI analysis with participants in future trials, to investigate whether there was a shift in RH activity after intervention.

As stated in Lancaster, Dodd, & Williamson (2004, p. 308), "a major reason for conducting a pilot study is to determine initial data for the primary outcome measure, in order to perform a sample size calculation for a larger trial”. Using the combination ratio mean difference score for the five participants of this case series (baseline average versus post-intervention probe; .03) and the standard deviation (.04) as an indicator of the combination ratio effect size, the minimum sample size for reaching a medium effect size in trial with a control group would be 23 (when using a one-tailed independent t-test), with an alpha level of .05 and power of .08. This estimation is based on G*Power (version 3.1.9, Erdfelder, Faul, Buchner, & Lang, 2009), where a hypothetical control group was assigned a mean of 0 and the same SD as in the current dataset.

Finally, although weekly samples of everyday conversations were collected during the pre- and post-intervention phase, the scope of the current report did not allow for their analysis. Future investigations could explore whether participants showed positive change in their conversational inventories within naturalistic conversations, using methods similar to Carragher et al. (2015).

5.5. Conclusion

The novel intervention for NFA was based on common constructions and their communicative functions. The intervention was acceptable to participants with NFA and their CPs, and shows promise to be effective for some individuals with NFA. Results from this case series suggest that the intervention may have the potential for transfer to participants’ connected speech. Outcome measures for future trials could focus on more general linguistic variables (e.g., grammatically well-formed multiword utterances in narratives) and examine further the intra- and inter-individual stability of connected speech.

This study should be further evaluated by taking into account everyday conversational data and a qualitative analysis of the availability of constructions before and after intervention. It could then be followed by a larger trial in which hypotheses, generated by the present case series, could be tested (Carey & Boden, 2003), for instance with regard to the relationships between aphasia severity, dose and outcome.

6. Discussion and future directions

To gain greater understanding of the importance of usage-based effects in aphasia, this thesis explored the use of familiar collocations in Broca's aphasia (study 1), examined the sensitivity to collocation strength in real-time processing by testing a large group of participants with various types of aphasia (study 2) and investigated the impact of a novel intervention aimed at training flexible use of semi-fixed constructions in NFA (study 3). All three studies used collocation strength to operationalize familiarity and to approximate formulaicity. Collocation strength was determined by using corpus-derived t-scores, an association measure calculated on the basis of the spoken BNC. In this way, more or less collocated combinations were identified in different speaker groups and used to design experimental stimuli and intervention targets.

In study 1, for the first time, everyday conversations of speakers with Broca's aphasia and their CPs were explored with regard to the degree to which residual multi-word utterances resemble strongly collocated phrases. Based on previous findings on FE usage, it was hypothesized that speakers with Broca's aphasia would rely more on familiar, strongly collocated word combinations. As predicted, word combinations in Broca's aphasia largely consisted of strongly associated bi- and trigrams, as compared to the combinations employed by neurotypical speakers. Furthermore, these frequency-based profiles were stable across a second group of 39 speakers with Broca's aphasia engaged in semi-structured interviews. These findings suggest that residual word combinations in aphasia can be explained by their frequency-based properties, a conclusion that is in accordance with Zimmerer et al. (2018). If t-scores are accepted as indicative of formulaicity, the findings also support previous literature (Van Lancker Sidtis & Postman, 2006) where, using raters to identify FEs, higher proportions of FEs were found in speakers with aphasia compared to participants with RHD. However, all of the PWA reported by Van Lancker Sidtis & Postman (2006) were individuals with fluent aphasia. Study 1 therefore extends our knowledge of FEs in NFA.

Study 1 provides novel evidence of the proposal that familiar collocations in aphasia act as residual constructions (e.g., "*I suppose*"-construction, Beeke, 2003), where linguistic form is paired with semantic-pragmatic meaning. In the literature, strongly collocated word combinations, acting as conversationally rich building blocks in neurotypical discourse, are often termed ‘automatisms’ or ‘stereotypical language’ when they occur in aphasic spoken language output. However, these terms are associated with disordered language. In this dataset, where speakers with Broca’s aphasia engaged in dyadic discourse, conversational functions of one prominent phrase, IDK, were studied by applying a rating scheme motivated by previous studies (Diani, 2004; Hildebrand-Edgar, 2016; Pichler & Hesson, 2016). Results showed that IDK fulfilled a variety of conversational functions across speakers with Broca’s aphasia and their CPs, which confirms previous findings of IDK usage (Diani, 2004; Grant, 2010). However, the functional profile of IDK was shown to be adapted to aphasic difficulties. For instance, IDK was used by CPs as a strategy to signal difficulties with understanding the meaning of a previous aphasic turn. A higher proportion of turn-constructional functions such as turn-yielding was identified in the group of speakers with aphasia, relative to the neurotypical group. Moreover, the interactional linguistic analysis provided new evidence of interpersonal and multifunctional uses of IDK in aphasia, although these instances were relatively rare. These findings confirm and extend previous reports, for example a single case report by Simmons-Mackie & Damico (1997), where IDK in a speaker with NFA was described as a tool to yield the conversational floor. Taken together, these results indicate that a larger study, deploying both corpus-based and interactional linguistic methods is likely to yield further insights into the discourse profiles of familiar collocations such as IDK.

The results from study 1 further indicate that, in more severe aphasia, IDK functions might reduce to purely turn-constructional roles. These findings tentatively suggest an important role of FEs in severe aphasia, where they may become crucial to participate in conversation. This is in line with Simmons-Mackie et al. (2004) who found that a speaker with severe aphasia was able to place FEs strategically into a turn to help the CP to interpret the intended message based on shared knowledge and the conversational context. Moreover, the present study suggests that IDK, used to take a turn, can become more meaningful by combining it with gestures that add semantic content, although there was only a small number of examples in the current dataset. Further, not every IDK token could be assigned a function. While in some cases, this was due to a lack of conversational context, it may also be indicative of over-reliance on such phrases in more severe aphasia, given the limited inventory of constructions. In this context, it is important to keep in mind that familiar collocations represent a conversational resource in most cases, but are also associated with disadvantages if overused in discourse. Wray (2012a) discusses the consequences of over-reliance on FEs in speakers with AD, a condition associated with abnormally high proportions of FEs (Van Lancker Sidtis, Choi, Alken, & Sidtis, 2015). Wray (2012a) emphasizes that formulaic language can be a barrier to conversation. Her argument is that an abnormally high amount of FEs, particularly repetitions and 'empty' language, leads to negative feelings in an interlocutor, who might feel "diminished and undervalued" (Wray, 2012a, p. 173). As a consequence, the interlocutor may use more FEs as an unconscious way of distancing and protecting themselves. This leads to what Wray (2012a, p. 173) calls a "downward spiral" which can cause a breakdown of meaningful conversational exchanges. Thus, it is important that regular CPs of people with communication difficulties (such as AD or aphasia), where FEs are a main resource, understand the positive and negative effects of FEs.

Together, the findings from study 1 suggest that strongly collocated word combinations, residual to aphasic language difficulties, may be processed as formulas. This interpretation is supported by some of the assumptions made by Ruiter et al. (2010) who argue that speakers with Broca's aphasia employ neurotypical elliptical utterances such as "*washing hands"* as they require less processing load than full sentences (Ruiter et al., 2010). Similarly, some familiar phrases in aphasia, for example IDK, which seem to be fixed in form, may not require grammatical processing. Instead, they "have become grammaticised as discourse particles by virtue of the fact that they form such regular and frequent combinations" (Beeke et al., 2007a, p. 265). The results from study 1 support this proposal, which may explain why more strongly collocated phrases are often retained in Broca's aphasia. Overall, this shows how association measures in combination with a qualitative, interactional linguistic analysis can provide new insights into grammatical behaviour in aphasia.

Relatively little is known about the recognition of familiar collocations in listeners with aphasia. Study 2 presented a combination of on-line, psycholinguistic (word monitoring) and corpus-based methods (t-scores) to measure sensitivity to collocation strength in three participant groups: younger adults, older adults and PWA. The study explored how listeners with aphasia recognized final words of stronger versus weaker collocations embedded in sentences, as compared to an age-matched group of neurotypical listeners. In particular, word monitoring latencies to the final words of stronger, relative to weaker collocated trigrams were measured. In line with predictions, all participant groups showed shorter monitoring latencies for stronger collocations, suggesting a processing speed advantage of stronger compared to weaker collocations. In the aphasic group, however, effects were smaller than in neurotypical participants. Further, the trigram pairs in study 2 were either modified by manipulating the first or the second word of a trigram. While these two experimental conditions affected the size of the effect in both control groups such that facilitatory effects were stronger when the critical word directly preceded the target, PWA, somewhat surprisingly, did not seem to show a boost in this condition. These results point to weakened anticipatory activation in the aphasic group across first- and second-word modifications. This tentatively suggests that the presence of aphasia leads to more effortful listening in real-time, such that anticipatory activation of strongly collocated words seems slower or diminished compared to neurotypical listeners. However, it should be noted that the aphasic group consisted of participants with a variety of underlying language and cognitive profiles and a relatively low proportion of participants with severe aphasia. Thus, it is necessary to design future experiments that explore factors such as aphasia type, severity, or the role of word class and distance between the critical word and the target, which may have influenced these results.

The results in both younger and older adults replicate findings from previous studies (e.g., Arnon & Snider, 2010; Jiang & Nekrasova, 2007), where a processing speed advantage for stronger collocations over less common/formulaic counterparts was found. Moreover, in the current study, sentences where there was a distance between the critical and the target words seemed to be more sensitive to education-related differences in older neurotypical participants, as well as off-line processing in PWA. This indicates that monitoring for a word that occurs with a distance to the critical word may make a WMT more sensitive to the amount of linguistic experience, and severity of language difficulties in aphasia. This distance effect could be further investigated in future research.

Phrases can be more or less strongly collocated, or, to put it differently, we can be more or less certain about the presence of a collocation. In study 2, this degree of association strength was operationalized by ranking pairwise t-score differences (i.e., between the stronger and weaker collocation of a pair). The degree of collocation strength appeared to influence the processing advantage in adults between 18 and 30 years of age. On the other hand, in adults older than 40, stronger collocations seemed to influence RTs regardless of the size of the t-score difference. While this was an unexpected result, it seems to be in line with Dąbrowska (2014b) who found a peak of collocational knowledge (as measured by an off-line test using multiple choice questions) at around 32 years of age, while participants older than 50 years showed a decline in collocational knowledge. However, an explanation for Dąbrowska's result could be that age results in more exposure, and as a consequence, people accept more potential collocations. This may lead to a weaker effect if the task is to select one collocation. Hence, the age difference in the current study may reflect more robust sensitivity to collocation strength in older adults, independent of the difference in degree of collocation strength between the trigram pair. Moreover, older adults, as compared to younger adults, showed stronger effects for binomials. This may be indicative of a different mechanism (e.g., semantic or associative priming) that explains facilitation in semantically salient word combinations such as binomials.

In study 2, average z-scores across speakers were used to perform mixed ANOVAs. These traditional ANOVAs and correlational analyses were complemented with Bayesian equivalents, using BFs, methods which are relatively new to the area of speech and language therapy research. BFs enable quantification of how much evidence there is in favour of the null and alternative hypotheses (Marsman & Wagenmakers, 2017). They are useful to complement traditional hypothesis testing methods such as ANOVAs.

A common focus of research into NFA is to develop effective intervention programmes that enhance the ability to access and produce multi-word utterances relevant to everyday life. Study 3 presented a case series of a novel computerised intervention for NFA. Underpinned by a usage-based approach, the intervention aimed to increase participants' repertoires of constructions and to enhance their connected speech. Although familiar, high-frequency phrases have high functional value, discourse situations, topics and contexts vary across every individual language user. Hence, practicing a pre-determined set of phrases is not an ideal method to improve a speaker's inventory of conversational building blocks (Linebarger et al., 2008). This problem was addressed by focusing on superimposition, a mechanism that facilitates productivity and thus communicative flexibility. The target constructions were grounded in generic, high-frequency phrases such as "*I like it*" or "*it's alright*", with high functional relevance to everyday conversation. Lexical items were inserted into open slots (e.g., "*they like flowers*", "*she's nice*"). Participants also practiced personally relevant variations of constructions.

The intervention program consisted of a combination of usage-based principles and neuroscientific and psycholinguistic learning principles (self-voice, structural priming, errorless learning and enhancing dose through home practice). The combination of these principles, new to aphasia intervention, was found to be acceptable to participants. While it remains unclear to what degree each of these principles contributed to intervention outcomes, they offer potential for further applications in aphasia therapies. For instance, self-voice is a useful motivational component that can be harnessed to model fluent production of word combinations, for example for individuals with non-fluent language output.

Outcomes of the intervention study were evaluated by measuring frequency-related variables in personal and picture-based narrative samples. In addition, an adapted Story Completion Test, normed with a large sample of neurotypical adult speakers, was used to investigate the availability of constructions targeted in the intervention. The intervention was acceptable to all participants and resulted in increased connected speech for some. More generic linguistic measures, i.e., number of grammatically well-formed utterances, were the most sensitive in capturing change after intervention, as compared to more targeted measures such as the ability to form word combinations (combination ratio) or to access constructions targeted in the intervention. This was a somewhat unexpected finding which, however, needs to be interpreted in light of the small case series, and needs to be followed up by additional analysis of conversation samples and potentially a larger trial using a refined version of the intervention. This will be necessary to explore which candidates benefit most from this type of intervention, and to refine suitable outcome measures that can reliably and sensitively capture change in connected speech output.

The present intervention suggests that usage-based theory, where constructions map onto semantic-pragmatic functions, represents a fruitful model to use in intervention for chronic aphasia. It is compatible with principles employed in previous interventions (e.g., lexical insertion used by Carragher et al., 2015, or training phrases with everyday relevance as used in script training for aphasia, e.g., Kaye & Cherney, 2016), but embeds these techniques in a theoretical model that is able to account for aphasic language production and comprehension at the level of bi- and trigrams. Furthermore, one interesting finding from the case series was that most participants self-rated their well-being and emotional state more positively after as compared to before the intervention. This result provides support for Stahl & Van Lancker Sidtis' (2015, p. 3) hypothesis, that using constructions such as FEs in aphasia intervention may have “a possible beneficial influence on motivation, subjective well-being, and quality of life”.

In summary, the empirical findings demonstrate that a frequency- or usage-based perspective on language in aphasia is useful to profile aphasic language output at the level of multi-word utterances and to research grammatical behaviour and collocational processing in aphasia. Usage-based principles are suitable to inform aphasia interventions. In the following paragraphs, theoretical and clinical implications will be further discussed, and suggestions for future research will be made.

6.1. Theoretical implications

Wray (2012b, p. 245) explains that “formulaic language has become very productively associated with the new generation of grammatical theories that locate multiword strings at the center, rather than the periphery, of our linguistic experience”. Referring to Ellis (2008b) and Taylor (2002), Wray’s statement is well-suited to highlight a way to theorise about grammar that can account for FEs. There is an overlap between the presence of fixed phrases or collocations in the speech produced by speakers with aphasia (study 1), and experimental evidence which shows that usage-based information at the multiword level plays a role in aphasic language processing (study 2). The findings from this thesis suggest that frequency at the multi-word level, one aspect of language usage, is not only important in neurotypical language processing, but it is one driver of language production and recognition in aphasia. This confirms and expands previous findings on the importance of usage-based factors to aphasic language processing such as single word frequency and AoA (e.g., Brysbaert & Ellis, 2016; Kittredge et al., 2008; Nickels & Howard, 1995), frequency effects in sentences (DeDe, 2013; Gahl et al., 2003; Knilans & DeDe, 2015), and the fact that speakers with aphasia often rely on FEs when they converse (e.g., Van Lancker Sidtis, 2012).

Moreover, this thesis contributes to exploring the syntax-lexicon continuum in aphasia, at the level of bi- and trigrams. The syntax-lexicon continuum is a useful conceptualization to explain why individuals with Broca's aphasia are able to access single words as well as familiar well-formed word combinations when they produce language, and why familiar word combinations such as IDK are used in interactionally beneficial ways (as form-meaning pairings). More strongly associated word combinations, as opposed to less associated combinations, are processed more easily by individuals with aphasia because they occur more frequently in language usage, and thus are thought to be fused more strongly. Since association measures can be used to approximate the level of entrenchment (Stefanowitsch & Flach, 2016), these findings suggest that combinations with high t-scores, i.e., those that are likely more entrenched, are more readily available. This may lead to differential, easier processing of collocations compared to more creative, novel combinations. As indicated by study 2, the first words of strongly collocated word combinations may activate subsequent words by spreading activation due to their relatively high association strength (e.g., Collins & Loftus, 1975; Vilkaite & Schmitt, 2017). This was especially true for binomials, where a semantically rich word (noun) seemed to prime the final word of the trigram (noun). Results from study 2 suggest that while association strength contributes to a processing advantage in aphasia, additional semantic relatedness and saliency of collocations may increase this processing advantage. The contribution of collocation strength on its own, versus collocation- and semantic association needs to be investigated in further experiments with PWA.

Overall, the building blocks of language output in NFA seem to be represented as lexically concrete forms such as words and fixed units (e.g., "*it's alright*"). At the same time, many of these multiword units have the potential to map onto more abstract patterns (e.g., "X *is alright*" / "*it's* X"), which would allow for communicative flexibility. Study 3 attempted to bridge a gap between familiar collocations as fixed expressions versus familiar collocations as exemplars of more abstract schematic frames with open slots. However, these relationships have been postulated based on theoretical assumptions, and more work needs to be done with regard to applying suitable methods to verify how constructions are mental representations of individuals with aphasia. We know that "*it's*" represents a strongly collocated combination that is frequently used by individuals with aphasia, but a purely frequency-based approach does not answer the question of which construction this combination is a part of (e.g., whether the "*it's alright*" construction maps onto a "X *is alright*" or "*it's* X" schema), and how the type and severity of aphasia may influence representations and processing of constructions. One hypothesis arising from the present thesis is that the idiom principle (i.e., retrieving stored units such as formulas) may operate at the uni- ("*look*"), bi- ("*this one*") and trigram level ("*don't know*", "*I want* X") in aphasia, with the potential to teach an open choice pathway via superimposition, utilizing unigrams or chunks at the bi- and trigram level (e.g., "*I want* [*this one*]").

6.2. Clinical implications

The frequency-based method used in this thesis is an objective and effective way to study FE usage and has the potential to be developed for application to clinical research. Past methods, using rater reliance on meta-knowledge of their language (e.g., Van Lancker Sidtis & Postman, 2006), classified a given utterance either as a formulaic or a nonformulaic utterance. In contrast, the frequency-based approach allows for estimation of the degree of familiarity or collocation strength. This is an advantage over having raters decide about formulaic status, given that classification of FEs may not be binary.

As shown by Dąbrowska (2014b), not every corpus-derived collocation may be represented in a speaker's mind. Thus, a purely frequency-based approach has its limitations. While it is one tool to extract and identify strongly associated chunks, it does not tell us about how such chunks might be used by an individual engaged in dyadic discourse, or how the meaning of such word combinations may be adjusted to aphasic difficulties or the conversational context. In study 1, analysis of the conversational functions of IDK complemented the frequency-based analysis in that it shed light on the 'full picture' of use of this collocation. This mixed-methods approach can inform development of assessments that take into account naturalistic language use. For instance, narrative samples could first be analysed from a frequency-based perspective, followed by extracting specific familiar collocations to examine the conversational functions.

The main advantage of using the word monitoring paradigm is that it allows examination of real-time processing of spoken language. Thus, it is a useful method to study the processing of connected speech, reflecting the speed with which typical spoken discourse happens. Similarly, a recent study by Conroy, Sotiropoulou Drosopoulou, Humphreys, Halai, & Lambon Ralph (2018) tested whether participants with different types of aphasia (including Broca's, anomic and transcrortical motor aphasia) showed improved naming abilities following a combined speed- and accuracy-focused treatment, as opposed to a more traditional treatment focusing on accuracy only. The primary aim was to improve participants' real-time lexical retrieval (i.e., decreased speed, improved accuracy and greater use of target words in connected speech samples). Findings showed that the novel speed- and accuracy training was more effective in increasing performance in naming pictures compared to a traditional approach, and led to enhanced vocabulary within connected speech production. The word monitoring paradigm represents a suitable on-line processing measure that may be useful as an instrument evaluating outcomes of such novel speed- and accuracy treatments with regard to real-time recognition of words occurring in sentences.

