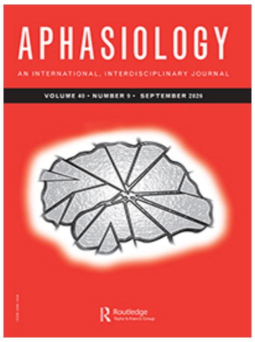


Routledge
Taylor & Francis Group

Aphasiology

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: www.tandfonline.com/journals/paph20

Measuring proximal outcomes of sentence production therapy: probing treated and untreated constructions with a story completion test

Claudia Bruns, Victoria Fleming, Fern Rodgers, Kerry Dathan, Jane Warren, Amir-Homayoun Javadi & Rosemary Varley

To cite this article: Claudia Bruns, Victoria Fleming, Fern Rodgers, Kerry Dathan, Jane Warren, Amir-Homayoun Javadi & Rosemary Varley (29 Jul 2026): Measuring proximal outcomes of sentence production therapy: probing treated and untreated constructions with a story completion test, Aphasiology, DOI: 10.1080/02687038.2026.2706852

To link to this article: https://doi.org/10.1080/02687038.2026.2706852

APHASIOLOGY
https://doi.org/10.1080/02687038.2026.2706852

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Measuring proximal outcomes of sentence production therapy: probing treated and untreated constructions with a story completion test

Claudia Bruns[a], Victoria Fleming[a], Fern Rodgers[a], Kerry Dathan[a], Jane Warren[a], Amir-Homayoun Javadi[a,b,c] and Rosemary Varley[a]

[a]Language & Cognition, Psychology & Language Sciences, University College London, London, UK; [b]School of Psychology, University of Kent, Canterbury, UK; [c]School of Rehabilitation, Tehran University of Medical Sciences, Tehran, Iran

ABSTRACT

Background: Many of the methods devised to assess aphasic sentence production involve spoken narratives (e.g. retelling a story, explaining a procedure) with high functional validity. However, when used to measure change following therapy, such unconstrained production tasks may fail to elicit the structures targeted by intervention. An alternative approach in targeting production of treated constructions is a story completion test (SCT), where the speaker completes a story stem, designed to elicit a specific sentence. In this study, we used a SCT to explore the outcomes of a new sentence therapy for aphasia (Unification Therapy Integrating LexIcon and Sentences – UTILISE), an intervention motivated by usage-based Construction Grammar. This approach allows creation of sets of frequency-matched treated and untreated sentences, permitting an exploration of generalisation, analogous to probing treated vs untreated pairs in lexical therapy studies.

Aims: To determine (1) whether speakers with aphasia produce more treated constructions and (2) more grammatically well-formed responses in a SCT following UTILISE therapy; and (3) whether therapy effects generalise to untreated sentences. In addition, we explored (4) whether baseline SCT performance was associated with behavioural profiling measures, and whether aphasia severity could predict therapy response on the SCT.

Methods & Procedures: Participants with chronic aphasia ($N = 39$) were recruited to a two-arm randomised control trial, with 31 completing the SCT at both baseline and outcome timepoints. Participants heard a story stem and attempted to produce a final sentence to complete the story, with stories probing treated and matched untreated sentences. UTILISE therapy involved two auditory tasks and one spoken sentence production task, all of which

ARTICLE HISTORY
Received 4 February 2026
Accepted 15 July 2026

KEYWORDS
Aphasia; therapy; assessment; story completion; sentence production

CONTACT Claudia Bruns c.bruns@ucl.ac.uk Language & Cognition, Psychology & Language Sciences, University College London, 2 Wakefield Street, Chandler House, London WC1N 1PF, UK

This article has been corrected with minor changes. These changes do not impact the academic content of the article.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/02687038.2026.2706852

targeted 20 high-frequency constructions. Therapy was delivered across 12 sessions over a four-week period. Audio-recordings of SCT responses were rated by masked scorers on two scales: whether (a) the target (expected) construction and (b) a grammatically well-formed response were produced.
Outcomes & Results: Participants showed significant increases in both production of target constructions and grammatically well-formed responses. In both cases, improvement generalised to untreated sentences. A significant association between sentence comprehension and SCT expected structure scores was found, and regression analysis indicated that severity may be a predictor of SCT change.
Conclusions: Applying usage-based principles to a sentence-level intervention facilitated the creation of a SCT with treated and untreated sentences to explore generalisation effects. Results revealed change in sentence production following UTILISE therapy, with generalisation to matched constructions of similar structure, length and usage frequency.

Introduction

Difficulties in sentence processing are pervasive in aphasia. These impairments are striking in the spoken output of those with non-fluent forms of aphasia that features reduced sentence length, simplified structures, and errors and omissions, particularly affecting verbs and function words. However, disruption of sentence processing is not restricted to non-fluent aphasias. Impairment of sentence comprehension is common across the aphasia subtypes (Bastiaanse & Edwards, 2004). Although production appears less impaired in fluent forms of aphasia, when comparisons are made to normative performance, there are signs of reduced sentence complexity (e.g., Zimmerer et al., 2018).