The intervention tested in study 3 aimed to increase a speaker's repertoire of constructions, by practicing familiar phrases using superimposition. The approach of inserting lexical units into a sentential frame is not new to aphasia therapy. Carragher et al. (2015), for example, used lexical insertion, motivated by mapping principles (where words were mapped onto roles in a sentence such as agent or theme). Thus, practicing flexible use of functionally relevant constructions is similar to this approach, but here the focus was on semi-lexicalised frames with pragmatic functions in conversation, rather than on abstract thematic roles. A focus on the mapping between linguistic form (a specific trigram such as "*I like* X") and semantic-pragmatic meaning (expressing an opinion) is similar to and may be compatible with script training in aphasia, but is embedded in a theoretical framework that allows for incorporation of principles such as lexical insertion or juxtaposition. Thus, it helps to explain outcomes within a plausible theoretical linguistic framework. Future developments of this novel intervention could consider implementing dialogic settings, to emphasize real-life applications of the linguistic material even more.

Finally, the intervention was designed to be computerised, which offers a useful format to implement principles such as increased dose, errorless learning or structural priming training, all of which are considered effective learning strategies in aphasia (e.g., Lee & Man, 2017; Varley, 2011). For the first time, this thesis evaluated an intervention using a combination of such learning strategies and self-voice, and outcomes of the case series indicate that this intervention has the potential to motivate future development of usage-based interventions for aphasia.

6.3.Future research

While the current thesis informed our knowledge of the inventory and processing of constructions in aphasia, it stimulated further research questions. These include:

- Is the t-score the most suitable measure to operationalise degree of familiarity / collocation strength of word combinations?
- Does AoA of word combinations influence language processing in aphasia?
- What is the most comprehensive way to investigate use of constructions?
- What is the role of familiar collocations in people with different degrees of aphasia severity and types of aphasia?

The techniques used in the present thesis, where frequency characteristics of word combinations were derived from the BNC, have proven to be useful to analyse speech samples of individuals with aphasia engaged in discourse, and to characterise different speaker groups. Moreover, t-scores are a rich resource to derive experimental stimuli. However, frequency characteristics of word combinations can be quantified using a number of measures, and t-scores are only one option of association measures. As stated in Chapter 2, another variable commonly reported in corpus-based research is MI. While "the MI-score tends to highlight infrequent combinations" (Gablasova et al., 2017, p. 172), i.e., is well-suited for words with a very low frequency, t-scores are said to be more robust for high co-occurrence frequencies (Gries, 2010). However, it should be noted that neither of these two take into account directionality which becomes important to explore how much one word might prime the next word (Durrant & Doherty, 2010). Such processes may be relevant in study 2 in relating facilitatory effects to anticipatory activation. In such cases, probability-based measures such as transitional probability (i.e., the likelihood that word b follows word a), could be useful to consider. Another disadvantage of the t-score is that it cannot be used to compare collocations across different corpora (Hunston, 2002). Instead, it is recommended to rank t-scores within each individual study. This needs to be kept in mind when selecting frequency-based measures in future usage-based experiments. There are other measures of association strength such as Log Dice, recommended by Gablasova et al. (2017). Log Dice addresses the above-mentioned weaknesses of t- and MI-scores. It operates on a fixed scale with a maximum of 14 which allows the comparison of values across corpora (a weakness of the t-score), and it is not biased towards rewarding lower-frequency combinations (a disadvantage of MI).

AoA is another usage-based variable which has been shown to be linked to single-word frequency. Likewise, AoA of word combinations might intercorrelate with phrase frequency and collocation strength. Future work could study the influence of AoA of word combinations on aphasic language processing. Arnon et al. (2017) provided a first report of how to operationalize and study the effect of AoA on multiword processing in neurotypical adults. These methods could form a starting point for incorporating multiword AoA in experimental aphasia research and for studying its possible influence on the processing of multiword sequences.

Use of frequency-related measures could be combined with other existing usage-based methods such as traceback analysis (Lieven et al., 2009). This could inform exploration of a speaker's individual language experience and use of fixed phrases / frames with slots, for example in the context of evaluating intervention outcome. However, in order to use traceback, a given corpus needs to be of a certain size to enable division into a main and a test corpus. Thus, relatively large corpora are required (as was the case in Dąbrowska & Lieven, 2005, for example, where the corpus consisted of approximately 30 hours per child at each time point). One can imagine that in order to study non-fluent language output in aphasia, it is likely that corpora would have to be even bigger due to the limited amount of word combinations produced by speakers with NFA as compared to typically developing children. On the other hand, such methods could be appropriate for examining language produced by individuals with fluent aphasia. One existing collection that offers a relatively large sample of aphasic speech is AphasiaBank (MacWhinney et al., 2011), a data source that was successfully incorporated into the current thesis (study 1). Moreover, there is a trend to collect naturalistic data such as everyday conversations over long periods of time, for example to evaluate conversation-based intervention (e.g., Best et al., 2016; Carragher et al., 2015). As demonstrated in study 1, such data are well-suited for usage-based analysis, to study which constructions are available to a speaker.

Although the present thesis was largely focused on NFA (study 1 and 3), there are reports suggesting that speakers with fluent aphasia make use of a high proportion of FEs (Van Lancker Sidtis & Postman, 2006), many of which may represent semi-fixed frames with slots (Wray, 2002a). However, these speakers might have greater difficulty to fill open slots with lexical units due to their word finding difficulties. Thus, one area for future research would be to explore to what degree familiar collocations might be a conversational tool in fluent aphasia. Furthermore, future research could include measures of listener perceptions on narratives (Carragher, Talbot, Devane, Rose, & Marshall, 2018) to investigate how use of familiar collocations such as IDK or "*I went to* ___" are perceived. Such methods could expand our understanding of when these constructions are perceived as 'empty' speech, as helpful to keep conversations going, or as a tool for a speaker with (more severe) aphasia to contribute to the conversation (e.g., IDK + gesture).

Finally, just as collocations represent appropriate targets for second language learning (Durrant & Doherty, 2010), individuals with aphasia may benefit from being encouraged via therapy to use their collocational language as a resource. One suggestion from this thesis is to use those collocations that are available to an individual as intervention targets, and to combine these with techniques to enhance language productivity. In this context, one main suggestion for further developing the findings from this thesis is to evaluate study 3 outcomes from a quantitative (frequency-based) and qualitative perspective (i.e., which constructions were available before versus after the intervention, did the productivity of these change following intervention), along with the analysis of everyday conversational data.

6.4. Conclusion

Usage-based theory has recently attracted more and more attention in aphasiology. Previous studies show how this framework can account for the presence of familiar collocations, which are often observed in the speech of individuals with aphasia. Study 1 showed that speakers with Broca's aphasia rely on strongly collocated word combinations when engaged in everyday conversations. Moreover, the common phrase IDK, a ubiquitous feature of both neurotypical and aphasic speech, represents a construction with multiple pragmatic functions, adapted to aphasic difficulties. The results of a word monitoring experiment in study 2 suggest that collocation strength in language use affects ease of recognizing final words of combinations in listeners with aphasia. However, the effects were weaker compared to neurotypical controls. Finally, a novel intervention for NFA, underpinned by usage-based principles, was developed and piloted within a case series (study 3). Results show promising trends with regard to enhanced connected speech within narratives. Together, this thesis provides further support that usage-based accounts are well-suited for examining language processing in aphasia at the multi-word level.

7. References

Abbot-Smith, K., & Behrens, H. (2006). How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. *Cognitive Science*, *30*, 995–1026. https://doi.org/10.1207/s15516709cog0000_61

Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, *88*(1), 148–159. https://doi.org/10.1016/S0093-934X(03)00296-7

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273. https://doi.org/10.1017/S030500091400049X

Anderson, E. C. (2017). *Frequency effects in the processing of verbs and argument structure: Evidence from adults with and without acquired aphasia. Doctoral Thesis.* University of Sheffield.

Arnon, I., & Cohen Priva, U. (2013). More than Words: The Effect of Multi-word Frequency and Constituency on Phonetic Duration. *Language and Speech*, *56*(3), 349–371. https://doi.org/10.1177/0023830913484891

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220. https://doi.org/10.1080/02687038.2016.1147767

Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning: The Effect of Familiarity on Children’s Repetition of Four-Word Combinations.

Psychological Science, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Baum, S. R. (1989). On-line sensitivity to local and long-distance syntactic dependencies in Broca's aphasia. *Brain and Language*, *37*(2), 327–338. https://doi.org/10.1016/0093-934X(89)90021-7

Beales, A., Whitworth, A., Cartwright, J., Panegyres, P. K., & Kane, R. T. (2018). Determining stability in connected speech in primary progressive aphasia and Alzheimer's disease. *International Journal of Speech-Language Pathology*, *20*(3), 361–370. https://doi.org/10.1080/17549507.2018.1442498

Beeke, S. (2003). "I suppose" as a resource for the construction of turns at talk in agrammatic aphasia. *Clinical Linguistics & Phonetics*, *17*(4–5), 291–298. https://doi.org/10.1080/0269920031000080055

Beeke, S., Maxim, J., Best, W., & Cooper, F. (2011). Redesigning therapy for agrammatism: Initial findings from the ongoing evaluation of a conversation-based intervention study. *Journal of Neurolinguistics*, *24*(2), 222–236. https://doi.org/10.1016/j.jneuroling.2010.03.002

Beeke, S., Wilkinson, R., & Maxim, J. (2003). Exploring aphasic grammar 2: do language testing and conversation tell a similar story? *Clinical Linguistics & Phonetics*, *17*(2), 109–134. https://doi.org/10.1080/0269920031000061786

Beeke, S., Wilkinson, R., & Maxim, J. (2007a). Grammar without sentence structure: A conversation analytic investigation of agrammatism. *Aphasiology*, *21*(3–4), 256–282. https://doi.org/10.1080/02687030600911344

Beeke, S., Wilkinson, R., & Maxim, J. (2007b). Individual variation in agrammatism: a single case study of the influence of interaction. *International Journal of Language & Communication Disorders*, *42*(6), 629–647. https://doi.org/10.1080/13682820601160087

Best, W., Maxim, J., Heilemann, C., Beckley, F., Johnson, F., Edwards, S. I., … Beeke, S. (2016). Conversation Therapy with People with Aphasia and Conversation Partners using Video Feedback: A Group and Case Series Investigation of Changes in Interaction. *Frontiers in Human Neuroscience*, *10*(562). https://doi.org/10.3389/fnhum.2016.00562

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of aphasia therapy, impact on recovery. *Stroke*, *34*(4), 987–992. https://doi.org/10.1161/01.STR.0000062343.64383.D0

Bilda, K. (2011). Video-based conversational script training for aphasia: A therapy study. *Aphasiology*, *25*(2), 191–201. https://doi.org/10.1080/02687031003798254

Bishop, D. V. M. (2003). Test for Reception of Grammar, Version 2 (TROG-2). London: Pearson.

Blanken, G. (1991). The functional basis of speech automatisms (recurring utterances). *Aphasiology*, *5*(2), 103–127. https://doi.org/10.1080/02687039108249477

Blanken, G., & Marini, V. (1997). Where do lexical speech automatisms come from? *Journal of Neurolinguistics*, *10*(1), 19–31. https://doi.org/10.1016/S0911-6044(97)00001-8

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*(3), 665–670. https://doi.org/10.3758/BRM.42.3.665

Blumenthal-Dramé, A. (2016). Entrenchment From a Psycholinguistic and Neurolinguistic Perspective. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* (pp. 129–152). Washington, DC: de Gruyter. https://doi.org/10.1515/9783110341423-007

BNC. (2007). BNC, version 3 (BNC XML Edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from http://www.natcorp.ox.ac.uk/

Boas, H. C. (2003). *A constructional approach to resultatives*. Stanford, CA: CSLI Publications.

Boas, H. C. (2010). The syntax–lexicon continuum in Construction Grammar: A case study of English communication verbs. *Belgian Journal of Linguistics*, *24*, 54–82. https://doi.org/10.1075/bjl.24.03boa

Brekelmans, G., & Meitanis, V. (2016). Phonetics for Word Monitoring manual. *Unpublished manuscript.*

Bridges, K. A., & Van Lancker Sidtis, D. (2013). Formulaic Language in Alzheimer's disease. *Aphasiology*, *27*(7), 799–810. https://doi.org/10.1080/02687038.2012.757760

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: are early-learned words more resilient? *Aphasiology*, *30*(11), 1240–1263. https://doi.org/10.1080/02687038.2015.1106439

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711–733. Retrieved from http://www.jstor.org/stable/4490266

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Cacciari, C., Reati, F., Colombo, M. R., Padovani, R., Rizzo, S., & Papagno, C. (2006). The comprehension of ambiguous idioms in aphasic patients. *Neuropsychologia*, *44*(8), 1305–1314. https://doi.org/10.1016/j.neuropsychologia.2006.01.012

Carey, T. S., & Boden, S. D. (2003). A Critical Guide to Case Series Reports. *Spine*, *28*(15), 1631–1634. https://doi.org/10.1097/01.BRS.0000083174.84050.E5

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), 846–887. https://doi.org/10.1080/09602011.2013.832335

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), 322–336. https://doi.org/10.1111/1460-6984.12135

Carragher, M., Talbot, R., Devane, N., Rose, M., & Marshall, J. (2018). Delivering storytelling intervention in the virtual world of EVA Park. *Aphasiology*, *32*(S1), 37–39. https://doi.org/10.1080/02687038.2018.1484880

Cherney, L. R., Halper, A. S., Holland, A. L., & Cole, R. (2008). Computerized Script Training for Aphasia: Preliminary Results. *American Journal of Speech-Language Pathology*, *17*(1), 19–34. https://doi.org/10.1044/1058-0360(2008/003)

Church, K. W., Gale, W., Hanks, P., & Hindle, D. (1991). Using Statistics in Lexical Analysis. In U. Zernik (Ed.), *Lexical acquisition: exploiting on-line resources to build a lexicon* (pp. 115–164). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Code, C. (1982). Neurolinguistic analysis of recurrent utterance in aphasia. *Cortex*, *18*, 141–152.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Conklin, K., & Schmitt, N. (2012). The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, *32*, 45–61. https://doi.org/10.1017/S0267190512000074

Conrad, S., & Biber, D. (2004). The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. *Lexicographica*, *20*, 56–71.

Conroy, P., & Scowcroft, J. (2012). Decreasing cues for a dynamic list of noun and verb naming targets: A case-series aphasia therapy study. *Neuropsychological Rehabilitation*, *22*(2), 295–318. https://doi.org/10.1080/09602011.2011.641434

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*, 1815–1827. https://doi.org/10.1093/brain/awy087

Couper-Kuhlen, E., & Selting, M. (2001). Introducing Interactional Linguistics. In M. Selting & E. Couper-Kuhlen (Eds.), *Studies in Interactional Linguistics* (pp. 1–22). John Benjamins.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253. https://doi.org/10.1075/lab.2.3.01dab

Dąbrowska, E. (2014a). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics*, *25*(4), 617–653. https://doi.org/10.1515/cog-2014-0057

Dąbrowska, E. (2014b). Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon*, *9*(3), 401–418. Retrieved from http://dx.doi.org/10.1075/ml.9.3.02dab

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, *16*(3), 437–474. https://doi.org/10.1515/cogl.2005.16.3.437

DeDe, G. (2013). Verb transitivity bias affects on-line sentence reading in people with aphasia. *Aphasiology*, *27*(3), 326–343. https://doi.org/10.1080/02687038.2013.843151

Diani, G. (2004). The discourse functions of I don't know in English conversation. In K. Aijmer & A.-B. Strenstrom (Eds.), *Discourse Patterns in Spoken and Written Corpora* (pp. 157–171). John Benjamins.

Dienes, Z. (2011). Bayesian Versus Orthodox Statistics: Which Side Are You On? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920

Diessel, H. (2015). Usage-based construction grammar. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 295–321). Berlin: Mouton de Gruyter.

Drew, P., & Holt, E. (1998). Figures of speech: Figurative expressions and the management of topic transition in conversation. *Language in Society, 27*(4), 495–522. https://doi.org/10.1017/S0047404598004035

Druks, J., & Masterson, J. (2000). An object and action naming battery. Hove: Psychology Press.

Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*(2), 125–155. https://doi.org/10.1515/CLLT.2010.006

Edmonds, A. (2014). CONVENTIONAL EXPRESSIONS: Investigating Pragmatics and Processing. *Studies in Second Language Acquisition*, *36*(1), 69–99. https://doi.org/10.1017/S0272263113000557

Edwards, S., & Tucker, K. (2006). Verb retrieval in fluent aphasia: A clinical study. *Aphasiology*, *20*(7), 644–675. https://doi.org/10.1080/02687030600631827

Ellis, N. C. (1996). Sequencing in SLA: Phonological Memory, Chunking and Points of Order. *Studies in Second Language Acquisition*, *18*(1), 91–126. https://doi.org/10.1017/S0272263100014698

Ellis, N. C. (2008a). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 63–103). Oxford: Blackwell.

Ellis, N. C. (2008b). Phraseology: The periphery and the heart of language. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 1–13). Amsterdam, the Netherlands: John Benjamins.

Ellis, N. C. (2015). Implicit AND explicit learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3–23). Amsterdam: John Benjamins.

Ellis, N. C. (2016). Online processing of Verb–Argument Constructions: Lexical decision and meaningfulness. *Language and Cognition*, *8*(3), 391–420. https://doi.org/10.1017/langcog.2016.18

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics*, *25*(1), 55–98. https://doi.org/10.1515/cog-2013-0031

Erdfelder, E., Faul, F., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text - Interdisciplinary Journal for the Study of Discourse*, *20*(1), 29–62. https://doi.org/10.1515/text.1.2000.20.1.29

Field, A. (2009). *Discovering statistics using SPSS* (Third edit). SAGE Publications.

Fillmore, C. J. (1988). The Mechanisms of "Construction Grammar." *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, 35–55.

Fletcher, M., & Birt, D. (1983). Storylines: Picture sequences for language practice. London: Longman.

Forrin, N. D., & MacLeod, C. M. (2017). This time it's personal: the memory benefit of hearing oneself. *Memory*, *26*(4), 574–579. https://doi.org/10.1080/09658211.2017.1383434

Forsberg, F. (2010). Using conventional sequences in L2 French. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(1), 25–51. https://doi.org/10.1515/iral.2010.002

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116–124.

Franklin, S., Turner, J., & Ellis, A. (1992). The A.D.A. Auditory Comprehension Battery. York, UK: University of York.

Friederici, A. D. (1983). Aphasics' perception of words in sentential context: Some real-time processing evidence. *Neuropsychologia*, *21*(4), 351–358. https://doi.org/10.1016/0028-3932(83)90021-0

Friederici, A. D. (1985). Levels of processing and vocabulary types: Evidence from on-line comprehension in normals and agrammatics. *Cognition*, *19*(2), 133–166. https://doi.org/10.1016/0010-0277(85)90016-2

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, *67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Gahl, S., & Menn, L. (2016). Usage-based approaches to aphasia. *Aphasiology*, *30*(11), 1361–1377. https://doi.org/10.1080/02687038.2016.1140120

Gahl, S., Menn, L., Ramsberger, G., Jurafsky, D. S., Elder, E., Rewega, M., & Holland Audrey, L. (2003). Syntactic frame and verb bias in aphasia: Plausibility judgments of undergoer-subject sentences. *Brain and Cognition*, *53*(2), 223–228. https://doi.org/10.1016/S0278-2626(03)00114-3

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*(5), 576–593. https://doi.org/10.1016/0749-596X(89)90014-4

Gleason, J. B., Goodglass, H., Green, E., Ackerman, N., & Hyde, M. R. (1975). The Retrieval of Syntax in Broca's Aphasia. *Brain and Language*, *2*, 451–471. https://doi.org/10.1016/S0093-934X(75)80083-6

Goldberg, A. E. (1995). *A Construction Grammar Approach to Argument Structure*. Chicago, IL: The University of Chicago Press.

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9

Goldberg, S., Haley, K. L., & Jacks, A. (2012). Script Training and Generalization for People With Aphasia. *American Journal of Speech-Language Pathology*, *21*, 222–238. https://doi.org/10.1044/1058-0360(2012/11-0056)

Goodglass, H., Gleason, J. B., Ackerman Bernholtz, N., & Hyde, M. R. (1972). Some Linguistic Structures in the Speech of a Broca's Aphasic. *Cortex*, *8*(2), 191–212. https://doi.org/10.1016/S0010-9452(72)80018-2

Goodglass, H., Kaplan, E., & Barresi, B. (2001a). Interpretative Summary: The Major Aphasic Syndromes. In H. Goodglass, E. Kaplan, & B. Barresi (Eds.), *The Assessment of Aphasia and Related Disorders* (3rd ed, pp. 57–83). Philadelphia: Lippincott Williams & Wilkins.