A range of interventions have been developed for sentence processing impairments in aphasia. They all face a common challenge of showing that post-treatment the person with aphasia (PwA) has more resources to support communication in everyday settings and made gains in functional communication. In the case of sentence production therapy, outcomes are often measured on narrative tasks, for example, through retelling a well-known story (e.g., of Cinderella; Saffran et al., 1989), or personal narratives, such as an individual's stroke story (Forbes et al., 2012). Such discourse-level tasks may tap critical outcomes but represent remote or distal measures of language change. If a therapy is directed at a particular grammatical structure, such as passives or negation, these constructions may not appear in an unconstrained narrative task or appear too rarely to evaluate meaningful change. There is a similar problem in determining the hierarchy of difficulty for grammatical constructions for aphasic speakers. Output might consist of canonical sentence structures (in English, a structure such as an active transitive sentence), or fragments of those structures, while other constructions remain un-surveyed. Khoshhal Mollasaraei et al. (2025) discuss further disadvantages of unconstrained elicitation tasks, including tense bias and avoidance of certain lexical items.

In response, researchers have developed tasks that constrain production options. Fink and colleagues (Fink et al., 1995) and Cho-Reyes and Thompson (2012) describe tasks where probe questions are presented that constrain responses (e.g., *What happened to X?*

to elicit a passive) and/or model the desired structure. Caplan and Hanna (1998) presented an event picture in which the key protagonists and objects were marked by arrows. One arrow was highlighted and the participant was required to begin a sentence with this component (e.g., in a passive construction the highlight would be placed over the image corresponding to the theme). Wilshire et al. (2014) report a task using event pictures without such visual cues to elicit a range of sentences, including rarer structures such as passive, ditransitive and embedded constructions. A difficulty with such constrained tasks is a reliance on extra-linguistic cues (such as arrows to mark protagonist roles), as well as dependence on visual processing in addition to linguistic demands. As a result, failure on these tasks might result from additional cognitive impairments rather than language disruption (Khoshhal Mollasaraei et al., 2025).

Another approach to constraining sentence production is the story completion test (SCT). This task was first described by Goodglass et al. (1972) as a method to survey the range of constructions available to a speaker with agrammatic aphasia (Barde et al., 2006; Gleason et al., 1975; Zatorski & Lesser, 1981). Listeners hear a short story stem which they attempt to complete by producing a single sentence (e.g., to elicit a declarative transitive construction: *Dogs always chase cats. A dog is in the street. A cat comes along. What happens?)*. Helm-Estabrooks and Ramsberger (1986) adapted the SCT method to form a two-component syntax stimulation therapy. The first stage involved delayed repetition: the therapist modelled the target sentence within a short story stem and this was then repeated by the PwA. The second stage employed the same story stems, but this time without therapist modelling the target sentence.

These constrained tasks are useful in probing target constructions but involve different linguistic and cognitive components from those required to produce everyday narratives. For example, a SCT places demands on auditory comprehension and retention of the story stem. Despite these additional elements, Rochon et al. (2005) observed more signs of improvement using constrained and proximal measures following a mapping therapy than on a more distal/unconstrained Cinderella narrative task. More proximal measures may play a useful role in measuring therapy outcomes by enabling detection of early-stage changes that might indicate system reorganisation and support for continuing with a particular therapeutic strategy. The SCT methodology is also employed in functional communication tests such as the Amsterdam-Nijmegen Everyday Language Test (ANELT, Blomert et al., 1995) to probe the combination of comprehension and production demands. Participants listen to descriptions of everyday scenarios and produce utterance(s) describing what they would say in that situation. In a recent study by Bailey and colleagues (2024) the story completion method was used to evaluate an adapted version of Verb Network Strengthening Treatment (VNeST, Edmonds et al., 2009); however, the task aimed to elicit a single verb rather than a whole phrase or sentence.

In experimental studies employing a SCT, the task is often used to elicit grammatical structures at an abstract level, for example, active, passive or ditransitive constructions that are not lexically specific. A newer approach to grammatical behaviour – usage-based Construction Grammar (CxG) – proposes that abstract constructions, such as an active declarative sentence [*X VERBS Y*], or a ditransitive construction [*X TRANSFER VERB Y to Z*], are just one level of grammatical representation (Bybee, 2010; Goldberg, 1995). In addition to abstract representations, other constructions with varying degrees of abstraction and size are stored in long-term memory. A construction consists of a pairing of linguistic form and semantic-pragmatic meaning/use – a familiar notion regarding lexical knowledge, but these pairings can be larger and more abstract than a single word. Large and fully concrete constructions can include multiword idioms (e.g., *the long and the short of it*) but also strings that are of high usage frequency such as *I don't know*. Partially specified strings may also be stored, such as *Where's the [X]*, with the empty slot being filled by a variety of other constructions (e.g., *Where's the key; where's the man who lost his key*) (Diessel, 2015). High-frequency strings are often retained in stroke aphasia (Bruns et al., 2019; Gahl et al., 2003; Zimmerer et al., 2018), as well as in Alzheimer's dementia and primary progressive aphasias (Wray, 2002; Zimmerer et al., 2020).