Goodglass, H., Kaplan, E., & Barresi, B. (2001b). *The assessment of aphasia and related disorders* (3rd ed). Philadelphia: Lippincott Williams & Wilkins.

Goss-Sampson, M. A. (2018). Statistical analysis in JASP: A students guide. Retrieved from https://jasp-stats.org/jasp-materials/

Grande, M., Hussmann, K., Bay, E., Christoph, S., Piefke, M., Willmes, K., & Huber, W. (2008). Basic parameters of spontaneous speech as a sensitive method for measuring change during the course of aphasia. *International Journal of Language & Communication Disorders*, *43*(4), 408–426. https://doi.org/10.1080/13682820701685991

Grant, L. E. (2010). A corpus comparison of the use of I don't know by British and New Zealand speakers. *Journal of Pragmatics*, *42*(8), 2282–2296. https://doi.org/10.1016/j.pragma.2010.01.004

Gries, S. T. (2010). Useful statistics for corpus linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: selected approaches* (pp. 269–291). Frankfurt am Main: Peter Lang.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on "alternations." *International Journal of Corpus Linguistics*, *9*(1), 97–129. https://doi.org/10.1075/ijcl.9.1.06gri

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283.

Hallowell, B. (2017). *Aphasia and other acquired neurogenic language disorders: A guide for clinical excellence*. San Diego, CA: Plural Publishing.

Harmon, T. G., Jacks, A., Haley, K. L., & Faldowski, R. A. (2016). Listener perceptions of simulated fluent speech in nonfluent aphasia. *Aphasiology*, *30*(8), 922–942. https://doi.org/10.1080/02687038.2015.1077925

Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic Facilitation in Agrammatic Sentence Production. *Brain and Language*, *62*(2), 221–254. https://doi.org/10.1006/brln.1997.1905

Hatchard, R. (2015). *A construction-based approach to spoken language in aphasia. Doctoral Thesis.* University of Sheffield. Retrieved from http://etheses.whiterose.ac.uk/10385/

Helm-Estabrooks, N. (1981). Helm Elicited Language Program for Syntax Stimulation. Austin, TX: Pro-Ed.

Helm-Estabrooks, N., & Albert, M. L. (2004). *Manual of aphasia and aphasia therapy* (2nd ed.). Austin, TX: Pro-Ed.

Helm-Estabrooks, N., & Nicholas, M. (2000). Sentence Production Program for Aphasia. Austin, TX: Pro-Ed.

Helm-Estabrooks, N., Nicholas, M., & Morgan, A. (1989). *Melodic Intonation Therapy. Manual.* Austin, TX: Pro-Ed.

Helm, N. A., & Barresi, B. (1980). Voluntary control of involuntary utterances. In *Clinical Aphasiology Conference* (pp. 308–315).

Hesson, A. M., & Pichler, H. (2016). Interpreting “I don’t know” use by persons living with dementia in Mini-Mental State Examinations. *Patient Education and Counseling*, *99*(9), 1534–1541. https://doi.org/10.1016/j.pec.2016.07.001

Hildebrand-Edgar, N. (2016). *Disentangling frequency effects and grammaticalization* (Working Papers of the Linguistics Circle of the University of Victoria No. 26,1). Retrieved from https://journals.uvic.ca/index.php/WPLC/article/view/15301

Hilpert, M., & Diessel, H. (2016). Entrenchment in Construction Grammar. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* (pp. 57–74). Berlin, Germany: de Gruyter Mouton.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). Der Aachener Aphasie Test (AAT). Göttingen: Hogrefe.

Huck, A., Thompson, R. L., Cruice, M., & Marshall, J. (2017). Effects of word frequency and contextual predictability on sentence reading in aphasia: an eye movement analysis. *Aphasiology*, *31*(11), 1307–1332. https://doi.org/10.1080/02687038.2017.1278741

Huckvale, M. (2016). ProRec. Retrieved from https://www.phon.ucl.ac.uk/resource/prorec/

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Israel, M., Johnson, C., & Brooks, P. J. (2001). From states to events: The acquisition of English passive participles. *Cognitive Linguistics*, *11*(1–2), 103–130. https://doi.org/10.1515/cogl.2001.005

Janssen, N., & Barber, H. A. (2012). Phrase Frequency Effects in Language Production. *PloS One*, *7*(3), 1–11. https://doi.org/10.1371/journal.pone.0033202

JASP-Team. (2017). JASP [Computer software]. Retrieved from https://jasp-stats.org/

Jiang, N., & Nekrasova, T. M. (2007). The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal*, *91*(3), 433–445. https://doi.org/10.1111/j.1540-4781.2007.00589.x

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

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, *3*(3), 218–223. https://doi.org/10.1093/scan/nsn014

Kaschak, M. P., Kutta, T. J., & Coyle, J. M. (2014). Long and Short Term Cumulative Structural Priming Effects. *Language, Cognition and Neuroscience*, *29*(6), 728–743. https://doi.org/10.1080/01690965.2011.641387

Kay, J., Coltheart, M., & Lesser, R. (1992). Psycholinguistic Assessments of Language Processing in Aphasia. Hove, UK: Lawrence Erlbaum Associates.

Kaye, R. C., & Cherney, L. R. (2016). Script Templates: A Practical Approach to Script Training in Aphasia. *Topics in Language Disorders*, *36*(2), 136–153. https://doi.org/10.1097/TLD.0000000000000086

Kecskés, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics*, *32*(5), 605–625. https://doi.org/10.1016/S0378-2166(99)00063-6

Kertesz, A. (2007). Western Aphasia Battery-Revised. San Antonio: PsychCorp.

Kilborn, K., & Moss, H. (1996). Word Monitoring. *Language and Cognitive Processes*, *11*(6), 689–694. https://doi.org/10.1080/016909696387105

Kirk, R. E. (1996). Practical Significance: A Concept Whose Time Has Come. *Educational and Psychological Measurement*, *56*(5), 746–759. https://doi.org/10.1177/0013164496056005002

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492. https://doi.org/10.1080/02643290701674851

Knilans, J., & DeDe, G. (2015). Online sentence reading in people with aphasia: evidence from eye-tracking. *American Journal of Speech-Language Pathology*, *24*(4), S961–S973. https://doi.org/10.1044/2015_AJSLP-14-0140

Kolk, H. H. J. (1995). A Time-Based Approach to Agrammatic Production. *Brain and Language*, *50*(3), 282–303. https://doi.org/10.1006/brln.1995.1049

Kopenhaver Haidet, K., Tate, J., Divirgilio-Thomas, D., Kolanowski, A., & Happ, M. B. (2009). Methods to improve reliability of video-recorded behavioral data. *Res Nurs Health*, *32*(4), 465–474. https://doi.org/10.1002/nur.20334

Kuiper, K. (2000). On the Linguistic Properties of Formulaic Speech. *Oral Tradition*, *15*(2), 279–305. Retrieved from http://journal.oraltradition.org/issues/15ii/kuiper

Laganaro, M. (2005). Syllable frequency effect in speech production: evidence from aphasia. *Journal of Neurolinguistics*, *18*(3), 221–235. https://doi.org/10.1016/j.jneuroling.2004.12.001

Laganaro, M., & Alario, F. X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, *55*(2), 178–196. https://doi.org/10.1016/j.jml.2006.05.001

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago; London: University of Chicago Press.

Lancaster, G., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: recommendations for good practice. *Journal of Evaluation in Clinical Practice*, *10*(2), 307–312. https://doi.org/10.1111/j..2002.384.doc.x

Langacker, R. W. (1987). *Foundations of cognitive grammar. Volume 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. W. (1991). *Foundations of cognitive grammar. Volume II. Descriptive Application*. Stanford, CA: Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar: A basic introduction*. New York: Oxford University Press.

Lee, J., & Man, G. (2017). Language recovery in aphasia following implicit structural priming training: a case study. *Aphasiology*, *31*(12), 1441–1458. https://doi.org/10.1080/02687038.2017.1306638

Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, *20*(3), 481–507. https://doi.org/10.1515/COGL.2009.022

Linebarger, M. C. (2015). Speech technology for aphasic sentence production disorders. In D. Dahl, K. Beals, M. Linebarger, & R. Fink (Eds.), *Speech and Language Technology for Language Disorders* (pp. 155–194). Berlin; Boston: de Gruyter.

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. *Journal of Rehabilitation Research and Development*, *45*(9), 1401–1414. https://doi.org/10.1682/JRRD.2007.10.0171

Lum, C., & Ellis, A. W. (1999). Why do some aphasics show an advantage on some tests of nonpropositional (automatic) speech? *Brain and Language*, *70*(1), 95–118. https://doi.org/10.1006/brln.1999.2147

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 139–162). Amsterdam: John Benjamins.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

Marslen-Wilson, W., & Tyler, L. K. (1981). Central Processes in Speech Understanding. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *295*, 317–332.

Marsman, M., & Wagenmakers, E. J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, *14*(5), 545–555. https://doi.org/10.1080/17405629.2016.1259614

Matsuno, K. (2017). Processing collocations: Do native speakers and second language learners simultaneously access prefabricated patterns and each single word? *Journal of the European Second Language Association*, *1*(1), 61–72. https://doi.org/10.22599/jesla.17

McElduff, K. M., & Drummond, S. S. (1991). Communicative functions of automatic speech in non-fluent dysphasia. *Aphasiology*, *5*(3), 265–278. https://doi.org/10.1080/02687039108248528

Meinzer, M., Djundja, D., Barthel, G., Elbert, T., & Rockstroh, B. (2005). Long-Term Stability of Improved Language Functions in Chronic Aphasia After Constraint-Induced Aphasia Therapy. *Stroke*, *36*(7), 1462–1466. https://doi.org/10.1161/01.STR.0000169941.29831.2a

Meunier, F. (2012). Formulaic Language and Language Teaching. *Annual Review of Applied Linguistics*, *32*, 111–129. https://doi.org/10.1017/S0267190512000128

Mikesell, L. (2009). Conversational Practices of a Frontotemporal Dementia Patient and His Interlocutors. *Research on Language & Social Interaction*, *42*(2), 135–162. https://doi.org/10.1080/08351810902864552

Millar, N. (2010). The Processing of Malformed Formulaic Language. *Applied Linguistics*, *32*(2), 129–148. https://doi.org/10.1093/applin/amq035

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/0033-295X.101.2.343

Mondini, S., Jarema, G., Luzzatti, C., Burani, C., & Semenza, C. (2002). Why Is “Red Cross” Different from “Yellow Cross”?: A Neuropsychological Study of Noun–Adjective Agreement within Italian Compounds. *Brain and Language*, *81*(1–3), 621–634. https://doi.org/10.1006/brln.2001.2552

Mondini, S., Luzzatti, C., Saletta, P., Allamano, N., & Semenza, C. (2005). Mental representation of prepositional compounds: Evidence from Italian agrammatic patients. *Brain and Language*, *94*(2), 178–187. https://doi.org/10.1016/j.bandl.2004.12.005

Newton, C., Kirby, P., & Bruce, C. (2017). Getting into shape: the effect of Shape Coding on the spoken language production of two men with chronic aphasia. *Aphasiology*, *31*(12), 1459–1481. https://doi.org/10.1080/02687038.2017.1306639

Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, *33*(10), 1281–1303. https://doi.org/10.1016/0028-3932(95)00102-9

Nobis-Bosch, R., Springer, L., Radermacher, I., & Huber, W. (2011). Supervised Home Training of Dialogue Skills in Chronic Aphasia: A Randomized Parallel Group Study. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1118–1136. https://doi.org/10.1044/1092-4388(2010/09-0204)

Noël, G. (2008). Language impairment, family interaction and the design of a game. *Visible Language*, *42*(2), 143–157.

Orsmond, G. I., & Cohn, E. S. (2015). The Distinctive Features of a Feasibility Study: Objectives and Guiding Questions. *OTJR: Occupation, Participation and Health*, *35*(3), 169–177. https://doi.org/10.1177/1539449215578649

Papathanasiou, I., Coppens, P., & Potagas, C. (2013). Aphasia and Related Neurogenic Communication Disorders: Basic Concepts and Operational Definitions. In *Aphasia and Related Neurogenic Communication Disorders* (pp.

xix–xxiii). Burlington: Jones & Bartlett Learning.

Pekarek Doehler, S. (2016). More than an epistemic hedge: French je sais pas “I don’t know” as a resource for the sequential organization of turns and actions. *Journal of Pragmatics, 106*, 148–162. https://doi.org/10.1016/j.pragma.2016.06.014

Pichler, H., & Hesson, A. (2016). Discourse-pragmatic variation across situations, varieties, ages: I DON’T KNOW in sociolinguistic and medical interviews. *Language & Communication*, *49*, 1–18. https://doi.org/10.1016/j.langcom.2016.04.002

Pinker, S. (1999). *Words and rules: The ingredients of language*. London: Weidenfeld and Nicolson.

Potagas, C., Kasselimis, D. S., & Evdokimidis, I. (2013). Elements of neurology essential for understanding the aphasias. In I. Papathanasiou, P. Coppens, & C. Potagas (Eds.), *Aphasia and Related Neurogenic Communication Disorders* (pp. 23–48). Burlington: Jones & Bartlett Learning.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, processing and use* (pp. 23–35). Amsterdam, the Netherlands: John Benjamins Press. https://doi.org/10.1075/lllt.9.03rea

Riches, N. G. (2013). Treating the passive in children with specific language impairment: A usage-based approach. *Child Language Teaching and Therapy*, *29*(2), 155–169. https://doi.org/10.1177/0265659012466667

Rochon, E., Laird, L., Bose, A., & Scofield, J. (2005). Mapping therapy for sentence production impairments in nonfluent aphasia. *Neuropsychological Rehabilitation*, *15*(1), 1–36. https://doi.org/10.1080/09602010343000327

Rodrigues, I. T., & Castro-Caldas, a. (2014). Aphasia with recurring utterances: Old syndrome, new perspectives. *Aphasiology*, *28*(11), 1350–1363. https://doi.org/10.1080/02687038.2014.921884

Ruiter, M. B., Kolk, H. H. J., & Rietveld, T. C. M. (2010). Speaking in ellipses: The effect of a compensatory style of speech on functional communication in chronic agrammatism. *Neuropsychological Rehabilitation: An International Journal*, *20*(3), 423–458.

Ruiter, M. B., Kolk, H. H. J., Rietveld, T. C. M., & Feddema, I. (2013). Combining possibly reciprocally dependent linguistic parameters in the quantitative assessment of aphasic speakers' grammatical output. *Aphasiology*, *27*(3), 293–308. https://doi.org/10.1080/02687038.2012.710319

Ruiter, M. B., Rietveld, T. C. M., Hoskam, V., & Van Beers, M. M. A. (2016). An Exploratory Investigation of e-REST: Teletherapy for Chronically Agrammatic Speakers. *International Journal of Telerehabilitation*, *8*(1), 21–28. https://doi.org/10.5195/IJT.2016.6191

Saffran, E. M., & Martin, N. (1997). Effects of Structural Priming on Sentence Production in Aphasics. *Language and Cognitive Processes*, *12*(5–6), 877–882. https://doi.org/10.1080/016909697386772

Schlenck, K.-J., Schlenck, C., & Springer, L. (1995). Die Behandlung des schweren Agrammatismus: Reduzierte Syntax-Therapie (REST). Stuttgart: Georg Thieme Verlag.

Schmid, H.-J. (2016). Linguistic entrenchment and its psychological foundations. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* (pp. 435–452). Berlin, Germany: de Gruyter.

Schmitt, N., & Carter, R. (2004). Formulaic Sequences in Action: An Introduction. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 1–22). Amsterdam, the Netherlands: John Benjamins.

SentenceShaper®. Psycholinguistic Technologies, Inc. http://www.sentenceshaper.com/

Shaoul, C., Baayen, R. H., & Westbury, C. F. (2014). N-gram probability effects in a cloze task. *The Mental Lexicon*, *9*(3), 437–472. https://doi.org/10.1075/ml.9.3.04sha

Shapiro, L., Swinney, D., & Borsky, S. (1998). Online Examination of Language Performance in Normal and Neurologically Impaired Adults. *American Journal* *of Speech-Language Pathology*, *7*(1), 49–60. https://doi.org/10.1044/1058-0360.0701.49

Sidtis, D., Canterucci, G., & Katsnelson, D. (2009). Effects of neurological damage on production of formulaic language. *Clinical Linguistics & Phonetics*, *23*(4), 270–284. https://doi.org/10.1080/02699200802673242

Silagi, M. L., Hirata, F. N., & Mendonça, L. I. Z. de. (2014). Sentence production in rehabilitation of agrammatism: A case study. *Dementia & Neuropsychologia*, *8*(3), 297–301. https://doi.org/10.1590/S1980-57642014DN83000015

Simmons-Mackie, N., & Damico, J. S. (1997). Reformulating the definition of compensatory strategies in aphasia. *Aphasiology*, *11*(8), 761–781. https://doi.org/10.1080/02687039708250455

Simmons-Mackie, N., Kingston, D., & Schultz, M. (2004). "Speaking for Another": The Management of Participant Frames in Aphasia. *American Journal of Speech-Language Pathology*, *13*(2), 114–127.

Sinclair, J. M. (1987). Collocation: Progress report. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday* (pp. 319–331). Amsterdam: John Benjamins.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, *175*, 111–122. https://doi.org/10.1016/j.bandl.2017.10.004

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 776–784. https://doi.org/10.1037/a0022531

Siyanova-Chanturia, A., & Martinez, R. (2015). The Idiom Principle Revisited. *Applied Linguistics*, *36*(5), 549–569. https://doi.org/10.1093/applin/amt054

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism*, *18*(3), 419–437. https://doi.org/10.1017/S1366728914000674

Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: collocations involving the word of. *Brain and Language*, *83*(2), 227–236. https://doi.org/10.1016/S0093-934X(02)00032-9

Springer, L., Huber, W., Schlenck, K.-J., & Schlenck, C. (2000). Agrammatism: Deficit or compensation? Consequences for aphasia therapy. *Neuropsychological Rehabilitation: An International Journal*, *10*(3), 279–309. https://doi.org/10.1080/096020100389165

Stahl, B., Henseler, I., Turner, R., Geyer, S., & Kotz, S. A. (2013). How to engage the right brain hemisphere in aphasics without even singing: evidence for two paths of speech recovery. *Frontiers in Human Neuroscience*, *7*, 35. https://doi.org/10.3389/fnhum.2013.00035

Stahl, B., & Kotz, S. A. (2014). Facing the music: three issues in current research on singing and aphasia. *Frontiers in Psychology*, *5*, 1033. https://doi.org/10.3389/fpsyg.2014.01033

Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G., & Pulvermüller, F. (2016). Using language for social interaction: Communication mechanisms promote recovery from chronic non-fluent aphasia. *Cortex*, *85*, 90–99. https://doi.org/10.1016/j.cortex.2016.09.021

Stahl, B., & Van Lancker Sidtis, D. (2015). Tapping into neural resources of communication: formulaic language in aphasia therapy. *Frontiers in Psychology*, *6*, 1526. https://doi.org/10.3389/fpsyg.2015.01526

Stefanowitsch, A., & Flach, S. (2016). The corpus-based perspective on entrenchment. In H.-J. Schmid (Ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge* (pp. 101–127). Berlin, Germany: de Gruyter.

Swinburn, K., Best, W., Beeke, S., Cruice, M., Smith, L., Pearce-Willis, E., … McVicker, S. (2018). A concise patient reported outcome measure for people with aphasia: the aphasia impact questionnaire 21. *Aphasiology*. https://doi.org/10.1080/02687038.2018.1517406

Swinburn, K., Porter, G., & Howard, D. (2004). The Comprehensive Aphasia Test.

Hove, UK: Psychology Press.

Taylor, J. R. (2002). *Cognitive grammar*. Oxford, UK: Oxford University Press.

Theakston, A. L., Ibbotson, P., Freudenthal, D., Lieven, E., & Tomasello, M. (2015). Productivity of Noun Slots in Verb Frames. *Cognitive Science*, *39*(6), 1369–1395. https://doi.org/10.1111/cogs.12216

Tomasello, M. (2003). Usage-Based Linguistics. In M. Tomasello (Ed.), *Constructing a language: A usage-based theory of language acquisition* (pp. 1–7). Cambridge, Mass.; London: Harvard University Press.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, Mass.; London: Harvard University Press.

Tremblay, A., & Baayen, H. (2010). Holistic Processing of Regular Four-word Sequences: A Behavioural and ERP study of the effects of structure, frequency and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151–173). London: The Continuum International Publishing Group.

Tremblay, A., Derwing, B., & Libben, G. (2009). Are Lexical Bundles Stored and Processed as Single Units? In *Proc. 23rd Northwest Linguistics Conference - Working Papers of the Linguistics Circle of the University of Victoria* (Vol. 19, pp. 258–279). Victoria BC CDA. Retrieved from https://journals.uvic.ca/index.php/WPLC/article/view/5666/2191

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. *Language Learning*, *61*(2), 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Trueswell, J. C. (1996). The Role of Lexical Frequency in Syntactic Ambiguity Resolution. *Journal of Memory and Language*, *35*(4), 566–585. https://doi.org/10.1006/jmla.1996.0030

Tsui, A. B. M. (1991). The pragmatic functions of “I don’t know.” *Text - Interdisciplinary Journal for the Study of Discourse*, *11*(4), 607–622. https://doi.org/10.1515/text.1.1991.11.4.607

Tyler, L. K. (1985). Real-time comprehension processes in agrammatism: A case study. *Brain and Language*, *26*(2), 259–275. https://doi.org/10.1016/0093-934X(85)90042-2

Tyler, L. K., & Cobb, H. (1987). Processing bound grammatical morphemes in context: The case of an aphasic patient. *Language and Cognitive Processes*, *2*(3–4), 245–262. https://doi.org/10.1080/01690968708406934

Tyler, L. K., & Marslen-Wilson, W. (1981). Children's processing of spoken language. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 400–416. https://doi.org/10.1016/S0022-5371(81)90524-7

Tyler, L. K., Moss, H. E., Patterson, K., & Hodges, J. (1997). The Gradual Deterioration of Syntax and Semantics in a Patient with Progressive Aphasia. *Brain and Language*, *56*(3), 426–476. https://doi.org/10.1006/brln.1997.1857

Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726. https://doi.org/10.1038/35094573

Ullman, M. T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: Evidence from the production, reading, and judgment of inflection in aphasia. *Brain and Language*, *93*(2), 185–238. https://doi.org/10.1016/j.bandl.2004.10.001

Van Lancker, D. (1993). Nonpropositional speech in aphasia. In G. Blanken, J. Dittman, H. Grimm, J. C. Marshall, & C. W. Wallesch (Eds.), *Linguistic disorders and pathologies. An international handbook* (pp. 215–225). Berlin: Walter de Gruyter.

Van Lancker, D., & Kempler, D. (1987). Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain and Language*, *32*(2), 265–277. https://doi.org/10.1016/0093-934X(87)90128-3

Van Lancker Sidtis, D. (2004). When novel sentences spoken or heard for the first time in the history of the universe are not enough: toward a dual-process model of language. *International Journal of Language & Communication Disorders*, *39*(1), 1–44. https://doi.org/10.1080/13682820310001601080

Van Lancker Sidtis, D. (2012). Formulaic Language and Language Disorders. *Annual Review of Applied Linguistics*, *32*, 62–80. https://doi.org/10.1017/S0267190512000104

Van Lancker Sidtis, D., Choi, J., Alken, A., & Sidtis, J. J. (2015). Formulaic Language in Parkinson's Disease and Alzheimer's Disease: Complementary Effects of Subcortical and Cortical Dysfunction. *Journal of Speech, Language, and Hearing Research*, *58*, 1493–1507. https://doi.org/10.1044/2015_JSLHR-L-14-0341

Van Lancker Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology*, *20*(5), 411–426. https://doi.org/10.1080/02687030500538148

Van Lancker Sidtis, D., & Rallon, G. (2004). Tracking the incidence of formulaic expressions in everyday speech: methods for classification and verification. *Language & Communication*, *24*(3), 207–240. https://doi.org/10.1016/j.langcom.2004.02.003

Van Lancker Sidtis, D., & Yang, S. (2017). Formulaic language performance in left- and right-hemisphere damaged patients: structured testing. *Aphasiology*, *31*(1), 82–99. https://doi.org/10.1080/02687038.2016.1157136

Varley, R. (2011). Rethinking aphasia therapy: A neuroscience perspective. *International Journal of Speech-Language Pathology*, *13*(1), 11–20. https://doi.org/10.3109/17549507.2010.497561

Varley, R., Cowell, P. E., Dyson, L., Inglis, L., Roper, A., & Whiteside, S. P. (2016). Self-Administered Computer Therapy for Apraxia of Speech: Two-Period Randomized Control Trial with Crossover. *Stroke*, *47*(3), 822–828. https://doi.org/10.1161/STROKEAHA.115.011939

Vilkaite, L., & Schmitt, N. (2017). Reading Collocations in an L2: Do Collocation Processing Benefits Extend to Non-Adjacent Collocations? *Applied Linguistics*, *amx030*, 1–27. https://doi.org/10.1093/applin/amx030

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Wayland, S. C., Berndt, R. S., & Sandson, J. R. (1996). Aphasic patients' sensitivity to structural and meaning violations when monitoring for nouns and verbs in sentences. *Neuropsychology*, *10*(4), 504–516. https://doi.org/10.1037/0894-4105.10.4.504

Wechsler, D. (2011). Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II). San Antonio, TX: Psychological Corporation.

Whiteside, S. P., Inglis, A. L., Dyson, L., Roper, A., Harbottle, A., Ryder, J., … Varley, R. A. (2012). Error reduction therapy in reducing struggle and grope behaviours in apraxia of speech. *Neuropsychological Rehabilitation*, *22*(2), 267–294. https://doi.org/10.1080/09602011.2011.639614

Whitworth, A., Budiman, P., Beales, A., Wynn, S., Cartwright, J., & Claessen, M. (2018). Just how stable is our discourse? A hurdle for measuring therapeutic change (or not). *Aphasiology*, *32*(sup1), 246–247. https://doi.org/10.1080/02687038.2018.1493427

Wilder, W. I., & Diamond, A. L. (1990). Some Like It Hot. In S. Thomas (Ed.), *Best American Screenplays 2 (1st ed)* (pp. 80–146). New York: Crown Publishers.

Wray, A. (2002a). Patterns of Formulaicity in Aphasic Language. In *Formulaic Language and the Lexicon* (pp. 217–246). Cambridge: Cambridge University Press.

Wray, A. (2002b). The Whole and the Parts. In A. Wray (Ed.), *Formulaic Language and the Lexicon* (pp. 3–18). Cambridge: Cambridge University Press.

Wray, A. (2012a). Patterns of formulaic language in Alzheimer's disease: implications for quality of life. *Quality in Ageing and Older Adults*, *13*(3), 168–175. https://doi.org/10.1108/14717791211264034

Wray, A. (2012b). What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics*, *32*, 231–254. https://doi.org/10.1017/S026719051200013X

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, *20*(1), 1–28. https://doi.org/10.1016/S0271-5309(99)00015-4

Zimmerer, V. C., Coleman, M., Hinzen, W., & Varley, R. A. (in prep). Reliance on common word combinations correlates with degree of syntactic impairment in aphasia.

Zimmerer, V. C., Newman, L., Thomson, R., Coleman, M., & Varley, R. (2018). Automated analysis of language production in aphasia and right hemisphere damage: frequency and collocation strength. *Aphasiology*, *32*(11), 1267–1283. https://doi.org/10.1080/02687038.2018.1497138

Zimmerer, V. C., Spencer, H., Douglas, M., Turkington, D., Hinzen, W., & Varley, R. A. (2017). Online sensitivity to linguistic information distinguishes people with and without schizophrenia. Newcastle, UK.

Zimmerer, V. C., & Wibrow, M. (2015). Frequency in Language Analysis Tool.

Zimmerer, V. C., Wibrow, M., & Varley, R. A. (2016). Formulaic language in people with probable Alzheimer's Disease: a frequency-based approach. *Journal of Alzheimer's Disease*, *53*(3), 1145–1160. https://doi.org/10.3233/JAD-160099

Appendix

A.Study 1 Appendices

Appendix A.1: *Number of uni-, bi- and trigrams produced by PWA and CPs in the naturalistic dataset (all utterances appear in the spoken BNC).*

	BL	**DM**	**DC**	**GL**	**JH**	**KK**	**PM**	**PG**	**SC***
Bigram tokens: **CPs**	414	1268	1106	731	695	1054	1028	1228	3636
(types)	(317)	(872)	(836)	(556)	(534)	(709)	(699)	(862)	(2082)
Bigram tokens: **PWA**	101	113	413	145	193	224	109	264	247
(types)	(30)	(74)	(262)	(83)	(68)	(115)	(66)	(158)	(132)
Trigram tokens: **CPs**	236	781	653	415	413	607	632	747	2160
(types)	(210)	(674)	(610)	(378)	(384)	(507)	(525)	(666)	(1751)
Trigram tokens: **PWA**	57	14	196	49	62	102	43	79	96
(types)	(14)	(11)	(172)	(34)	(19)	(58)	(31)	(69)	(61)
Average trigram t-scores (token-based): ***CPs***	*16*	*16*	*13*	*14*	*11*	*13*	*17*	*14*	*14*
Average trigram t-scores (token-based): ***PWA***	*89*	*26*	*18*	*13*	*83*	*49*	*19*	*18*	*38*
Average trigram t-scores (type-based): ***CPs***	*12*	*10*	*11*	*12*	*11*	*9*	*10*	*10*	*9*
Average trigram t-scores (type-based): ***PWA***	*20*	*14*	*16*	*11*	*20*	*20*	*17*	*16*	*16*

In the last four rows, average trigram t-scores are shown, note the high inter-subject variability in the aphasic group.
** Based on eight conversation samples; whereas the values of the remaining dyads are based on four conversation samples.*

Appendix A.2: *Rating system for IDK conversational functions.*

<table>
<tr><th colspan="2">Code</th><th>Conversational function</th><th>Examples
(PWA = participant with aphasia;
CP = conversation partner)</th></tr>
<tr><td>LOK</td><td>Lack of knowledge</td><td>---</td><td>(KK_2.2_3_4)
CP: where do the other people live from the stroke club are they all round [NAME] area?
PWA: ⌈ ehm: I don’t know ⌉ where
⌊ ((shakes head and folds arms)) ⌋
PWA: ⌈no ⌉
CP: ⌊you don’t⌋ know↓ (.)°ri:ght°

(BL_4.2_7_8)
CP: what will you have for y' pudding
(1.4)
PWA: I don't know
(3.6)
PWA: ⌈bu:m bu:m bu:m bu:m ⌉
| ((interrogative gesture with left hand, palm |
⌊facing up)) ⌋
(.)
CP: whatever they've got
(0.3)
PWA: °yep°
(3.6)
PWA: ice cream
(0.3)
CP: ice cream hm</td></tr>
</table>

Code		Conversational function	Examples (PWA = participant with aphasia; CP = conversation partner)
INT	Interpersonal	- Avoiding disagreement - Prefacing disagreement - Avoiding assessment - Avoiding commitment to the answer / statement - Minimising compliments - Hedging	*Avoiding disagreement (DM_1.2_1)* PWA: pork (1.8) nice (.) ⌈we:: ⌉ CP: ⌊wha ⌋ roast pork (0.6) PWA: u::m dunno:, (0.8) (gristles) (1.6) CP: grill⌈ed,⌉ ⌈like ⌉ a chop, PWA: ⌊gri⌋ lls ⌊grilled⌋ PWA: gris (.) yeah *Avoiding assessment (JH_4.3_2_3_4_5)* CP: what you could do is see if the railway's running. ⌈(2.3) ⌉ ⌊ ((CP and PWA start eating)) ⌋ PWA: mmm (1.5) ⌈°I don't⌉ know° CP: ⌊then ⌋ ⌈ (1.5) ⌉ ⌊ ((cutlery noise))⌋ PWA: (about nine or ten ↑pound) CP: don't matter (how much it is) does ↑it?

Revised definition of "hedging":

A speaker distances themselves from what they are saying, but they don't have any direct evidence for it, i.e. the speaker is not fully committed to a proposition. Speculation is different from hedging because a speaker acknowledges that they don't have enough knowledge to form a definite opinion. Raters should distinguish between an IDK token that **functions** as a hedge versus one that **accompanies** it.

Practical implications for final IDK ratings:

After assessing IRR, all instances marked as hedges by R1 were subsequently checked against the revised definition. As a result, 12 out of 16 CP group cases were designated to not be instances of hedging (but three of the 12 remained in the INT domain); in the aphasic group, three out of six cases, were no longer judged to be hedges (but two of the three remained in the INT domain). Thus, final IDK ratings consist of both individual and consensus ratings.

<table>
<tr><th colspan="2">Code</th><th>Conversational function</th><th>Examples
(PWA = participant with aphasia;
CP = conversation partner)</th></tr>
<tr><td>TC</td><td>Turn-constructional</td><td>- Turn holding device (e.g., turn construction difficulty / verbalize on-line planning processes)

- Floor yielding device

- Taking a turn

- Other</td><td>Verbalizing on-line planning process (KK_1.6_1_2_3)

CP: well there’s you (1.0) what job was you doing?
PWA: right ⌈ (2.3) ⌉ oil
⌊ ((puts pen and paper at the side)) ⌋
CP: oil
PWA: oil
CP: what did (.) what did you do in oil?
PWA: (1.8) I don’t know↑
CP: °yourr:↑°
PWA: riggers
CP: rigger
PWA: ⌈yeah ⌉ but↓
CP: ⌊°yeah° ⌋
CP: you’re a rigger
PWA: ⌈ rigger ⌉
⌊ ((reaching for his mug)) ⌋
(4.2)

Yielding conversational floor (SC_7_1_2)

PWA: erm (.) Catherine
CP: yeah↑
are you <u>fil</u>ming?
PWA: yeah
CP: oh (.) Catherine what Catherine <u>Jon</u>es?
(1.0)
PWA: yeah
CP: yeah?
(1.0)
PWA: erm (.) I don’t know↓
(1.5)
CP: don’t know about <u>what</u> love↓
(1.0)
oh what the (.) the (.) erm
PWA: e-mail</td></tr>
</table>

<table>
<tr><th colspan="2">Code</th><th>Conversational function</th><th>Examples
(PWA = participant with aphasia;
CP = conversation partner)</th></tr>
<tr><td>M</td><td>Multifunctional</td><td>e.g., mixed turn-constructional and interpersonal</td><td>Mixed TC and INT (SC_1_1)

CP: when the pict=when she showed you the pictures
PWA: yeah
(2.0)
CP: and you had to say some=say what it was
PWA: hmm
CP: you got those (.) you did those really well
PWA: er yeah
PWA: one or two but (.) but er: (2.0) picture
CP: hmm?
(.)
PWA: er above or
⌈ (4.5) ⌉
CP: ⌊ ((nods)) ⌋
PWA: but I (.) I don’t know ((looks thoughtful))
CP: the one that said underneath
CP: and on top of
PWA: yeah
CP: did you find that hard?
(3.5)
PWA: ye:s↓ sort of</td></tr>
<tr><td>U</td><td>Unclear function</td><td>---</td><td>Not enough context - beginning of clip (BL_3.2_1_2)

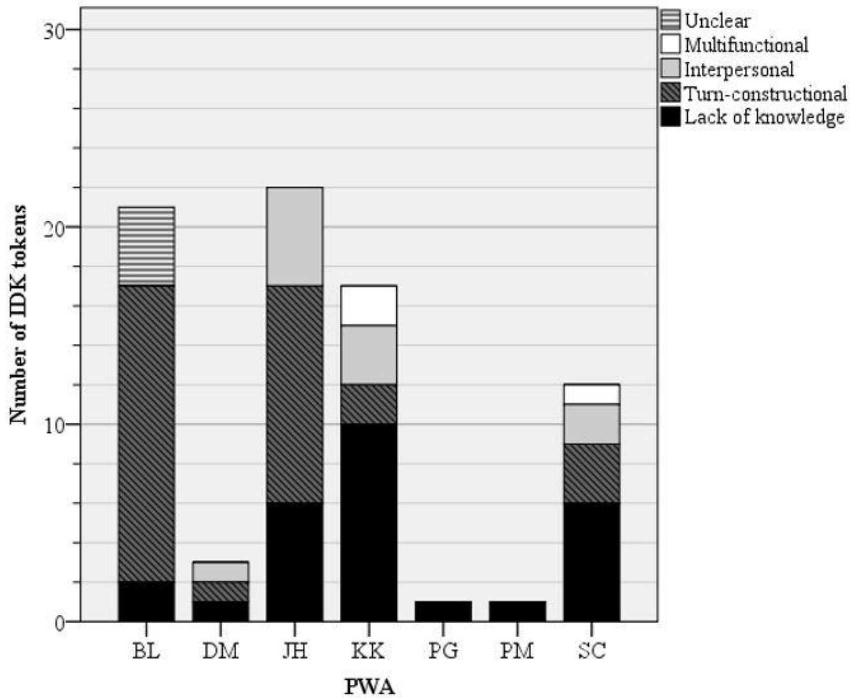
PWA: I don't know for- er=
CP: =I don't know how to do the(se)
(0.4)
PWA: ⌈ di di di di::::: ⌉
⌊ ((sings and shakes head)) ⌋</td></tr>
</table>

Appendix A.3: *CA transcription symbols used in extracts 1-3 and in Appendix A.2.*

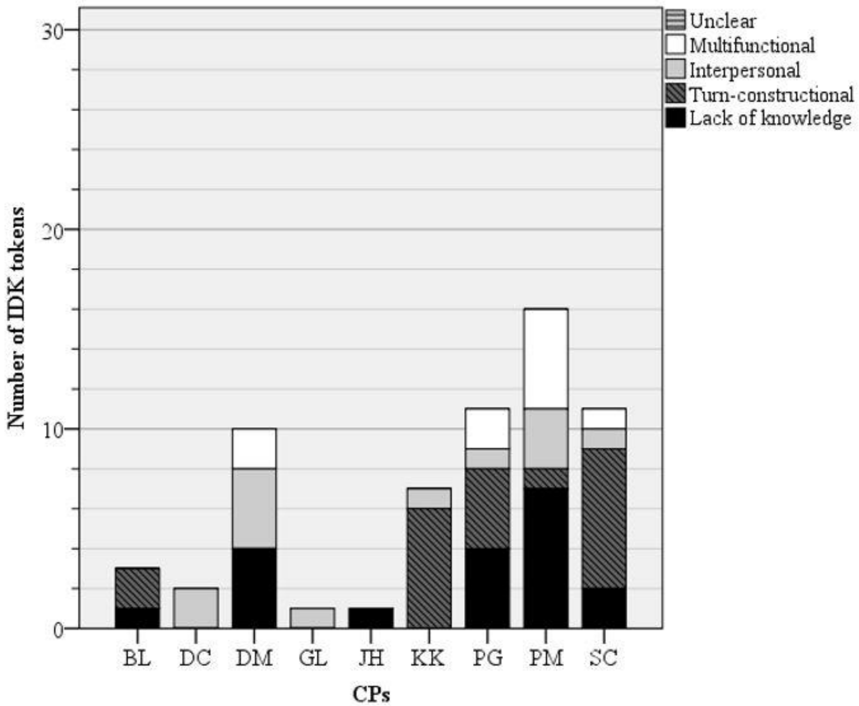
PWA:	participant with aphasia
CP:	conversation partner
⌈I don't know ⌉ ⌊((shakes head)) ⌋	square brackets show where there is overlap
((looks at CP))	text in double brackets gives a description of what people are doing
what (will)	single brackets indicate a syllable/word/phrase that is hard to understand
(2.5)	a number in single brackets denotes a pause in seconds, e.g. 2.5 seconds
(.)	micropause of less than 1 second
er:	colons indicate a lengthening of the sound or syllable they follow
?	a question mark indicates a rising tone
.	a full stop indicates a falling tone
,	a comma indicates a continuing tone, as if a speaker will say more
=	an equals sign marks where there is no hearable gap between two words
but-	a single dash indicates a word or sound that is abruptly cut off
I ↑need	an upward arrow marks a noticeable upward shift in tone
ye:s↓ sort of	a downward arrow marks a noticeable downward shift in tone
I don't know	underlining indicates emphasis
°yeah what°	degree signs indicate quiet speech, two or more indicate very quiet speech
OOH	capital letters indicate loud speech
>em, em, em<	lesser than/greater than signs indicate sections of speech that are faster

Appendix A.4: *IDK profiles for individual speakers in the two speaker groups*

Participants with aphasia (PWA)



Conversation partners (CPs)