Bruns et al. (2025) employed principles from usage-based CxG to design a new therapy for aphasic sentence processing impairments (Unification Therapy Integrating LexIcon and Sentences – UTILISE). The therapy primed 20 high-frequency constructions in two listening tasks and then supported their production in a speaking task. After stimulation of the base form, new lexical items were inserted into the agent and theme slots around the verb, or an adjunct was added (e.g., from *I made it* to *He made coffee yesterday*). They report outcomes in sentence comprehension (Test for Reception of Grammar, TROG-2, Bishop, 2003) and narrative production (personal and retelling cartoon narratives), both of which represent distal outcomes as these tasks did not necessarily probe the constructions that appeared in therapy. Modest yet significant improvement was observed in connected speech.

The therapy differs from other sentence interventions in that it begins from the "bottom-up", by focusing on high frequency, lexically specific strings and then progressing to more abstract constructions. Key differences to other therapies include no requirement for metalinguistic awareness, e.g., explicit modelling of syntactic movement as in Treatment of Underlying Forms (Thompson et al., 2010), or of the mapping between syntactic and thematic structure (Rochon et al., 2005). The intervention focuses on semantically light, low-imageability and therefore versatile verbs (e.g., *want, like, make*) compared with established programmes such as VNeST (Edmonds et al., 2009). Furthermore, focus on high-frequency items not only results in targeting functional frames, but also forms that are more likely to be resilient in the face of aphasic disruption and thus easier to access. The intervention also differs in delivery, including extensive auditory stimulation prior to production.

An advantage of taking a usage-based approach to sentence therapy is that it allows selection of sentence pairs matched on key variables such as frequency of use, structure and length. For example, a treated Subject-Verb-Object construction *I had it* can be matched in length, structure and frequency to the untreated (control) sentence *I found it*. Frequency values (t-scores[1]) for this pair are 8.12 versus 7.72 respectively, extracted from the spoken sub-corpus of the British National Corpus (BNC Consortium [BNC], 2007). In this way, the generalisation of therapy effects from treated to untreated items can be explored in a manner similar to matched treated/untreated pairs in lexical studies. Typically, in word-level therapies, item-specific effects are observed, with gains often limited to treated items (e.g., Palmer et al., 2019). In sentence therapies, generalisation might be assessed on untreated structures, but such assessments sometimes explore items that are not matched on critical variables. Similarly, more distal tasks such as picture descriptions or narratives can be used, but with the risk that targeted constructions may not appear, meaning potential positive therapy effects may be missed.

This study reports outcomes on a SCT for treated and untreated structures following a sentence processing intervention (UTILISE). The SCT was a secondary outcome measure in this study. Results on the primary outcomes of spoken sentence comprehension and connectivity of output are reported in Bruns et al. (2025), which provides detail on the therapy protocol (see also Bruns et al., 2025, supplemental materials), but for convenience a TiDiER summary (Hoffmann et al., 2014) is provided in Appendix 1. The study employed a randomised control trial (RCT) design, with participants randomly assigned to either immediate or deferred trial entry. There were two baseline assessments followed by a four-week intervention phase. Participants were re-assessed immediately following the intervention, and again after an eight-week no treatment phase. Due to the length of the testing protocol (and the length of a 40-item SCT), performance was measured on the SCT only at Baseline 1 (B1) and Outcome 1 (O1 – immediately after therapy). As a result, we were unable to exploit the main study's RCT design and instead we combined data across the two groups in a pre-/post-treatment comparison.

Here we examine whether a SCT provided indications of change in sentence production following therapy. We addressed the following research questions (RQs):

(1) Does treatment directed at high-frequency constructions result in improved ability to retrieve those constructions in a SCT?
(2) Does the intervention result in an increase in the number of grammatically well-formed SCT responses?
(3) Do participants show change in untreated frequency/length/structure matched constructions?
(4) Exploratory: (a) Is SCT performance at baseline associated with sentence comprehension and/or digit span performance; and (b) are SCT change scores (post- minus pre-therapy) predicted by baseline demographic or behavioural profiling variables such as age or aphasia severity?

Methods

Ethical approval for the research was granted by the Department of Language & Cognition Research Ethics Committee for a test norming study (LCD-2019–13) and the UCL Research Ethics Committee for the therapy study (8123/001). All participants gave written informed consent prior to taking part in the research.

SCT materials

In devising the SCT, we employed the methodology described by Goodglass et al. (1972), but the content was based on the constructions targeted in the current therapy and their untreated twins. The intervention targeted 20 high-frequency constructions. Each treated construction (e.g., *I like it*) was matched to an untreated item (e.g., *I keep it*). Pairs ($n = 20$) were matched on string frequency in the BNC, and sentence structure. This resulted in matching on other key variables that might influence sentence processing such as argument structure, verb ergativity, and length (in words). Appendix 2a lists the story stems, the target constructions and their bi-/trigram (two/three-word combinations) association strength (t-scores); Appendix 2b contrasts each treated vs. untreated item pair. Regarding the treated vs. untreated pairs, a Mann-Whitney U test demonstrated no difference in string frequency between treated and untreated items, $U = 174$, $p = .495$. We devised 40 story stems to elicit treated and untreated constructions. Target constructions did not appear in story stems; thus, production could not be due to immediate memory recall or repetition. Story stem length ranged between 14–25 words, and stems were matched in length (as measured by number of words) across treated (average length: 20.7 words) and untreated items (average length: 20.8 words), $U = 198$, $p = .968$. Story stems were audio-recorded to ensure stimuli were identical across participants and assessment points to eliminate any extralinguistic cues, and then randomly assigned to a single list. All participants completed the test in this fixed random order at both baseline and outcome.