B. Study 2 Appendices

Appendix B.1: *Underlying trigram structures of WMT stimuli*

	Stronger collocation	Weaker collocation	Underlying structure: Stronger collocation	Underlying structure: Weaker collocation	Same structure?
First-word manipulations	**at** the back	**up** the back	prep + det(article) + noun		yes
	on the bottom	**with** the bottom	prep + det(article) + noun		yes
	in the city	**at** the city	prep + det(article) + noun		yes
	round the corner	**near** the corner	prep + det(article) + noun		yes
	in the country	**from** the country	prep + det(article) + noun		yes
	on the door	**from** the door	prep + det(article) + noun		yes
	in the future	**on** the future	prep + det(article) + noun		yes
	in the garden	**for** the garden	prep + det(article) + noun		yes
	on the ground	**from** the ground	prep + det(article) + noun		yes
	at this point	**from** this point	prep + det(demonstrative) + noun		yes
	in the process	**for** the process	prep + det(article) + noun		yes
	in the summer	**of** the summer	prep + det(article) + noun		yes
	on the table	**up** the table	prep + det(article) + noun		yes
	in the town	**from** the town	prep + det(article) + noun		yes
	go to sleep	**have** to sleep	verb + prep + noun	verb + *to*-infinitive + verb	**no**

	a few days	**the** few days	det(article) + det(quantifier) + noun		yes
	the other side	**that** other side	det(article) + det + noun	det(demonstrative) + det + noun	yes
	cup of tea	**box** of tea	noun + prep + noun		yes
Second-word manipulations	in **the** bin	in **that** bin	prep + det(article) + noun	prep + det(demonstrative) + noun	yes
	in **the** car	in **this** car	prep + det(article) + noun	prep + det(demonstrative) + noun	yes
	in **this** case	in **one** case	prep + det(demonstrative) + noun	prep + det(number) + noun	yes
	in **some** cases	in **three** cases	prep + det(quantifier-unknown amount) + noun	prep + det(number - specified amount) + noun	yes
	all **the** people	all **three** people	det(quantifier) + det(article) + noun	det(all) + det(number) + noun	yes
	in **a** position	in **no** position	prep + det(article) + noun	prep + det(negation) + noun	yes
	in **my** view	in **the** view	prep + det(possessive) + noun	prep + det(article) + noun	yes
	on **the** bus	on **his** bus	prep + det(article) + noun	prep + det(possessive) + noun	yes
	on **the** floor	on **your** floor	prep + det(article) + noun	prep + det(possessive) + noun	yes
	go **to** bed	go **in** bed	verb + prep + noun		yes
	a **little** bit	a **lovely** bit	det(article) + det(quantifier) + noun	det(article) + adjective + noun	**no**
	the **other** day	the **only** day	det(article) + det + noun	det(article) + adjective + noun	**no**
	a **great** deal	a **fair** deal	det(article) + adjective + noun		yes
	the **same** sort	the **light** sort	det(article)+ adjective + noun		yes

	a **long** time	a **nice** time	det(article) + adjective + noun		yes
	a **hundred** percent	a **twenty** percent	det(article) + det(quantifier) + noun		yes
	the **first** place	the **one** place	det(article) + det(ordinal) + noun	det(article) + det(demonstrative) + noun	yes
	on **the** phone	on **your** phone	prep + det(article) + noun	prep + det(possessive) + noun	yes
Binomials	**body** and soul	**spirit** and soul	noun + coordinating conjunction + noun		yes
	bride and groom	**horse** and groom			
	shirt and tie	**hat** and tie			
	milk and sugar	**fruit** and sugar			
	fish and chips	**bread** and chips			
	knife and fork	**plate** and fork			
	ladies and gentlemen	**scholars** and gentlemen			
	salt and pepper	**juice** and pepper			
	shoes and socks	**books** and socks			
	wine and cheese	leeks and cheese			
	gin and tonic	**lime** and tonic			
	cats and dogs	**birds** and dogs			

Note. Critical words of trigrams are highlighted in bold; det = determiner; prep = preposition

Appendix B.2: *Complete list of stimuli including the frequency-related measures of trigrams and target words*

Condition	Pair #	Stronger collocation	Weaker collocation	Freq of critical word strong	Freq of critical word weak	Sentential context stronger coll.	Sentential context weaker coll.	Target position (both sent.)	Trigram freq (strong / weak)	Trigram t-score (strong / weak)
F	1	at the back	up the back	47616	34930	I can see the mouse. It's running at the back of the park.	I can see the rat. It's running up the back of the park.	10	589 / 20	24 / 4
F	2	on the bottom	with the bottom	81083	47043	We should complete all the paperwork. We could perhaps start on the bottom of this form.	We should complete all the paperwork. We could perhaps start with the bottom of this form.	13	151 / 5	12 / 2
F	3	in the city	at the city	142193	47616	We had a great evening. The pub in the city of London was very cosy.	We had a great afternoon. The pub at the city of London was very cosy.	10	150 / 12	12 / 3
F	4	round the corner	near the corner	7171	1403	It's a tricky situation. The lorry round the corner of the side street was parked there since yesterday.	It's a tricky situation. The lorry near the corner of the side street was parked there since yesterday.	9	138 / 3	12 / 2
F	5	in the country	from the country	142193	24121	I'm interested in wildlife photography. There is a great exhibition in the country next to Brazil.	I'm interested in ceramic sculptures. There is a great exhibition from the country next to Peru.	13	212 / 7	15 / 3
F	6	on the door	from the door	81083	24121	Claire's new mirror looks lovely. I can see it on the door of her room.	Gwen's new bookshelf looks lovely. I can see it from the door of her room.	12	132 / 7	11 / 3

F	7	in the future	on the future	142193	81083	We see each other twice in June. In the future plans, we will discuss our ideas.	We see each other twice in May. On the future plans, we will discuss our ideas.	10	230 / 14	18 / 4
F	8	in the garden	for the garden	142193	67047	I think your shopping trip was a success. The chair in the garden is comfortable.	I think your shopping trip was a success. The chair for the garden is beautiful.	13	194 / 9	14 / 3
F	9	on the ground	from the ground	81083	24121	I will take care of the laundry tonight. The picnic blanket on the ground looks dirty.	I will take care of the laundry tomorrow. The picnic blanket from the ground looks dirty.	14	141 / 19	12 / 4
F	10	at this point	from this point	47616	24121	The new hotel looks massive already. At this point, I even think it is higher than the church.	The new building looks massive already. From this point, I even think it is higher than the city hall.	9	115 / 8	11 / 3
F	11	in the process	for the process	142193	67047	We should check whether we have enough flowers. They need them in the process of decorating.	We should check whether we have enough ivy. They need them for the process of decorating.	14	93 / 3	10 / 2
F	12	in the summer	of the summer	142193	174549	I like iced coffee a lot. This was my favourite drink in the summer when it was really hot.	I like mango juice a lot. This was my favourite drink of the summer when it was really warm.	14	193 / 20	14 / 4
F	13	on the table	up the table	81083	34930	I have seen a spider. It was crawling on the table in the living room.	I have seen a beetle. It was crawling up the table in the kitchen.	11	242 / 5	16 / 2
F	14	in the town	from the town	142193	24121	The results from the election are official. Many people in the town support the new mayor.	The results from the election are official. Many people from the town support the new president.	12	156 / 15	12 / 4

F	15	go to sleep	have to sleep	31180	77272	Alex and Katy's twins are adorable. When they go to sleep, however, they cry a lot.	Paul and Anne's children are adorable. When they have to sleep, however, they cry a lot.	11	114 / 5	11 / 2
F	16	a few days	the few days	206203	409715	We could go to the supermarket now. It will not be too crowded except a few days before Christmas.	We could go to the lido now. It will not be too crowded except the few days before Easter.	17	100 / 2	10 / 1
F	17	the other side	that other side	409715	227030	There are three clients visiting this afternoon. They are interested in the other side of the housing market.	There are three customers visiting this morning. They are interested in that other side of the housing market.	14	560 / 4	24 / 2
F	18	cup of tea	box of tea	1160	1356	Mike is queuing up. He's buying a cup of tea for his sister.	Chris is queuing up. He's buying a box of tea for his daughter.	10	452 / 2	21 / 1
S	19	in the bin	in that bin	409715	227030	There is one more task I have for you. Please put the letter in the bin for me.	There is one more task I have for you. Please put the letter in that bin for me.	16	93 / 3	10 / 2
S	20	in the car	in this car	409715	58180	Rachel and I spoke about the Christmas presents. She rang me when she was in the car yesterday evening.	Lucy and I spoke about the birthday presents. She rang me when she was in this car yesterday morning.	17	392 / 7	20 / 3
S	21	in this case	in one case	58180	58149	Let's check the paperwork. In this case he received a bill.	Let's check the folder. In one case he received a bill.	7	181 / 6	13 / 2
S	22	in some cases	in three cases	20590	17738	The organisation of the concert seems strange. In some cases they sold cheaper tickets.	The organisation of the festival seems odd. In three cases they sold cheaper tickets.	10	92 / 2	10 / 1

S	23	all the people	all three people	409715	17738	It's an exciting programme. All the people from our team are looking forward to the event.	It's an interesting programme. All three people from our team are looking forward to the conference.	7	101 / 5	10 / 2
S	24	in a position	in no position	206203	59809	Let me summarize the situation with the house. We're in a position to exchange contracts at any time.	Let me summarize the situation with the flat. We're in no position to exchange contracts at any time.	12	152 / 3	12 / 2
S	25	in my view	in the view	23518	409715	A lot of people went to vote on Monday. In my view of the situation this was a good result.	A lot of people went to vote last week. In the view of the government, this was a good result.	12	104 / 11	10 / 3
S	26	on the bus	on his bus	409715	14047	Tom just called. He is going to come to me on the bus on Sunday.	Marc just called. He is going to come to me on his bus on Monday.	13	118 / 3	11 / 2
S	27	on the floor	on your floor	409715	29544	I know what I would do. If I found a pound on the floor I'd pick it up.	I know what I would do. If I found a pound on your floor I'd pick it up.	14	376 / 2	19 / 1
S	28	go to bed	go in bed	233692	142193	Remember to turn the heating off. You can do that when you go to bed for a nap.	Remember to close all the windows. You can do that when you go in bed for a nap.	15	185 / 3	14 / 2
S	29	a little bit	a lovely bit	9071	2398	We went to the restaurant yesterday. We had a little bit of fish and some potatoes.	We went to the pub yesterday. We had a lovely bit of fish and some potatoes.	11	1716 / 2	41 / 1

S	30	the other day	the only day	15093	13663	I'm happy about the result of the exam. I was helping John the other day I saw him.	I'm happy about the result of the essay. I was helping Phil the only day I saw him.	15	591 / 7	24 / 3
S	31	a great deal	a fair deal	3878	1416	I went to the garage yesterday evening. The salesman luckily offered me a great deal to replace the engine.	I went to the garage this morning. The salesman luckily offered me a fair deal to replace the engine.	15	404 / 10	20 / 3
S	32	the same sort	the light sort	6843	1462	I'm not going to the shop. The same sort of material is available online.	I'm not going to the market. The light sort of material is available online.	9	145 / 2	12 / 1
S	33	a long time	a nice time	6165	6247	Your plan looks good. Six weeks is a long time for travelling in Italy.	Your plan looks good. Eight weeks is a nice time for travelling in Australia.	10	667 / 22	26 / 5
S	34	a hundred percent	a twenty percent	10513	9507	I have checked the radiator. I would say there is a hundred percent chance that it works.	I have checked the radiator. I would say there is a twenty percent chance that it works.	13	86 / 5	9 / 2
S	35	the first place	the one place	10027	58149	We might move to South London. There was a balcony in the first place that we saw yesterday.	We might move to Southampton. There was a balcony in the one place that we saw yesterday.	13	260 / 4	16 / 2
S	36	on the phone	on your phone	409715	29544	I can't believe it. I wish to have our boss here on the phone and talk about this proposed pay rise.	I can't believe it. I wish to have our manager here on your phone and talk about this proposed pay rise.	14	378 / 5	19 / 2
B	37	body and soul	spirit and soul	1124	443	I like running. It is good for body and soul and is fun.	I like hiking. It is good for spirit and soul and is fun.	10	-	-

B	38	bride and groom	horse and groom	30	636	I don't like this picture. The bride and groom both look tired.	I don't like this painting. The horse and groom both look tired.	9	-	-
B	39	shirt and tie	hat and tie	219	335	Peter looked great. He had his best shirt and tie on.	Andrew looked great. He had his best hat and tie on.	10	-	-
B	40	milk and sugar	fruit and sugar	667	330	I've sent Jamie to the shop. Milk and sugar are an essential addition to our afternoon tea.	I've sent Johnny to the shop. Fruit and sugar are an essential addition to our afternoon tea.	9	-	-
B	41	fish and chips	bread and chips	750	628	It's a great place. I love the fish and chips served in this pub.	It's a great place. I love the bread and chips served in this pub.	10	-	-
B	42	knife and fork	plate and fork	185	310	Please bring along my rucksack. And don't forget your knife and fork for the BBQ.	Please bring along your jacket. And don't forget your plate and fork for the BBQ.	12	-	-
B	43	ladies and gentlemen	scholars and gentlemen	437	5	Thanks for coming. Good morning ladies and gentlemen and welcome to our weekly jazz concert.	Thanks for coming. Good evening scholars and gentlemen and welcome to our weekly jazz concert.	8	-	-
B	44	salt and pepper	juice and pepper	294	163	That's unbelievable. I forgot the salt and pepper for our breakfast.	That's embarrassing. I forgot the juice and pepper for our breakfast.	8	-	-
B	45	shoes and socks	books and socks	553	894	Let's go to the shopping centre. We need to buy shoes and socks for him.	Let's go to the city centre. We need to buy books and socks for him.	13	-	-
B	46	wine and cheese	leeks and cheese	406	12	We need to remember to buy butter. Wine and cheese is already on the shopping list.	We need to remember to buy bread. Leeks and cheese is already on the shopping list.	10	-	-

B	47	gin and tonic	lime and tonic	54	38	This is a great barman. I want to go and get gin and tonic for all of us.	This is a great barman. I want to go and get lime and tonic for all of us.	14	-	-
B	48	cats and dogs	birds and dogs	173	281	It's common sense. Cats and dogs should not be kept indoors for too long.	It's common sense. Birds and dogs should not be kept indoors for too long.	6	-	-

Note. F = First-word manipulations; S = Second-word manipulations; B = Binomials

Appendix B.3: *Cloze task score sheet*

ID: _________________, Date: __________________

Cloze task with familiar expressions

Formulas listed in Van Lancker-Sidtis & Rallon (2004, pp. 22 ff.), extracted from 'Some Like It Hot':

	Circle: correct	/ incorrect
	(self-correction = SC)	
1) The train is about to depart. The station master shouts at the last few passengers: "All _________!"	✓	×
2) Peter knocks on the door of his manager's office. He hears the manager say: "Come ________!"	✓	×
3) James fears that he is late for the wedding. The cab driver looks at him and says: "We'll be there in no time. Don't _________."	✓	×
4) The fire is getting out of control and the room is filled with smoke. The fireman shouts: "Get out of __________."	✓	×
5) Caroline just finished reading a bedtime story to her children. When she leaves their room, she says: "Good ___________."	✓	×
6) John is offering some birthday cake to everyone. He puts the cake-stand on the table and says: "Help _______________."	✓	×
7) Julia is looking for a new outfit in the shop. The saleswoman points to a dress and asks: "How about that __________?"	✓	×
8) The teacher just explained the exercises to the students. One student asks what to do and the teacher replies: "I already told ________."	✓	×
9) Marc looks for his keys. His son asks where they are and Marc replies: "I don't _____________."	✓	×
10) Mary asks Hannah what time the concert starts. Hannah says: "I'm not quite _____________."	✓	×

ID: _______________, Date: ________________

11)
Kate asks the shop assistant if there is any fresh bread.
He says: "It's all gone. I'm terribly __________." ✓ ×

12)
The children are late for school.
Their mother shouts: "Let's _________." ✓ ×

13)
Paula is feeling full after eating dinner.
Joe offers her a piece of cake. She says: "No, thank ________." ✓ ×

14)
Tom asks his wife if she would mind buying some milk on her way home.
She replies: "Not at _________." ✓ ×

15)
Anna arrives at the birthday party and meets Jane for the first time.
She says "Pleased to meet _________." ✓ ×

16)
Ben apologized to Jeremy for keeping him waiting.
Jeremy smiled and said: "That's _________." ✓ ×

17)
Carol asks the policeman if she should go left or right.
He points to the right and says: "This _________." ✓ ×

18)
Rachel joins her friends in the café. They are in the middle of chatting about their holiday plans.
Rachel asks: "What are you talking _________?" ✓ ×

19)
Tina's daughter comes home from the playground with a bruised knee.
Tina asks her: "What _________?" ✓ ×

20)
Jane tells Claire about her idea to dye her hair pink.
Claire says: "You're out of your ___________." ✓ ×

Sum correct: ____

Sum SC: ____

Appendix B.4: *Pilot data: z-score differences per trigram pair and participant*

Condition	Pair	Participant										Average z-score difference	Stronger collocation of pair
		1	2	3	4	5	6	7	8	9	10		
First-word manipulations	1	0.53	1.01	-1.51	0.17	-1.13		0.31	0.57	0.35	0.03	0.04	*at the back*
	2	-1.17	-1.44	-1.22	1.03	0.01	0.48	0.33	0.69	0.11	0.71	-0.05	*on the bottom*
	3	0.76	0.54		0.11	0.61	-0.09	-0.36	0.28	0.06	-2.31	-0.04	*in the city*
	4	0.20	0.26	0.26	-0.55	0.58	5.17	-0.25	0.16		-1.05	0.53	*round the corner*
	5	1.86	1.01	0.01	-0.31	0.11	-1.21	2.12	-0.02	0.27	0.86	0.47	*in the country*
	6	0.74	-0.11	-0.57	1.36	-0.42	-0.73	1.82	-0.58	1.26	0.66	0.34	*on the door*
	7	0.93			-0.44	1.06	1.53	0.93	0.61	1.18		0.83	*in the future*
	8	0.46	0.43	0.38	0.14	0.73	0.23	-0.32	0.33	0.14	-0.55	0.20	*in the garden*
	9		1.24	-0.21	0.99	-0.44	-0.32	1.60	-1.52	1.15	3.87	0.71	*on the ground*
	10	0.09	-0.42	0.54	0.67	-1.97	-0.38	0.98	0.23	-0.53	-0.47	-0.13	*at this point*
	11	0.91	-0.12	-0.92	0.09	1.04	0.19	-0.87	-0.59	-1.03	-3.15	-0.44	*in the process*
	12	0.01		1.06	0.04	2.11	0.39	0.04	-0.32	1.32	1.02	0.63	*in the summer*
	13	-0.15	2.07	-0.14	0.10	-1.38	-0.01	-0.75	-0.42	-0.84	0.18	-0.13	*on the table*
	14	0.55	1.25	0.58	0.71	-0.01		0.20	-0.54	0.69	0.47	0.43	*in the town*
	15	-0.87	0.50	1.19	1.92	0.95	0.30	-0.02	6.06	2.78	1.60	1.44	*go to sleep*
	16	0.43	-0.76	0.06	-0.97	0.66	-0.30	0.05	-0.84	0.20	-1.11	-0.26	*a few days*
	17	0.37		-0.87		-1.62	-0.36	-1.05	0.60	0.16	0.52	-0.28	*the other side*
	18	0.54	0.18	0.44	0.08	0.61		1.14	0.71	1.71	-0.45	0.55	*cup of tea*