SCT norming study

Prior to conducting the intervention study, twenty-two neurotypical adults were recruited (14 males, 8 females, mean age = 57 years, range 39–69, $SD = 7.49$). They were all native English speakers with no history of speech, language or cognitive impairment. Data were collected face-to-face for 12 participants and via Microsoft Teams for the remaining participants due to COVID restrictions in place at the time. Participants heard the recorded story stems and were asked to complete them with a single sentence. Responses were audio-recorded and subsequently transcribed orthographically. The number of responses that included the lexically-specific target construction was counted. Some stories were more successful in eliciting the target construction than others. Therefore, each item was weighted by normative cloze probability, by dividing the number of target answers by the total number of responses for that item. For example, the story stems for *I see that, I keep it, Do you have coffee*, resulted in production of the target construction in all participants and thus were weighted at 1. By contrast, *I found it*, and *I want to look* had lower probabilities of target production (both 0.19). The mean cloze probability for the treated items was 0.64 ($SD = 0.22$), and for the untreated control constructions 0.69 ($SD = 0.26$). An independent samples t-test showed that the cloze probabilities of the two lists (treated vs untreated) were similar, $t(38) = .756$, $p = .454$, $d = .239$ (small effect). Appendix 2a lists the normative cloze probabilities for each item. Total values for treated and untreated structures were 12.72 and 13.87, respectively. In this way, the scoring of PwA responses (for the expected structure scale) was informed by normative cloze probabilities.

Aphasia intervention study

The UTILISE study recruited 39 participants with chronic aphasia, with 33 completing the intervention. Recruitment was not restricted to people with non-fluent, agrammatic aphasia; participants had moderate-to-severe aphasia after a single stroke (full inclusion criteria can be found in Bruns et al., 2025). Two participants had a significant amount of missing SCT data at either one or both timepoints (due to recording error or frustration with the task). We report results from 31 participants for the SCT.

The 31 participants (19 males, 12 females; mean age = 60 years, $SD = 11$) all had chronic aphasia (mean time post-onset = 86 months, $SD = 76$) which varied in severity and aphasia

Participant demographics (= 31).

	Mean	Range	
Age at B1	59.81	25–87	11.03
Time post onset at B1 (months)	86.42	9–319	76.36
Years of education	15.29	11–18	2.15
Occupational classification[1]	3.71	1–8	1.94
Handedness (EHI)[2] (61–100 = right handers)	84.68	−100 – 100	39.88
Object naming (BNT)[3] (max. 60)	16.87	0–55	16.62
Spoken word comprehension (CAT)[4] (max. 30)	23.65	9–30	5.98
Sentence comprehension (TROG-2 blocks) (max. 20)[5]	4.29	0–14	4.02
Digit span (PALPA-13)[6,7] (max. 7)	2.63	0–6	1.71
Non-verbal IQ (WASI-II MR)[8] (max. 30)	14.29	6–23	5.20

Notes: 1: Based on Standard Occupational classification 2020, Office for National Statistics (2020); 2: EHI = Edinburgh Handedness Inventory – Short Form, adapted from Veale (2014); 3: BNT = Boston Naming Test, Kaplan et al. (2001); 4: CAT = Comprehensive Aphasia Test, Swinburn et al. (2004); 5: TROG-2 = Test for Reception of Grammar, Bishop (2003); 6: PALPA-13 by verbal repetition or pointing = Subtest 13 of the Psycholinguistic Assessments of Language Processing in Aphasia, Kay et al. (1992); 7: The range of digit span scores was wide in the sample (0–6). Seven participants had a score of 0. Digit span was associated with aphasia severity. The likely interpretation of low scores is that these participants had severe aphasia, reflecting a pragmatic sampling strategy in early-stage development of a therapy; 8: WASI-II MR = Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence, Wechsler (2011).

profile (Table 1; see Appendix 3 for more details). At trial entry, information was recorded on laterality, language(s) used, years of formal education and previous occupation, as well as digit span and non-verbal IQ (WASI-II Matrix Reasoning, Wechsler, 2011) (Table 1). We also collected repeated baseline scores on primary outcome measures of spoken sentence comprehension (TROG-2, Bishop, 2003) and connectivity of speech in narratives (see Bruns et al., 2025, for further information). We also created a composite measure of aphasia severity based on scores on spoken word comprehension (subtest from Comprehensive Aphasia Test, CAT, Swinburn et al., 2004) and naming (Boston Naming Test, BNT, Kaplan et al., 2001).