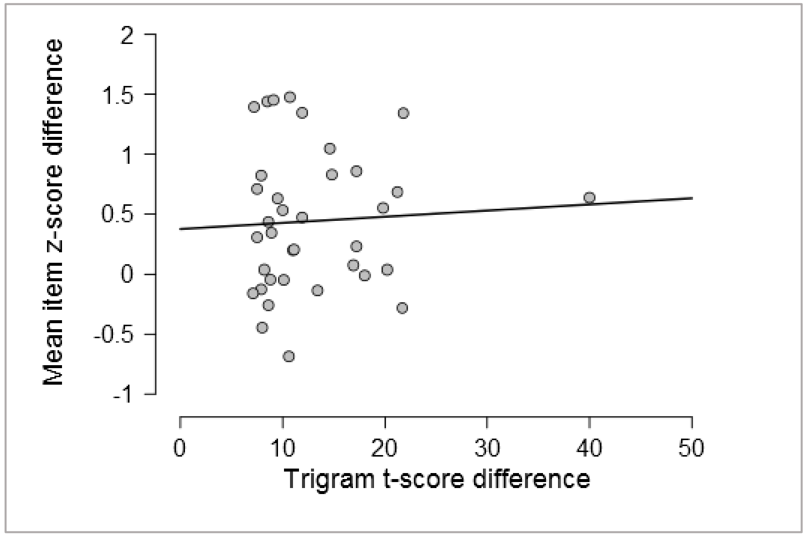
Second-word-manipulations	19	1.08		0.34	0.32	2.15	-0.05	0.26		1.39	1.07	0.82	*in the bin*
	20	0.77			1.52	1.44	-0.05	1.47	0.87	1.41	-0.57	0.86	*in the car*
	21	-0.44	1.95	-0.21	0.70	-1.90	-1.24	2.96	0.40	-0.41		0.20	*in this case*
	22	0.50	1.91	0.50	-0.17	-0.44	0.14	0.49		-3.05	0.45	0.04	*in some cases*
	23	-0.12	1.17	0.26	-0.64	0.46	-0.06	2.85	-0.89	1.98	-1.96	0.31	*all the people*
	24	-1.02	-1.26	0.93	-1.22	-0.38	-0.83	0.27	-2.43	-0.17	-0.73	-0.69	*in a position*
	25	4.07	0.73	0.03		1.92		2.39	1.27	2.58	-1.85	1.39	*in my view*
	26	0.71	1.25	1.92	3.20	1.98		1.32	0.57	0.40	1.72	1.45	*on the bus*
	27	0.89	-1.00	0.02	-0.64	0.73	0.10	0.83	0.21	-1.21		-0.01	*on the floor*
	28	2.28	0.45	1.56	2.10	1.46	0.77	1.64	0.63	0.69	1.86	1.35	*go to bed*
	29	2.74		1.30	-0.77	0.39	0.09	-1.72	3.69	0.66	-0.65	0.64	*a little bit*
	30	1.83	3.81	2.08	1.35	1.13	-0.61	2.18	0.66	-0.34		1.34	*the other day*
	31	-0.53	1.57	0.96	-1.63	0.69		-0.78	0.10	0.21		0.07	*a great deal*
	32						0.59		0.37	2.99	1.95	1.48	*the same sort*
	33	0.78		0.19	0.77	0.81	0.49	1.06	0.41	0.16	1.50	0.68	*a long time*
	34	0.43	-0.35	-0.66		-1.49	-0.05	-0.55	-0.83	1.70	0.38	-0.16	*a hundred percent*
	35			-0.03			0.77	0.54	1.36	2.59		1.05	*the first place*
	36	0.19	0.71	1.11	0.31	0.39	-1.87	0.81	0.32	0.55	-0.23	0.23	*on the phone*
Binomials	37	-1.29	2.62					2.26		2.32	0.51	1.28	*body and soul*
	38	0.01	2.13	-0.01		0.20	0.09	1.80	2.22	0.43	1.49	0.93	*bride and groom*
	39	0.89		0.05	-0.42	-0.33	0.14	0.46	-0.29	-0.28	-0.45	-0.02	*shirt and tie*
	40	0.51		0.38	1.87	0.65	0.03		1.04	2.38		0.98	*milk and sugar*
	41	1.26		3.34	1.46	-1.87		2.01	1.63	0.88	0.98	1.21	*fish and chips*

	42	0.19	1.14	-0.09	0.42	-0.38	-0.09	0.89	1.43	-1.56	0.51	0.25	*knife and fork*
	43	-0.40	-0.26	1.03	-0.25	2.02	-2.78	-0.85	0.62	1.75	-1.11	-0.02	*ladies and gentlemen*
	44	-0.08	1.13	2.18	2.03	1.97	0.30	-0.16	1.21	2.12	2.68	1.34	*salt and pepper*
	45	1.56	0.60	0.22	2.39	-0.31	1.07	1.14	-0.68	-1.71	0.69	0.50	*shoes and socks*
	46	-1.29	-0.51		-1.24	-2.14	-0.39	0.55	0.64	-2.32	-1.09	-0.87	*wine and cheese*
	47	-0.77	-0.70	2.37		-0.34	0.48	0.27	1.28	-0.43	-0.32	0.21	*gin and tonic*
	48	0.35	2.23		-0.86	1.61	-0.92	-0.45	2.32	1.01	0.16	0.61	*cats and dogs*

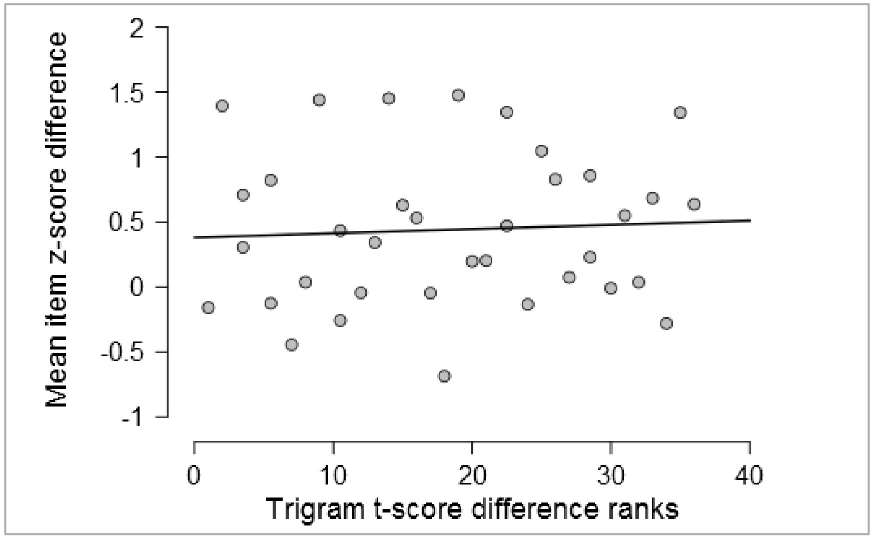
NB: Coloured numbers highlight positive z-score differences, i.e. shorter monitoring latencies in the stronger collocation compared to the weaker collocation.

Appendix B.5: *Pilot data: Correlation plots of a) raw trigram t-score difference and z-score difference and b) t-score difference ranks and z-score difference*

(a)



(b)



Appendix B.6: *Pilot data: Results of the cloze task*

	Target	Participant 1	2	3	4	5	6	7	8	9	10	Cloze probability
1	*aboard*	✓	✓	✓	onboard	✓	✓	✓	✓	✓	✓	**.9**
2	*in*	✓	✓	✓	✓	✓	here	✓	quickly	✓	✓	**.8**
3	*worry*	✓	✓	✓	panic	✓	✓	✓	✓	✓	✓	**.9**
4	*here*	there	here, there, I don't know	there	there	there	✓	✓	✓	there	✓	**.4**
5	*night*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
6	*yourself*	yourselves	yourselves	yourself	me	yourselves	me	yourself	yourself	yourself	me	**.7**
7	*one*	✓	✓	✓	dress	✓	✓	✓	✓	✓	✓	**.9**
8	*you*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
9	*know*	✓	✓	✓	care	✓	✓	✓	✓	✓	✓	**.9**
10	*sure*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
11	*sorry*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
12	*go*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
13	*you*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
14	*all*	✓	✓	✓	the moment [R] Tesco	✓	✓	✓	✓	✓	this time [R] I don't know	**.8**
15	*you*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
16	*alright*	fine	okay	okay	okay	okay	fine	✓	okay	✓	okay	**.2**
17	*way*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
18	*about*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
19	*happened*	✓	✓	✓	✓	✓	✓	✓	✓	✓	wrong	**.9**
20	*mind*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	**1**
Sum expected		18	18	18	12	18	17	20	18	19	16	
Percentage correct		90%	90%	90%	60%	90%	85%	100%	90%	95%	80%	

[R] = repetition of sentence; for item 6, both 'yourself' and 'yourselves' are accepted as the expected answer

Appendix B.7: *Example PALPA score sheet including scoring procedures*

PALPA

13

page 3 (of 3)

Auditory Digit Matching Span

Presenter's & Marking Form

Name: C17 Date: 3/11/17

Instructions to Presenter: Present each digit at a 1 second interval with flat intonation. Wait for 2 seconds, then present the same digits, either in identical order or with two adjacent digits reversed. Circle wrong responses in appropriate columns.

Start at left-most column. If subject is correct, move to top of next column and so on, till subject fails. At this point, move back to last correct span and present the next group of digits at that length. If the subject succeeds, move to next span length and so on. This is a stepwise method of assessing digit matching span. Span is the length at which the majority of correct same-different judgements is made.

Instructions to Subject: I'm going to say some numbers to you, one at a time. Then I'm going to say the same numbers again. Listen carefully. I want you to tell me whether I said them in the same order. Say "yes" if they were the same order, "no" if they were different.

2		3		4		5		6		7	
79	Y	467	N	1284	Y	38273	Y	156982	N	4236187	N
79		647		1284		38273		156928		4263187	
24	N	954	Y	5391	Y	27691	N	294356	N	9438756	Y
42		954		5391		72691		249356		9438756	
31	N	269	N	8297	N	98456	Y	947685	Y	2541398	Y
13		296		8927		98456		947685		2541398	
89	Y	325	N	6285	Y	68345	N	213847	Y	9862134	N
89		235		6285		68435		213847		8962134	
46	N	518	Y	8954	Y	98145	Y	859124	N	3142596	N
64		518		8954		98145		851924		3142956	
72	Y	635	N	4527	Y	62975	N	784916	N	2874512	Y
72		653		4527		69275		874916		2874512	
13	Y	298	Y	5691	N	76813	Y	368254	N	7286453	N
13		298		5619		76813		368524		7284653	
54	N	126	Y	9623	N	50962	N	613982	Y	1398724	Y
45		126		6923		50926		613982		1398724	
83	Y	987	N	1879	N	94127	Y	173859	Y	4392718	N
83		897		1789		94127		173859		4392781	
61	N	543	Y	2498	N	37514	N	875463	Y	5831467	Y
16		543		4298		35714		875463		5831467	
Span	7 1		1		1		1		3		8 + 2 mistakes

Appendix B.8: *Information Sheet: younger control participants*

LONDON'S GLOBAL UNIVERSITY

UCL

Information Sheet for Participants in Research Studies

You will be given a copy of this information sheet.

Title of Project: The recognition of words in sentences

Project ID No: LC/2013/05

Name, Address and Contact Details of Investigators:

Prof Rosemary Varley
UCL Division of Psychology and Language Sciences
Chandler House
2 Wakefield Street
London WC1N 1PF
rosemary.varley@ucl.ac.uk
Tel. 020 7679 4234

Claudia Heilemann
UCL Division of Psychology and Language Sciences
Chandler House
2 Wakefield Street
London WC1N 1PF
c.heilemann@ucl.ac.uk
Tel. 020 7679 4287

You are being invited to take part in a research study. Before you decide whether to take part, it is important to understand why the research is being done and what it will involve. Please take time to read the following information and discuss it with others if you wish. Ask us if anything is not clear or if you need more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
The main purpose of this study is to investigate the recognition of words in spoken sentences.

Do I have to take part?
No. It is up to you to decide whether or not to take part.
If you do decide to take part, you will be given this information sheet to keep. You will sign a consent form. If you decide to take part, you can stop the research at any time and without giving a reason.
Your decision on taking part will not influence your relationship with the UCL.

What will happen to me if I take part?
You will come to Chandler House at UCL and meet a researcher. There is a single session of 45-60 minutes. You will be able to take breaks within the session.

There are four tasks. All involve listening to and saying words. The words are neutral and will not be embarrassing or cause offence.
Task 1: Using a computer, you will read/listen to a word and then press a button when you hear that word in a sentence.
Task 2: You add a word to an incomplete sentence. This will be audio recorded.
Task 3: You listen to two words and say if they are the same or different.
Task 4: You describe a series of pictures in your own words. This will be audio recorded.

What are the possible disadvantages and risks of taking part?
The methods have been used for many years and are considered safe. You are free to stop your participation at any time and without having to give a reason. In addition, if you have any remaining questions, you may contact the experimenter at any time.

What if something goes wrong?
If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal UCL complaints mechanisms are available to you. In the first instance, complaints should be directed to Professor Rosemary Varley (Head of Department, UCL Department of Language & Cognition, Chandler House, 2 Wakefield Street, London WC1N 1PF).

Will my taking part in this study be kept confidential?
Yes. All information will be confidential. Your name and other personal details will not be revealed to any person outside the project. You will be given an anonymous identifier. All research information will be kept safely locked away.
All audio files will be stored on an encrypted hard drive in a locked filing cabinet at UCL. The audio files will be transcribed and only the transcriptions will be used for analysis. The audio files will only be accessed by the research team and no one will know who you are from the transcription.
We will destroy all identifiable information after 12 months and store only fully anonymised data in the longer term. However, you can choose to have the audio file deleted once it has been transcribed.

What will happen to the results of the research study?
The results of this study may be reported at conferences and in articles in academic journals. In any report of the study, we will ensure that you cannot be identified. If you would like a copy of any report, these can be obtained from Claudia Heilemann.

Who has reviewed the study?
This study has been reviewed by the UCL Division of Psychology & Language Sciences Research Ethics Committee.

Thank you for taking part in this study. You will be given a copy of the Information to keep.

Appendix B.9: *Consent form: younger control participants*

Consent form for participants in research studies

Title of Project: **The recognition of words in sentences**

Project ID No: LC/2013/05

Participant's Statement

I ………………………………………………………………………agree that I have

- read the information sheet and/or the project has been explained to me;
- had the opportunity to ask questions and discuss the study;
- received satisfactory answers to all my questions and my rights as a participant and whom to contact in the event of a research-related injury.

I understand that I am free to withdraw from the study without penalty.

I understand that I consent to the processing of my personal information for the purposes of this study only.

I understand that any such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

I give permission to be audio-recorded during parts of the session. The audio file will be stored securely by the research team. The audio recording will be destroyed when the analysis is complete.

However, if you would prefer the audio to be deleted immediately after transcription, please tick here. ☐

Signed: **Date:**

Investigator's Statement

I ………………………………………………………………………confirm that I have explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: **Date:**

Appendix B.10: *Information sheet: older control participants*

LONDON'S GLOBAL UNIVERSITY

UCL

Information Sheet for Participants in Research Studies

You will be given a copy of this information sheet.

Title of Project: The recognition of words in sentences

Project ID No: LC/2013/05

Name, Address and Contact Details of Investigators:

Prof Rosemary Varley
UCL Division of Psychology and Language Sciences
Chandler House
2 Wakefield Street
London WC1N 1PF
rosemary.varley@ucl.ac.uk
Tel. 020 7679 4234

Claudia Heilemann
UCL Division of Psychology and Language Sciences
Chandler House
2 Wakefield Street
London WC1N 1PF
c.heilemann@ucl.ac.uk
Tel. 020 7679 4287

You are being invited to take part in a research study. Before you decide whether to take part, it is important to understand why the research is being done and what it will involve. Please take time to read the following information and discuss it with others if you wish. Ask us if anything is not clear or if you need more information. Take time to decide whether or not you wish to take part.

What is the purpose of the study?
The main purpose of this study is to investigate the recognition of words in spoken sentences.

Do I have to take part?
No. It is up to you to decide whether or not to take part.
If you do decide to take part, you will be given this information sheet to keep. You will sign a consent form. If you decide to take part, you can stop the research at any time and without giving a reason.
Your decision on taking part will not influence your relationship with the UCL.

What will happen to me if I take part?
There are **two** sessions that take 45-60 and 45 minutes, respectively. There will be breaks during each session.
You will come to Chandler House at UCL and meet a researcher. The tasks involve listening to words and sentences. All the words and sentences are neutral and will not cause embarrassment. You will receive payment at a rate of £7.50 per hour after each session.

Session 1 (45-60 minutes):
There are three tasks.
Task 1: Using a computer, you will read/listen to a word and then press a button when you hear that word in a sentence.
Task 2: You add a word to an incomplete sentence. This will be audio recorded.
Task 3: You listen to two words and say if they are the same or different.

Session 2 (45 minutes):
You will listen to numbers, words and sentences, and select and describe pictures. Some of your picture descriptions will be audio recorded.

What are the possible disadvantages and risks of taking part?
The methods have been used for many years and are considered safe. You are free to stop your participation at any time and without having to give a reason. In addition, if you have any remaining questions you may contact the experimenter at any time.

What if something goes wrong?
If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal UCL complaints mechanisms are available to you. In the first instance, complaints should be directed to Professor Rosemary Varley (Head of Department, UCL Department of Language & Cognition, Chandler House, 2 Wakefield Street, London WC1N 1PF).

Will my taking part in this study be kept confidential?
Yes. All information will be confidential. Your name and other personal details will not be revealed to any person outside the project. You will be given an anonymous identifier. All research information will be kept safely locked away.
All audio files will be stored on an encrypted hard drive in a locked filing cabinet at UCL. The audio files will be transcribed and only the transcriptions will be used for analysis. The audio files will only be accessed by the research team and no one will know who you are from the transcription.
We will destroy all identifiable information after 12 months and store only fully anonymised data in the longer term. However, you can choose to have the audio file deleted once it has been transcribed.

What will happen to the results of the research study?
The results of this study may be reported at conferences and in articles in academic journals. In any report of the study, we will ensure that you cannot be identified. If you would like a copy of any report, these can be obtained from Claudia Heilemann.

Who has reviewed the study?
This study has been reviewed by the UCL Division of Psychology & Language Sciences Research Ethics Committee.

Thank you for taking part in this study. You will be given a copy of the Information to keep.

Appendix B.11: *Consent form: older control participants*

Consent form for participants in research studies

Title of Project: **The recognition of words in sentences**

Project ID No: LC/2013/05

Participant's Statement

I ……………………………………………………………………………agree that I have

- read the information sheet and/or the project has been explained to me;
- had the opportunity to ask questions and discuss the study;
- received satisfactory answers to all my questions and my rights as a participant and whom to contact in the event of a research-related injury.

I understand that I am receiving a (token) payment for participating in this research and that some of my personal details will be passed to UCL Finance for administrative purposes only.

I understand that I am free to withdraw from the study without penalty.

I understand that I consent to the processing of my personal information for the purposes of this study only.

I understand that any such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

I give permission to be audio-recorded during parts of the session. The audio file will be stored securely by the research team. The audio recording will be destroyed when the analysis is complete.

However, if you would prefer the audio to be deleted immediately after transcription, please tick here. ☐

Signed: **Date:**

Investigator's Statement

I ……………………………………………………………………………confirm that I have explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signed: **Date:**

Appendix B.12: *Aphasia-friendly information sheet*

LONDON'S GLOBAL UNIVERSITY

UCL

Information Sheet for Participants in Research Studies

Title of project: The recognition of words in sentences

(Project ID No: LC/2013/05)

We would like to invite you to take part in a research project.

Who we are

Prof Rosemary Varley

UCL Department of Language & Cognition

Chandler House

2 Wakefield Street

London WC1N 1PF

rosemary.varley@ucl.ac.uk

020 7679 4234

Claudia Heilemann

UCL Department of Language & Cognition

Chandler House

2 Wakefield Street

London WC1N 1PF

c.heilemann@ucl.ac.uk

020 7679 4287

What you need to know

This research does **not** include **therapy**.

Before you decide, you need to **understand**:

- **Why** we are doing the research.
- **What** you have to do.

Please read this information carefully.

Take your **time**.

Talk to your **family** and **friends** about the research.

Ask us **questions** if:

- You do **not understand** something.
- You need more **information**.

Why are we doing this research?

Some people find **sentences** hard after a **stroke**.

We want to find out about **sentence difficulties**.

What will I have to do?

You will have **two** meetings.

Meeting 1:

- You will **describe pictures**.

- Some of your descriptions will be **audio recorded**.

- You will **listen** to sentences and **point** at pictures.

- You will be asked if two **numbers** sound **same** or **different**.

Meeting 2:

You will be asked to do **three things**.

1) We use a **computer**.

You listen to **words** and **sentences**.

You push a **button** when you hear a certain word.

2) You **finish** a sentence that you hear. This will be **audio recorded**.

3) You will be asked if two **words** sound **same** or **different**.

Do I have to take part?

You can decide. You can say **no**. If you say **no**, this will **not** influence your **relationship** with UCL. If you say **no**, you will **still get your normal help**.	
If you decide to take part: You can **keep** this information sheet. A researcher will **invite** you to **Chandler House** at UCL. You will sign a **consent form** to show you're **happy** to take part. You will receive a **payment** of **£7.50** after each meeting as a contribution to travel and out-of-pocket expenses linked to attending the meeting.	  

The meetings

The meetings will take **60 - 70 minutes**.	
There will be **breaks** during the meetings.	
You can **stop** the research at **any time**, and without giving a reason.	

Who will see the results about me?

We will keep your **name**, **personal details** and **audio** recordings **safe, locked away**.

Only **researchers** will see the **information** about you.

Only **researchers** will access the **audio recordings.**

We will tell **other researchers** (in England and across the world) about the **results.**

When we **present** the results of this study, we will **take out** your **name.**

No one **outside the study** will know **who you are** from the results.