A full description of the UTILISE intervention can be found in Bruns et al. (2025), together with a TiDiER summary (Hoffmann et al., 2014), and the TiDiER summary (Hoffmann et al., 2014) is presented in Appendix 1 of this report. The trial was prospectively registered (ISRCTN14466044) and the pre-registered protocol can be found at https://osf.io/qvpjm. The SCT was a secondary outcome measure for the study. UTILISE-1 was a feasibility study of a novel usage-based sentence therapy. It was delivered at low dose (approximately 12 hours total dosage) in a university clinic setting. The therapy consisted of three computerised tasks: two auditory input tasks (task 1: judging if two sentence strings were same/different; task 2: button press response to a target word in a sentence), and a spoken sentence production task (task 3). The 20 target constructions appeared in all therapy tasks to prime/stimulate these structures. Only the treated constructions appeared in the therapy. The matched untreated constructions appeared only in baseline and outcome testing.

Participants completed the SCT at Baseline (B1) and Outcome (O1), with a gap of approximately 8 weeks for the immediate condition, and 12 weeks in the deferred condition (including a 4-week therapy phase). Participants listened to audio-recorded story stems and were asked to complete each story by producing a single sentence. Responses were audio-recorded and the resultant recordings were assigned a randomised tag to mask timepoint (Baseline/Outcome). Responses were transcribed by a researcher who did not administer the assessment/intervention and was thus masked to timepoint. Transcription was orthographic, with phonetic/phonological errors ignored if the intended target was identifiable.

Responses were coded by researchers masked to timepoint, on two binary scales: (1) presence of target structure; (2) whether the sentence produced was grammatically well-formed. Coding rules were developed using the responses of the neurotypical controls and subsequently manualised to support fidelity in scoring data from aphasic participants (Appendix 4). For target (expected) structure scoring, responses were scored as correct if they contained the same phrase structure elements as the target construction, irrespective of lexical selection. For example, for the target *I hate it*, the following responses would all be scored as correct: *I loathe it; He hates it*. For target structure scoring of untreated items, the response had to both match the target structure and include a different main verb from the corresponding treated construction. For example, for the untreated Subject-Verb-Object item "*I lost it*", which was paired with the treated item "*I said it*", the response "*someone nicked it*" was coded as correct as the main verb differed from the treated verb "*say*", but the response still reflected the intended underlying SVO structure. Without this criterion, a participant could be credited with successful generalisation through producing the treated exemplar of a construction.

Responses with the target inside an embedded clause were scored as correct (e.g., "*I said [I hate it]"*). Within-phrase omissions and errors on function words were ignored (e.g., "*I've got train"* was correct for target structure NP V NP). A response with omission of a phrasal head was scored as both off-target and ungrammatical (e.g., "*I've got a"*). Single word responses were scored as non-target and not grammatically well-formed. Social speech was ignored in scoring: e.g., for the target *help me*, "*help please"* and *"yes help thank you"* would be scored as incorrect on both target and well-formedness criteria. Unintelligible responses scored zero, as were perseverative responses, social and task-directed speech (e.g., "*can't do it"*). Where a participant produced multiple responses, the "best" attempt was scored for both target structure and grammatical well-formedness. More details can be found in Appendix 4.

One researcher (RV) coded all data, with a second researcher (KD) independently coding 14 samples ($n = 503$ items) to determine inter-rater reliability (23% of the total 62 samples). Both were masked to phase (Baseline/Outcome). The two raters had an agreement of 93% for expected structure, and 91% agreement for grammatical well-formedness ratings. An unweighted Cohen's kappa (JASP Team [JASP], 2025) revealed substantial agreement between raters (using benchmarks suggested by Landis & Koch, 1977) for both expected structure ($\kappa = .73$) and well-formedness ratings ($\kappa = .79$).

Coding of participant transcriptions resulted in four raw scores: Expected Structure for treated vs untreated items, and Well-formedness for treated vs untreated items. For Expected Structure, if the target was elicited, the participant was given the cloze probability weighted score from the normative sample (ranging from .19 to 1.00 across the 40 items, see Appendix 2). For Well-formedness, each grammatically correct multi-word utterance was assigned a score of 1, resulting in a maximum score of 20 for treated and untreated items, respectively.

For each participant, raw scores were converted into percentages. For example, Participant 20 had a raw score for Expected Structures (treated items) of 6.99 (maximum score = 12.72) which resulted in a percentage score of 55%. For well-formed responses (treated items), this participant scored 19 (maximum score = 20 items) which corresponded to a percentage of 95%. In this way, we were able to obtain scores for participants who had incomplete data sets (see Appendix 3).

We used JASP (version 0.95, JASP, 2025) and SPSS (Version 31, IBM Corp [IBM], 2025) to perform parametric assumption checks (e.g., normality), and R (https://www.r-project.org/) to test our main hypotheses, using the ANOVA-type statistic (ATS) and relative treatment effect (RTE), a non-parametric alternative to traditional ANOVA suitable for non-normal data (nparLD package in R; Langsdorf et al., 2022; Noguchi et al., 2012). Specifically, two within-subject analyses examined the effect of Condition (Treated vs Untreated items) and Time (B1 vs O1) on Expected Structures and Well-formedness. Two-sample non-parametric rank based post-hoc tests were used to follow up any significant results (nparLD).