Is the research safe?

A **committee** decides if research can happen.

This is the UCL Division of Psychology & Language Sciences Research Ethics Committee.

They say that this research can happen.

They say that it is **safe.**

They say that it has been **planned properly.**

What if something goes wrong?

If you **take part** in the research

and **if you think** you were **harmed**

there are **people to talk to.**

You can **contact** Prof Rosemary Varley

Head of Department

Department of Language & Cognition

2 Wakefield Street

London WC1N 1PF

Tel. 020 7679 4234

Appendix B.13: *Aphasia-friendly consent form*

LONDON'S GLOBAL UNIVERSITY

UCL

Consent form for participants in research studies

Project ID No: LC/2013/05

Taking part in the research project:

The recognition of words in sentences

Please **mark** yes or no in **each statement**.

	I have read the information about the research, or the project has been **explained** to me.	yes ☐ no ☐
	I have had the chance to ask **questions** about the research.	yes ☐ no ☐
	I am **happy** with the **answers** to my questions.	yes ☐ no ☐
	I am happy that parts of the meeting will be **audio recorded** for the study.	yes ☐ no ☐
	I understand that **information** about me will be kept **safe**, and it will not be shared with anyone outside research.	yes ☐ no ☐

	I know that when **results** are shared, researchers will **not** use **my name**.	yes ☐ no ☐
	I **choose** to **share** my information with the [redacted]	yes ☐ no ☐
 	I understand that I can **stop** being in the research project at any time. If I stop, I don't have to give any reason.	yes ☐ no ☐
	I understand that I am receiving a **payment** as a contribution to travel and out-of-pocket expenses linked to attending the meeting, and that some of my **information** will be passed to **UCL Finance**.	yes ☐ no ☐
	I **agree** to take part in the research.	yes ☐ no ☐

Name: ______________________________

Signature: ____________________ **Date:** ____________________

Researcher's statement:

I ....confirm that I have explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signature: ______________________ **Date:** ____________________

Appendix B.14: *Average z-score differences by participant and condition for younger adults*

ID	**Average z-score difference**		
	First-word manipulations	**Second-word manipulations**	**Binomials**
YC1	0.36	0.32	0.52
YC3	-0.07	0.61	0.54
YC4	0.76	0.23	0.64
YC5	0.36	0.48	1.13
YC6	0.27	0.95	0.12
YC7	0.24	0.72	0.60
YC8	0.59	0.41	0.36
YC9	0.32	0.27	0.23
YC10	0.59	0.26	0.22
YC11	0.09	0.27	0.83
YC12	0.72	0.99	0.71
YC13	0.21	0.67	0.79
YC14	0.18	0.05	0.33
YC15	0.32	0.51	0.57
YC16	-0.21	0.77	1.01
YC17	0.24	0.16	0.87
YC18	0.11	0.69	-0.12
YC19	0.33	0.07	0.41
YC20	0.00	0.83	0.27
YC21	0.33	0.67	0.54
YC22	0.81	0.69	0.51
YC23	-0.21	0.75	0.57
AVERAGE	***0.29***	***0.52***	***0.53***
SD	***0.28***	***0.28***	***0.30***
Effect of the manipulation?	*yes*	*yes*	*yes*

Highlighted cells: Effect in opposite direction compared to the prediction

Appendix B.15: *Average z-score differences by participant and condition for older adults*

ID	Average z-score difference		
	First-word manipulations	**Second-word manipulations**	**Binomials**
OC1	0.08	0.76	0.58
OC2	0.39	0.96	1.23
OC3	0.20	0.78	1.21
OC4	0.27	1.17	0.22
OC5	0.45	0.36	1.45
OC6	0.32	0.61	1.14
OC7	0.32	0.59	0.47
OC8	0.19	0.75	0.72
OC10	0.06	0.62	0.57
OC12	-0.22	0.52	1.09
OC13	0.15	0.40	0.76
OC14	0.08	0.55	0.77
OC16	0.39	0.84	0.67
OC17	-0.16	0.52	1.46
OC18	0.22	0.93	0.54
OC19	0.18	0.38	1.11
OC20	0.60	0.19	0.19
OC21	0.43	0.74	0.39
OC22	-0.07	-0.48	1.14
OC23	0.62	0.76	0.42
OC24	0.19	0.47	0.57
OC25	0.49	0.20	1.20
OC26	0.26	0.83	1.47
AVERAGE	***0.24***	***0.58***	***0.84***
SD	***0.22***	***0.33***	***0.40***
Effect of the manipulation?	*yes*	*yes*	*yes*

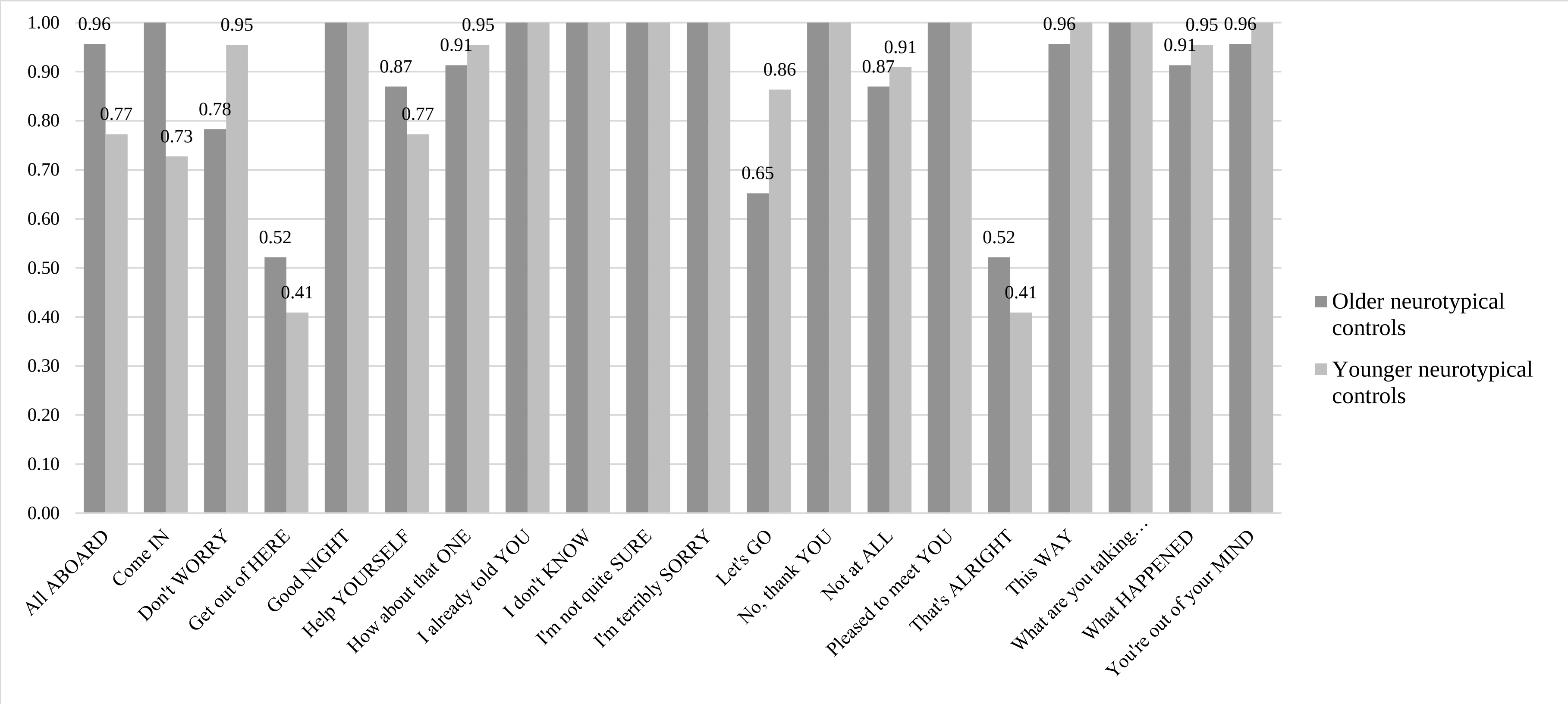
Highlighted cells: Effect in opposite direction compared to the prediction

Appendix B.16: *Average z-score differences by participant and condition for PWA*

ID	Average z-score difference		
	First-word manipulations	**Second-word manipulations**	**Binomials**
A1	0.42	0.85	-0.20
A2	-0.25	0.07	1.03
A3	0.14	0.54	0.44
A4	-0.22	0.23	0.29
A5	0.54	0.28	0.53
A6	0.30	0.79	0.14
A7	0.46	0.50	0.63
A8	0.29	0.55	1.01
A10	0.03	0.29	1.80
A11	0.12	0.43	0.66
A13	0.56	0.25	0.75
A14	0.33	0.56	1.26
A15	0.14	0.13	0.58
A16	-0.13	-0.32	0.44
A17	-0.20	0.68	0.09
A18	0.34	0.44	0.52
A19	0.06	-0.16	-0.11
A20	0.22	0.03	0.76
A21	0.30	-0.32	0.59
A23	0.35	0.49	0.27
A24	0.23	-0.19	-0.53
AVERAGE	***0.19***	***0.29***	***0.52***
SD	***0.24***	***0.34***	***0.52***
Effect of the manipulation?	*yes*	*yes*	*yes*

Highlighted cells: Effect in opposite direction compared to the prediction

Appendix B.17: *Cloze probabilities of 20 items across 22 younger and 23 older neurotypical control participants*



Item "***in***": the most common alternative answer was "*here*" (younger: 4, older: 0). Item "***worry***": the most common alternative answer was "*panic*" (older: 5, younger: 1). Item "***here***": the most common alternative answer was "*there*" (older: 9, younger: 12). Item "***go***": the most common alternative answer was "*hurry*" (older: 7, younger: 3). Item "***alright***": the most common alternative answer was "*okay*" (older: 11, younger: 11).

C.Study 3 Appendices

Appendix C.1: *Aphasia-friendly information sheet*

LONDON'S GLOBAL UNIVERSITY

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

Information Sheet for Participants in Research Studies

Title of project: From fixed to flexible: using multi-word utterances more creatively in aphasia (Project ID No: LCRD.2017.01)

We would like to **invite you and a family member or friend** to **take part in a research project**.

Who we are

Dr Suzanne Beeke

UCL Department of Language & Cognition
Chandler House
2 Wakefield Street
London WC1N 1PF

s.beeke@ucl.ac.uk

020 7679 4215

Claudia Heilemann

UCL Department of Language & Cognition
Chandler House
2 Wakefield Street
London WC1N 1PF

c.heilemann@ucl.ac.uk

020 7679 4287

What you need to know

This research **does include an intervention.**

Before you decide, you need to **understand**

- **Why** we are doing the research
- **What** you have to do

Please read this information carefully.

Take your **time** to think - do you want to take part in this project?

Talk to your **family** and **friends** about the research.

Ask us **questions** if

- You do **not understand** something
- You need more **information**

Why are we doing this research?

Some people find **talking in sentences** hard after a **stroke**.

We want to **help you using short sentences more flexibly**.

We want to **test out** a **new intervention.**

What will I have to do?

You will be involved in a **12-week programme**.

There will be **three phases**

12 x

Phase 1 (Week 1-3)

You do some **language and thinking tests**

You might **name or describe pictures**, and **point at pictures.**

We **ask you some questions** about your **interests.**

Parts of these meetings will be **audio recorded.**

Together with your conversation partner, you **video record 20 minutes of your conversations at home.** You can **decide when** to record and **what to talk about.**

Phase 2 (Week 4-9)

You **repeat some words** that we will **audio record.**

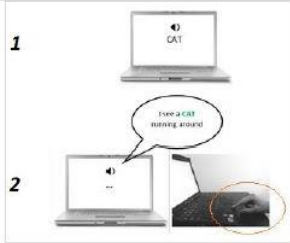
You **listen back to your recordings.**

We use a game on a **computer.**

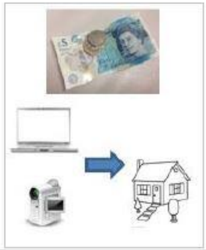
You listen to **words** and **sentences.**

You push a **button** when you hear a certain word.

You **take a computer home** to practice this game.



We can cover your **travel expenses** for **transporting** a **laptop computer** and **video camera** to your **home.**



You **practice** short **sentences** with a **computer.**

Phase 3 (Week 10-12)

We **repeat some of the tasks** from **Phase 1.**	 
You **video record one conversation per week** with your conversation partner **at home**. You **can decide when** to record and **what to talk about.**	
You **practice** short **sentences** with a **computer.**	
After 4 weeks, we will **repeat some tasks** again.	

Do I have to take part?

You can decide. You can say **no.** If you say **no,** this will **not** influence your **relationship** with UCL. If you say **no,** you will **still get your normal help.**	
If you decide to take part: You can **keep** this information sheet. A researcher will **invite** you to **Chandler House** at UCL. You will sign a **consent form** to show you're **happy** to take part.	  

The meetings

The meetings take **60 - 70 minutes.**	
You will have **2 meetings** each **week.**	
There will be **breaks** during each meeting.	
You can **stop** the research at **any time**, and without telling us why.	

Who will see the results about me?

We will keep your **name, personal details** and **audio and video** recordings **safe, locked away.** We will follow the rules of the Data Protection Act 1998. Only **researchers** will see the information about you. Only researchers will access the **audio** and **video recordings.**	
If you <u>prefer</u> to have your **audio recordings** **<u>destroyed</u> after analysis is complete**, you can let us know.	After analysis: 
The **videos** will also be kept on a computer in **CAVA (a human Communication Audio Visual Archive)** at the UCL Library, **either (1)** until the analysis is complete **or (2)** for as long as the library exists. **<u>You decide.</u>**	
The **video recordings** will be **accessible to other researchers via CAVA.** Your **real names and addresses** will **not be stored** in CAVA.	

The **recordings** may be used for **conferences, future research and teaching at UCL** by responsible staff members and research students.

Your face will **not** be **blanked out** on the videos because we need to see your **eyes** and **face** as you talk. This means people might **recognise** you, but it is very **unlikely**.

When we **present** the results of this study, we will **take out** your real **name**.

We will tell **other researchers** (in England and across the world) about the **results.**

No one **outside the study** will know **who you or your conversation partner are** from the results.

What might be difficult about taking part?

You may find it **tiring** or **distressing**.

It will **take up your time**.

The **intervention** might **not help** you. If it does **not help** you, things will **stay the same as they are now**.

What might be good about taking part?

You may **enjoy** taking part.

You may be **helped** by the intervention.

You will **help** us **learn more** about **what could be good about the intervention**.

Is the research safe?

A **committee** decides if research can happen.

This is the UCL Department of Language and Cognition Research Ethics Committee.

They say that this research can happen.

They say that it is **safe**.

They say that it has been **planned properly**.

What if something goes wrong?

If you **take part** in the research
and **if you think** you were **harmed**
there are **people to talk to.**

If you are **unhappy** with the project, **first** phone Project Leader, **Suzanne Beeke**.

Or you can **contact** Prof Rosemary Varley

Head of Department
Department of Language & Cognition
2 Wakefield Street
London WC1N 1PF
Tel. 020 7679 4234

Appendix C.2: *Information sheet for CPs*

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

INFORMATION SHEET FOR PARTICIPANTS IN RESEARCH STUDIES

You will be given a copy of this information sheet.

Project:
From fixed to flexible: using multi-word utterances more creatively in aphasia

This study has been approved by the UCL Language and Cognition Research Department Ethics Committee (ID number: LCRD.2017.01).

Researchers:
Claudia Heilemann (PhD student) and Dr Suzanne Beeke (Senior Lecturer, Project Supervisor)

Address:
Language & Cognition Research Department, University College London, Chandler House, 2 Wakefield Street, London, WC1N 1PF

Contact details:
✉ c.heilemann@ucl.ac.uk / ☎ 020 7679 4287
✉ s.beeke@ucl.ac.uk / ☎ 020 7679 4215

We would like to invite you to take part in a research study. Before you decide whether to take part, it is important to understand why the research is being done and what it will involve.

Our research project investigates whether people with aphasia benefit from a new intervention programme that focuses on using multi-word utterances (e.g. "*I like it*") more flexibly (e.g. "*You like coffee*"). The results of this study will inform the development of training packages for people with aphasia.

Choosing not to take part will not disadvantage you or your family member / friend with aphasia in any way. For example, it will not affect their access to any future speech and language therapy.

The following information is designed to help you decide whether to take part or not. You will be given the opportunity to take the information sheet away, have a think about it, discuss it with others and ask us for any further information if something is unclear or you have a question. Please use the contact details above if we can be of any help during this time.

1

What would the project involve?

The project runs from September 2017 to February 2019. Your family member/friend will be involved in a 12-week programme consisting of three phases. All meetings will take place at UCL in Chandler House (WC1N 1PF).

In the first phase (week 1-3), a range of assessments will be carried out by the student researcher whose background is in Speech and Language Therapy. The assessments will provide information about your family member's / friend's language and cognitive ('thinking') skills such as naming pictures or telling a story. Most of these assessments will be audio recorded. In addition, the research team will provide a video camera, to take to the home of your family member / friend. With this, you can record conversations between the two of you without a researcher present. The researcher will demonstrate how to use the video camera and answer any questions you may have. They will also provide a written instruction sheet. We would like you to record one recording per week, i.e. three recordings in total during Phase 1.

Both yourself and your family member / friend should be visible in the video. Each recording should last 20-30 minutes, and can be made up blocks of shorter conversations on different days. There are no restrictions on what you talk about, as comparing different interactions will add information to the study. But we do ask that the TV and radio are switched off when the video camera is running. The recordings will be analysed using a method that is informed by Conversation Analysis. We will focus on the communicative skills used by your family member / friend.

In the second phase (week 4-9), your family member / friend will take part in the actual intervention phase, for which we will use a laptop computer. The aim of the intervention is to help your family member / friend to use multi-word utterances more creatively, by practicing them in a structured way. We will also use your family member's / friend's own audio recordings of short sentences to practice certain utterances. One part of the intervention will involve additional practice at home (with a laptop computer provided by the research team).

In the third phase (week 10-12), we will repeat some of the assessments from Phase 1. We would like you to record one conversation per week, i.e. three recordings in total during Phase 3. Again, both yourself and your family member / friend should be visible in the video. Each recording should last 20-30 minutes, and can be made up blocks of shorter conversations on different days. There are no restrictions on what you talk about, as comparing different interactions will add information to the study. But we do ask that the TV and radio are switched off when the video camera is running. The recordings will be analysed using a method that is informed by Conversation Analysis. We will focus on the communicative skills used by your family member / friend.

After a break of four weeks, we will repeat some of the assessments and would like you to audio or video record one more conversation sample.

→ Travel expenses linked to transporting the video camera and laptop computer to your family member's / friend's home will be covered.

→ You do not have to be present at the meetings in Chandler House, as you would be involved in the conversation recordings only, at home.

Do I have to say 'yes'?

No. You can take your time to make your decision, and you can contact the researcher, Claudia Heilemann, if you have any questions. We suggest you talk to family and friends to help you decide.
If you decide to take part, you will sign a consent form so that we have a record of your decision. You can stop participating at any time without telling us why.

What if I change my mind?

You are completely free to withdraw your participation from the study at any time, without having to provide any reason for doing so. Choosing not to participate will not incur any penalty or loss of benefits to which you or your family member/friend are otherwise entitled.

If you stop taking part, we will keep the videos that you have already made.

How will my data be looked after?

All information will be confidential. All data will be collected, used and stored in accordance with the Data Protection Act 1998. It will be stored securely on an encrypted hard drive in a locked cabinet. With your permission, the video recordings will be securely stored in an online archive, CAVA (a human Communication Audio Visual Archive) at the University College London (UCL) Library for as long as the Library exists. They will be accessible to bona fide researchers via CAVA, who have agreed to follow the same confidentiality procedures. People who want to use the videos will sign a contract to respect your confidentiality, rights and dignity. They will only be able to use them for research that is sponsored by their employer and follows the Data Protection Act 1998.

The videos will be stored in CAVA at UCL Library, either (1) until the analysis is complete, or (2) for as long as the library exists.
You and your family member / friend with aphasia decide how long we can keep your videos.

All video recordings, assessment data, papers and electronic documents will be anonymous; neither you nor your family member / friend will be referred to by your real names.

If you choose to withdraw from the project before the end of therapy, or can no longer consent to take part, the videos you have already made will be kept in the CAVA archive. Your real names and addresses will not be stored in the CAVA archive.

With your permission, the video recordings may be used for future research and teaching purposes at UCL by responsible staff members and research students. Clips of your videos will be watched by students and other researchers when the research team present the results at conferences.