As use of a SCT as a therapy outcome measure is relatively novel, we also sought to explore relationships at baseline between performance and other language capacities that appear to be necessary for processing the story stem. In particular, we examined relationships with digit span as a measure of phonological working memory and spoken sentence comprehension (RQ 4a). To evaluate Research Question 4b, we examined if change in SCT scores could be linked to any predictors that are often employed in understanding therapy outcome. For this analysis, we examined associations between SCT change scores and aphasia severity (composite score calculated based on BNT and CAT spoken word-picture matching test) and age. In these exploratory analyses, we used JASP and SPSS to conduct correlational and stepwise multiple regression analyses.

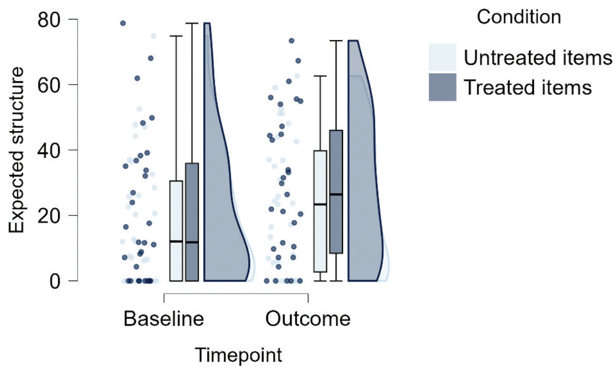
Results

We first analysed participants' ability to retrieve target structures (Expected Structure). At baseline, there was no difference between scores for Treated vs Untreated items. An increase in median scores for both Treated and Untreated items was observed after therapy (Table 2). The assumption of normality was not met across Expected Structure scores as Shapiro-Wilk tests were significant due to a small proportion of 0 scores (floor performance, see Figure 1). The non-parametric rank-based omnibus analysis (nparLD) indicated a significant main effect of Time on Expected Structure scores, $F(1) = 11.82$, $p < .001$, and a significant main effect of Item (Treated vs Untreated), $F(1) = 7.49$, $p = .006$

Descriptive statistics and effect sizes (relative treatment effects, RTE) for Expected structure and Well-formedness outcomes.

	Condition	Time point	Median(%)	IQR	RTE	CI Lower	CI Upper
Expected structure	Treated	B1	11.79	36.79	0.48	0.44	0.52
		O1	26.42	40.10	0.58	0.53	0.62
	Untreated	B1	12.05	32.61	0.44	0.41	0.47
		O1	23.38	42.57	0.51	0.48	0.54
Well-formedness	Treated	B1	45.00	55.00	0.47	0.44	0.50
		O1	55.00	70.00	0.55	0.52	0.58
	Untreated	B1	45.00	55.00	0.46	0.43	0.49
		O1	50.00	55.00	0.53	0.50	0.56

Notes: IQR = interquartile range, CI = confidence interval.

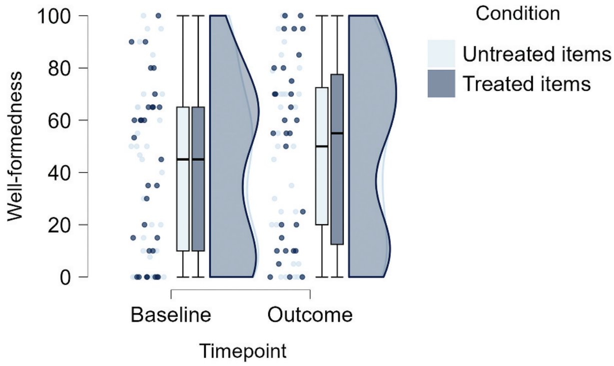


SCT Expected Structure percentage scores at Baseline and Outcome for Treated and Untreateditems (based on = 31 participants).

(Figure 1). There was no interaction effect, $F(1) = .99$, $p = .32$. Pairwise contrasts using rank-based tests with t-approximations (nparLD) indicated scores were significantly higher post-therapy for both Treated, $t(30) = 2.62$, $p = .014$, and Untreated items, $t(30) = 3.24$, $p = .003$. The relative treatment effects (RTE) indicate a greater probability of higher scores after therapy (versus B1) for Treated and Untreated items (Table 2). For instance, the RTE of .58 for Treated O1 can be interpreted as follows: if we selected a participant's score from Treated O1, this is more likely to be higher than a random score from the whole sample (probability of 58% that it will be a higher score).

Figure 2 visualises Well-formedness percentage scores by Time and Condition. Again, the baseline scores for Treated vs Untreated items were similar. An increase in median Well-formedness scores after therapy was observed for both Treated items and Untreated items (Table 2).

The rank-based omnibus analysis (nparLD) revealed a significant main effect of Time (B1 vs O1), $F(1) = 14.80$, $p < .001$. There was no significant effect of Item (Treated vs Untreated), $F(1) = .81$, $p = .37$, and no interaction effect, $F(1) = .21$, $p = .64$. Pairwise rank-based contrasts (nparLD) indicated improvements after therapy were significant for both Treated items (B1 vs O1), $t(30) = 3.17$, $p = .004$, and Untreated items (B1 vs O1), t



SCT Well-formedness percentage scores at B1 and O1 for Treated and Untreated items (based on = 31 participants).

(30) = 2.98, *p* = .006. Relative treatment effect (RTE) scores again indicated a greater probability of higher scores after therapy (versus B1) for both Treated and Untreated items (Table 2). Appendix 5 visualises the RTEs for Expected Structure and Well-formedness.

Since the SCT is a bespoke measure, we explored whether performance on standardised tests (sentence comprehension ability: TROG-2 scores, and digit span) was associated with baseline SCT performance. We averaged SCT scores across Treated and Untreated items at baseline for both Expected Structure and Well-formedness analyses. Table 3 lists the correlations. Due to non-normality, nonparametric correlations (Spearman's rho) were used. Results revealed a small significant association between SCT Expected Structure scores and TROG-2 scores. While TROG-2 was significantly associated with retrieving target constructions, there was no significant association with ability to form grammatically well-formed responses. Associations between both SCT measures and digit span scores were moderate but non-significant.

Correlations between Expected structure and Well-formedness baseline scores (averaged across treated and untreated items) and baseline profiling variables.

		Spearman's rho		Effect size (Fisher's)	SE Effect size
Well-formedness scores (baseline average)	31	.939***	<.001	1.729	0.208
Digit span score	30	0.298	.110	0.307	0.193
TROG-2 blocks at baseline (B1)	31	0.385*	.032	0.406	0.191
Digit span score	30	0.324	.081	0.336	0.194
TROG-2 blocks at baseline (B1)	31	0.342	.060	0.356	0.190

Note: * < .05, *** < .001.

Next, we explored whether Age or Aphasia Severity predicted change in SCT performance/therapy response. For this analysis, SCT change scores (Outcome-Baseline) were used as outcome variables. Hierarchical regression analysis was conducted to examine whether Severity (a composite score composed of BNT and CAT spoken word comprehension) predicted Expected Structure and Well-formedness Change, beyond a potential age effect. All assumptions for linear regression were met. In Model 1, Age was entered as a single predictor; in Model 2, Age and Severity were both entered as predictors. Age and Aphasia Severity were regarded as separate predictors, based on a correlational analysis showing a nonsignificant relationship between the two variables ($r(29) = -.090$, $p = .629$).

For Expected Structure Change, the model with Age alone (Model 1) was not statistically significant, $F(1, 29) = .102$, $p = .751$; Age did not significantly predict Expected Structure change, $\beta = -.059$, $t(29) = -.320$, $p = .751$, $R^2 = .004$. The addition of Severity (Model 2) resulted in an increase in explained variance of 13% (R^2 Change = .125); however, the model was not significant, $F(2, 28) = 2.065$, $p = .146$. Severity was a near-significant predictor of Expected Structure change, $\beta = .355$, $t(28) = 2.004$, $p = .055$, $R^2 = .129$, with a moderate effect size ($f^2 = 0.15$).

For Well-formedness Change, the results showed a similar pattern: the model with Age alone (Model 1) was not statistically significant, $F(1, 29) = .159$, $p = .693$, indicating that Age did not significantly predict change on Well-formedness, $\beta = -.074$, $t(29) = -.399$, $p = .693$, $R^2 = .005$. Again, Model 2 resulted in a similar increase in explained variance of 13% (R^2 Change = .129), $F(2,28) = 2.173$, $p = .133$. Severity was again a near-significant predictor of change in Well-formedness, $\beta = .361$, $t(28) = 2.042$, $p = .051$, $R^2 = .134$, with a moderate effect size ($f^2 = 0.15$).

Discussion

This study reports outcomes of a SCT for a therapy (UTILISE) directed at sentence processing impairments. A companion article (Bruns et al., 2025) described outcomes of this therapy on spontaneous narrative speech production and reported significant improvements in speech connectivity (proportion of grammatical three-word combinations, trigrams) after intervention. However, the constructions treated in therapy did not necessarily appear in the spontaneous narratives. Here we describe outcomes on a SCT that allowed access to treated constructions to be probed, thus representing a more proximal outcome measure of change. Furthermore, informed by a usage-based approach to sentence representation, the SCT method demonstrates it is possible to create a frequency/structure/length matched set of untreated items to explore generalisation of any therapy gains at sentence-level.

With regard to our main research questions, the results indicated that the intervention directed at high-frequency constructions such as *I don't know* resulted in improved ability to produce treated structures in a SCT (RQ1). In addition, there were significant increases in the number of grammatically well-formed responses post-therapy (RQ2). Participants also showed positive change in the production of target structures as well as well-formed responses in matched untreated items (RQ3). Relative treatment effects indicated greater probability for improvement in Expected Structure scores for Treated items in comparison to Untreated items, but there was no significant interaction between Time and Condition (Treated vs Untreated items). For Well-formedness scores, relative treatment effects were similar across Treated and Untreated items. This suggests that usage-based Construction Grammar is of value in designing sentence therapy for aphasia.