It is possible that a person who sees your video might recognise you, if they know you. The chance of this happening is actually very small. We do not blank out your face on the video, because we need to study your eyes and facial expression. In our extensive experience of research, only one person has ever been recognised at a conference from their video.

You will sign a consent form to show that you agree with how we will keep and use your personal details and videos.

What will happen to the results of the research study?

The results of this study may be reported at conferences and in articles in academic journals. In any report of the study, we will ensure that you cannot be identified. If you would like a copy of any report, these can be obtained from Claudia Heilemann.

Who has checked that this project is okay?

All research at UCL is looked at by an independent group of people, called a Research Ethics Committee. They protect your safety, rights, wellbeing and dignity. This study has been reviewed and approved by the UCL Language and Cognition Research Department Ethics Committee.

What if something goes wrong?

If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal UCL complaints mechanisms are available to you. In the first instance, complaints should be directed to Professor Rosemary Varley (Head of Department, UCL Department of Language & Cognition, Chandler House, 2 Wakefield Street, London WC1N 1PF).

What happens next?

We would like to give you the opportunity to have a think about the information and decide whether you would be interested in participating or not. Please use the contact details on page 1 if we can be of any help during this time. If you do decide to take part you will be asked to sign a consent form. Following this, we will arrange the most convenient days and times to meet your family member / friend at Chandler House.

Appendix C.3: *Aphasia-friendly consent form*

LONDON'S GLOBAL UNIVERSITY

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

UCL

Consent form for participants in research studies

Taking part in the research project:

From fixed to flexible: using multi-word utterances more creatively in aphasia

This project has been **approved** by the UCL Language and Cognition Research Department **Ethics** Committee (ID number: LCRD.2017.01).

Please **mark** **yes** ☐ or **no** ☐ in **each statement**.

	I have read the information about the research or the project has been **explained** to me.	**yes** ☐ **no** ☐
	I have had the chance to ask **questions** about the research.	**yes** ☐ **no** ☐
	I am **happy** with the **answers** to my questions.	**yes** ☐ **no** ☐
	I am happy that parts of the meetings will be **audio recorded** for the study.	**yes** ☐ **no** ☐
After analysis: 	Please **tick** this box if you prefer to have your audio data **destroyed after analysis**. ☐	

	I agree to be **videotaped** during this project.	yes no
	I agree that short bits of my **videos** can be used to **teach students** about aphasia at UCL or can be watched by other **researchers** at **conferences**.	yes no
	I know that **my face** will **not** be **blanked out** on the videos because researchers need to see my **eyes** and **facial expressions**. I know this means people might **recognise** me, but I understand that this is very **unlikely**.	yes no
  	I understand that the **videos** will be **stored safely at the UCL Library**. They will be stored in **CAVA** (a human Communication Audio Visual Archive) **EITHER a)** <u>for as long as the Library exists</u>, for future aphasia research. I understand that **other researchers can look at the videos for their studies**. Other researchers will sign a **contract** to **respect** my confidentiality, rights and dignity, and use videos in a responsible way.	yes no yes no

After analysis: 	**OR b)** <u>until the analysis is complete</u>. Then your **video data will be destroyed** after the analysis is finished.	yes ☐ no ☐
	I understand that **information** about me will be kept **safe**, and it will **not** be **shared** with **anyone outside research**.	yes ☐ no ☐
	I know that when **results** are shared, researchers will **not** use **my name**.	yes ☐ no ☐
	I **choose** to **share** my information with the [redacted]	yes ☐ no ☐
  	I understand that I can **stop** being in the research project at any time. If I stop, I **don't have to give any reason**. I know that if I choose to stop before the end, my **videos** will be kept in the CAVA Archive for **as long as I said above**.	yes ☐ no ☐
 	I understand that I am receiving a **payment** to cover the **travel** expenses linked to **transporting a laptop computer and video camera** to my home and back to UCL. Some of my **information** will be passed to **UCL Finance**.	yes ☐ no ☐

	I **agree** to take part in the research.	yes ☐ no ☐

Name: ______________________________

Signature: ____________________ **Date:** ____________________

Researcher's statement:

I ……………………………………………………………………..confirm that I have explained the purpose of the study to the participant and outlined any reasonably foreseeable risks or benefits (where applicable).

Signature: ________________________ **Date:** ____________________

Appendix C.4: *Consent form for CPs*

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

UCL

INFORMED CONSENT FORM FOR PARTICIPANTS IN RESEARCH STUDIES

Project: From fixed to flexible: using multi-word utterances more creatively in aphasia

This study has been approved by the UCL Language and Cognition Research Department Ethics Committee (ID number: LCRD.2017.01).

Researchers:
Claudia Heilemann (PhD student) and Dr Suzanne Beeke (Senior Lecturer, Project Supervisor)

Address:
Language & Cognition Research Department, University College London, Chandler House, 2 Wakefield Street, London, WC1N 1PF

Contact details:
✉ c.heilemann@ucl.ac.uk / ☎ 020 7679 4287
✉ s.beeke@ucl.ac.uk / ☎ 020 7679 4215

Participant's Statement:

I, ..., agree that I have read the information sheet about the study and/or the project has been explained to me orally, and that I have had the opportunity to ask questions and discuss the project with the researchers. I have either received satisfactory answers to all my questions, or have been advised of an individual to contact for answers to pertinent questions about the research. I have also been advised of my rights as a participant and whom to contact in the event of a research-related injury.

I understand that my participation will be video recorded. I am also aware of the use of these recordings for future research and teaching following the end of the project. I know that I will not be referred to by my real name in any video files, papers or electronic documents. I understand that the video recordings and results of the study will be stored securely on an encrypted hard drive in a locked cabinet and in an online repository, CAVA (a human Communication Audio Visual Archive) at University College London for as long as I specify below.

I consent to the processing of my personal information for the purposes of this study only, and not for any other purpose. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

Finally, I understand that I am free to withdraw from the study without any penalty, at any point.

Signed: .... Date: ....

Investigator's Statement:

I, ..., agree that I have carefully explained the purpose of the study to the participant.

Signed: .... Date: ....

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

UCL

CONSENT TO BE VIDEOED, AND FOR STORAGE AND FUTURE USE OF VIDEO AND INTERVIEW DATA

Please initial box

1. I agree to be videotaped during this project. ☐

2. I know that my videos will be stored in an online archive, CAVA (a human Communication Audio Visual Archive) at University College London:

Initial one box only to show how long you want videos to be kept for

EITHER 2a) for as long as the Library exists, for future aphasia research.
I know that the videos will be accessible to bona fide researchers via CAVA, who will sign a contract to respect my confidentiality, rights and dignity, and use videos in a responsible way. ☐

OR 2b) until the analysis is complete.
Then they will be destroyed. ☐

3. I know that if I choose to stop therapy before the end, my videos will be kept in the UCL CAVA Archive for the time indicated by me at point 2, above. ☐

4. I know that if I lose the ability to consent to take part, my involvement in this study will stop and my videos will be kept in the UCL Archive of Human Communication for the time indicated by me at point 2, above. ☐

5. I know that my face will not be blanked out on the videos because researchers need to see my eyes and facial expressions. I know this means people might recognise me, but I understand that this is very unlikely. ☐

6. I agree that short bits of video can be watched by students and researchers at conferences and for teaching purposes. ☐

7. I know that when the project team writes articles, talks about the intervention and labels my videotapes, they will not use my real name. ☐

Name of conversation partner	Date	Signature

Name of person taking consent	Date	Signature

Appendix C.5: *Frequency characteristics of the 12 source constructions*

Source construction	Raw frequency (spoken BNC)	Frequency per million words (spoken BNC)	t-score (as determined by FLAT Version 2)
I made it	35	4	5
I want that	58	6	6
I said it	222	22	15
I went to	638	64	25
I had it	106	11	8
I want to	1285	129	36
it's alright	766	77	28
I like it	164	17	11
I hate it	72	7	9
Where is it	166	17	13
When is it	43	4	5
don't know	9034	907	95
Average	**1049.1**	**105.3**	**21.3**
SD	**2543.8**	**255.3**	**25.2**

Appendix C.6: *Subset of A.D.A. Synonym Matching Task*

No	No in original version	Word pair		HI	LI	HF	LF
1	2	listen	attend		S		
2	3	rug	stomach	D			
3	5	gem	jewel				S
4	6	idle	heartless				D
5	8	lovely	nice			S	
6	15	smell	scent		S		
7	17	astute	stop		D		
8	21	dirt	filth				S
9	37	boulder	rock	S			
10	41	magician	wizard				S
11	42	gloom	deceit		D		
12	44	mail	post	S			
13	45	biscuit	crocodile				D
14	47	command	order		S		
15	54	jail	spook				D
16	56	stupid	genuine		D		
17	72	road	lane			S	
18	75	clumsy	frail		D		
19	78	stream	coat			D	
20	80	forbid	prohibit				S
21	87	courteous	gallant		S		
22	93	fraud	charlatan		S		
23	94	gorilla	ape				S
24	95	misery	orderly		D		
25	102	sever	lie		D		
26	105	elegant	graceful				S
27	111	robbery	cookie				D
28	115	couch	sofa				S
29	119	lazy	thrifty				D
30	121	rule	law		S		
31	122	film	boat	D			
32	123	dungeon	spade				D
33	129	yell	glutton				D
34	132	germ	bacteria				S
35	133	lack	deficiency		S		
36	136	clever	weak		D		
37	141	thick	authentic		D		
38	145	labour	toil		S		
39	149	lunch	river			D	
40	151	slap	ghost				D

HI / LI = high - / low-imageability pairs; HF / LF = high- / low frequency pairs;

S / D = similar / different meaning

Appendix C.7: *Acceptability questionnaire for participants with NFA*

LONDON'S GLOBAL UNIVERSITY

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

Project: From fixed to flexible: using multi-word utterances more creatively in aphasia
(Project ID No: LCRD.2017.01)

Feedback form after therapy

for participants

Your views are very important to us as they will help to develop future therapy programmes.

Your comments will be **stored safely** and **securely** as outlined in the consent forms.

Thank you for your time.

1

Feedback form for participants

Were **any parts** of the therapy **helpful** for you?

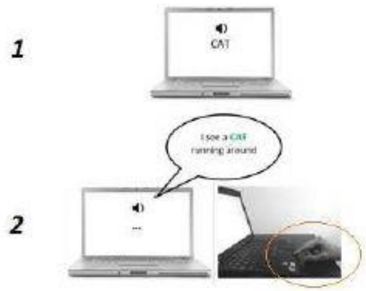
Please circle:

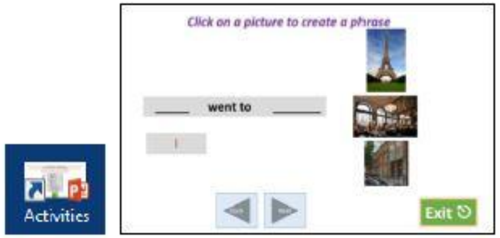
Yes **No**

2

Feedback form for participants

What did you find **helpful**?  Tick: ☐✓

  Making **recordings** of my own voice and **listening** to them ☐	 Playing the **game**  on the **laptop** ☐	 **Instructions** for the laptop ☐
Why?	**Why?**	**Why?**

		
Practicing speaking with **Claudia** ☐	**3 steps** to practice speaking ☐	Practicing speaking **at home** ☐
Why?	**Why?**	**Why?**

Were **any other parts helpful** for you? **Why**?

Overall, **how helpful** did you find this therapy?

Please circle:

1	2	3	4	5
Not at all helpful	**Not very** helpful	**Neutral**	**A little** helpful	**Very** helpful

Were **any parts** of the therapy **not helpful** for you? 

Please circle:

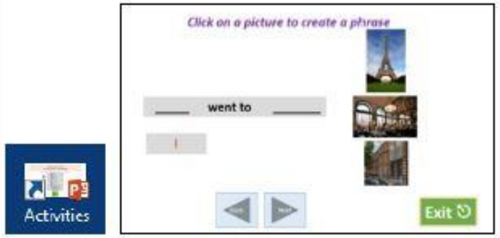
Yes **No**

What was **not helpful**?  Tick: ✓

  Making **recordings** of my own voice and **listening** to them ☐	1 2   Playing the **game** on the **laptop** ☐	 **Instructions** for the laptop ☐
Why?	**Why?**	**Why?**

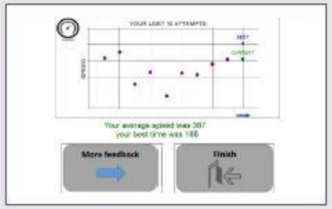
<table>
<tr>
<td>

Practicing speaking with Claudia ☐</td>
<td>

3 steps to practice speaking ☐</td>
<td>

Practicing speaking at home ☐</td>
</tr>
<tr>
<td>Why?</td>
<td>Why?</td>
<td>Why?</td>
</tr>
<tr>
<td colspan="3">Were there any other parts that you did not find helpful? Why?</td>
</tr>
</table>

How **easy** did you find the **game** that you practiced at home?



Please circle: 

1	2	3	5
Not at all easy	**Not very** easy	**Easy**	**Very** easy

How **easy** did you find practicing the **Activities** at home?

Please circle: 

1	2	3	5
Not at all easy	**Not very** easy	**Easy**	**Very** easy

Was the **therapy** what you **expected**?

Please circle: 

Yes **No**

Why?

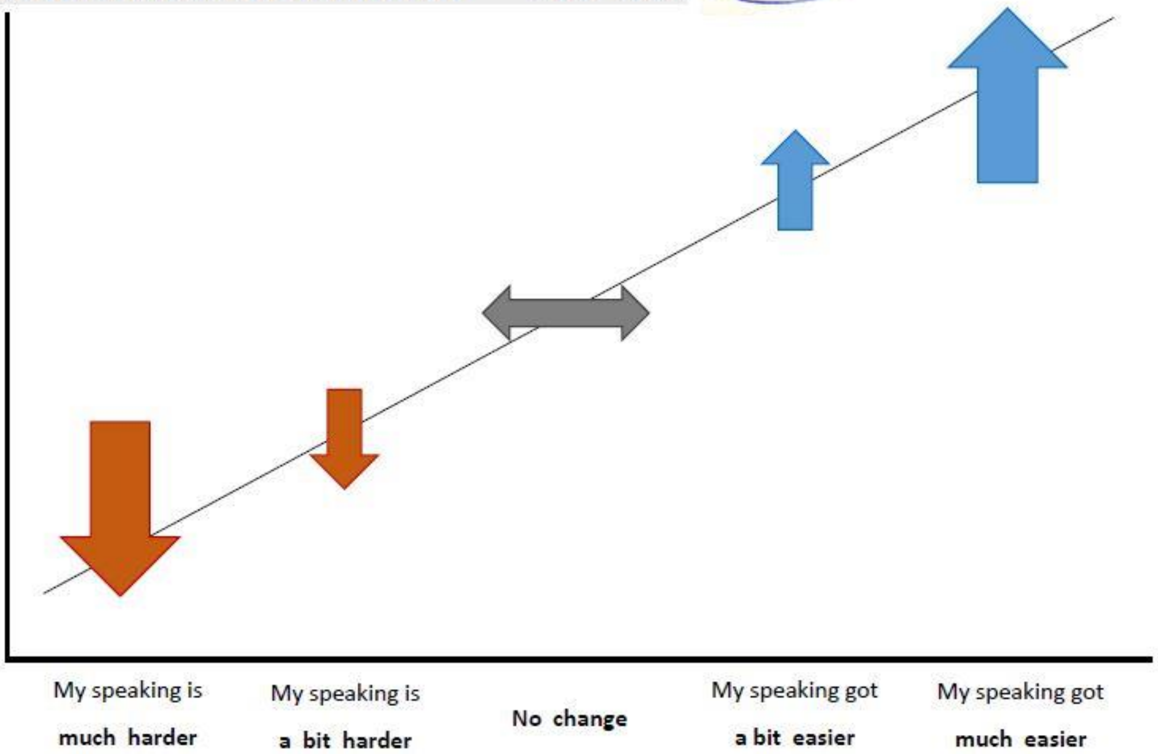
Do you think **anything else** should be **included** in the **therapy**? 

These days, **how often** do you **use the sentences, or parts of the sentences** practiced in **therapy?**

Please circle:

				
Never	**Rarely**	A few times a **month**	A few times a **week**	**Every day**

Did your speaking change after therapy? Please circle:



14

Feedback form for participants

Appendix C.8: *Acceptability questionnaire for CPs*

LONDON'S GLOBAL UNIVERSITY

UCL LANGUAGE AND COGNITION
DIVISION OF PSYCHOLOGY AND LANGUAGE SCIENCES

Project: From fixed to flexible: using multi-word utterances more creatively in aphasia

(Project ID No: LCRD.2017.01)

Feedback form after therapy

for Conversation Partners

Your views are very important to us as they will help to develop future therapy programmes.

Your comments will be stored safely and securely as outlined in the consent forms.

Thank you for your time.

Page 1 of 4

Conversation partner feedback form

1) Overall, **how helpful** do you think this therapy was for your friend / family member with aphasia? *(Please tick)*

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Not at all helpful	Not very helpful	Neutral	A little helpful	Very helpful

2) Were any parts of the therapy particularly **helpful** for your friend / family member with aphasia? *(Please tick)*

Yes ☐ No ☐ Not sure / can't comment ☐

What? Why?

----

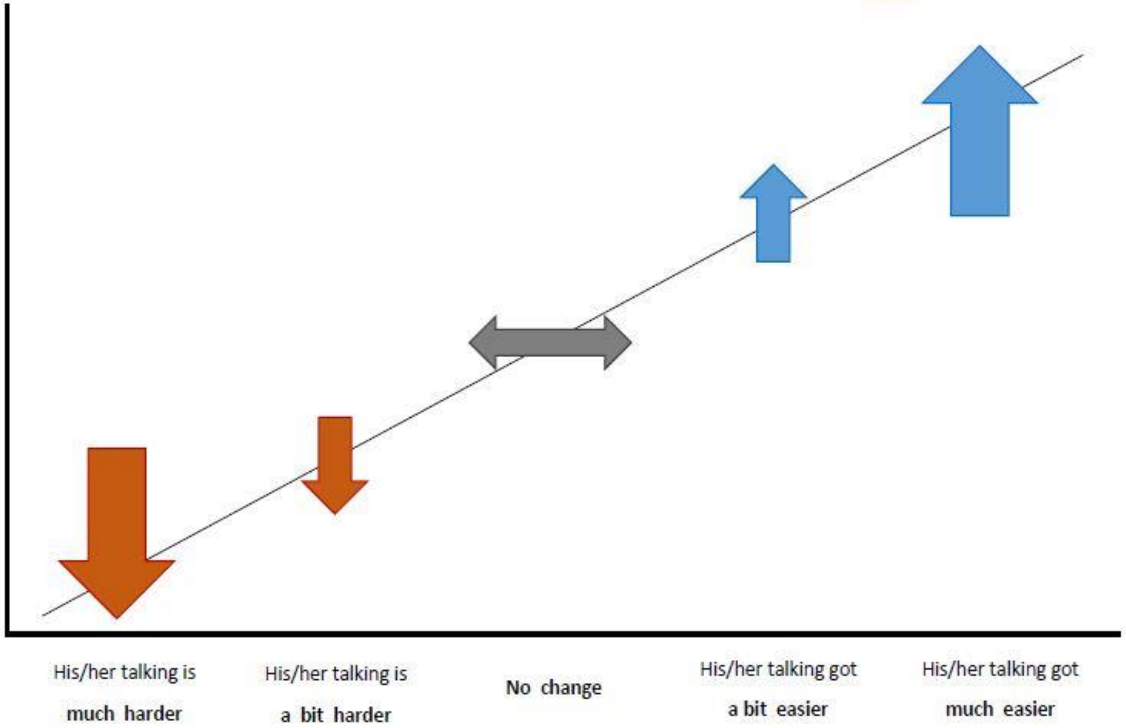
3) Were any parts of the therapy **not helpful** for your friend / family member with aphasia? *(Please tick)*

Yes ☐ No ☐ Not sure / can't comment ☐

What? Why?

----

4) Did your friend's / family member's **talking** change after therapy? *Please circle:*



5) Did **anything else** about your friend's / family member's **communication** (for example writing, reading or understanding what others say) change after therapy? *(Please tick)*

Yes ☐ No ☐ Not sure / can't comment ☐

What? Why?

---

6) Do you have **any other comments** about our therapy study?

Please post your feedback forms to:

Dr Suzanne Beeke, UCL, Department of Language & Cognition,

Chandler House, 2 Wakefield Street, London WC1N 1PF. **Thank you.**