However, some caution should be exercised in interpreting these results. The SCT was a secondary outcome measure in our study. Due to the length of the testing protocol, we made the decision only to obtain one pre-therapy baseline (Baseline 1). The 40-item SCT was demanding for many participants. It took between 15 to 40 minutes to administer and the task places demands on both auditory comprehension and sentence production. The assessment had to be shortened or abandoned with some participants due to frustration with the task. As a result of our design, we were unable to exploit the RCT structure of the study through a treatment/no-treatment comparison. Such a comparison would require participants to complete the SCT both at Baseline 1 and Baseline 2. Due to the absence of repeated baselines, one possible explanation of the increase in scores over time might be a general factor of increasing familiarity with the task and experimenter, or the general psychological effects of participation in a therapy study. As a result, there is a need for caution in assuming any causal relationship between the intervention and SCT results. However, in our companion article (Bruns et al., 2025), where repeated baselines were recorded on primary outcome measures thus allowing stronger causal inferences to be made, we observed modest improvements in speech connectivity following therapy. Future research could focus on increasing the number of baseline sessions rather than the number of tasks administered in a session in order to complete multiple evaluations of all outcome measures in addition to broader cognitive-linguistic profiling.

Despite the best efforts of the researchers to devise stories that would elicit the intended structures, some items were less effective than others. Since the high-frequency constructions targeted in therapy were often semantically neutral (e.g., "*I made it*"), it was challenging to create story contexts that would elicit a certain target structure with high probability even in neurotypical participants. This resulted in several SCT items with low cloze probabilities. For this study, the items that elicited cloze probabilities of or below 0.50 were considered "low cloze" (listed in Appendix 2). It may be the case that some targets are particularly difficult to elicit, regardless of story prompt. While other researchers omitted low-cloze items (e.g., Bailey et al., 2024), we accounted for this variability by adjusting the scoring system. The researchers may review the low-cloze items in future research, to investigate whether alternative probe stories would be more effective at eliciting the target structures.

Limitations of study design aside, we can speculate on the source of the observed generalisation of therapeutic effect to untreated constructions. Unlike a lexical therapy study where form of untreated items is entirely distinct from treated ones, in this study the verbs of the matched constructions were different, but there was lexical overlap on positions around the verb (e.g., *I had it/I found it*). One possibility is that this overlap facilitated transfer across constructions, for example, by offering an initiation cue from the first word. This account may not be fully satisfactory as an associative cue from the first item might lead to production of the verb trained in therapy and the scoring of the untreated items on Expected Structure required that the verb was different to that of its treated twin. Another alternative is that the therapy resulted in facilitation at the level of abstract frames. For example, *[X] [VERBS] [Y]* constructions were predominant in the intervention, and repeated stimulation might result in increased accessibility of the abstract frame. As the treated and untreated items of the SCT had shared structure, this resulted in facilitation for both sets of items. The additional lexical overlap across treated and untreated items, together with use of very high-frequency forms such as pronouns, might also enable easier allocation of words to the slots in the frame (Tooley, 2020; van Boxtel et al., 2025).

In more general terms, the positive outcomes of UTILISE therapy reported here and in Bruns et al. (2025) are unlikely to be due to a single mechanism of action. The intervention was based on a number of design principles. In addition to high-frequency content, the therapy involved both input and production components. The aim of the listening tasks was to prime target constructions prior to production based on connectivity between input and output systems (Varley, 2011). The listening phases were also part of a strategy to reduce errors in production and increase the likelihood of practice of fluent production (Whiteside et al., 2012). Finally, and unlike therapies which involve explicit instruction (e.g., on the formation of passive structures, or how words map to particular slots in a sentence frame), the therapy was implicit in the sense that the participant was not required to reflect or make conscious decisions on language. Any one or any combination of these components might contribute to the positive outcomes.

Our final exploratory research questions addressed associations between baseline SCT performance and sentence comprehension (TROG-2, Bishop, 2003) and digit span scores (RQ4a). We observed high correlations between SCT scores (Expected Structure and Well-formedness), indicating these measures capture largely overlapping information regarding SCT performance. We observed modest positive correlations between SCT and TROG-2, and SCT and digit span; however, only one of these reached statistical significance (TROG-2 and Expected Structure). This may be due to low power rather than lack of true associations (post-hoc power calculations indicate around 40% power for small associations (.30)). These associations suggest that the SCT is not a pure measure of production but, consistent with its use in the ANELT (Blomert et al., 1995) as a measure of functional communication, also reflects the ability of an individual to switch between listener/comprehender and speaker roles. Moreover, this result suggests that use of a SCT with individuals with profound comprehension impairment is unlikely to produce a valid measure of sentence production ability. For SCT and digit span, the moderate (but non-significant) associations suggest that recall of isolated lexical items (i.e., non-syntactically organised strings of numbers) as tested in a digit span task involves different processes to those employed in online sentence comprehension/production.

We also explored whether change in SCT performance from baseline to outcome could be predicted by factors of Age and Aphasia Severity (RQ4b). Our aim here was to understand more about possible predictors of positive therapy outcomes (Doogan et al., 2018). Identification of such predictors would allow allocation of patients to specific therapy regimes on a systematic basis versus costly trial-and-error allocation. The latter is particularly problematic when receipt of therapy services is often limited to a short period in the acute/sub-acute period of living with aphasia. In line with the findings of Ellis and Urban (2016), we found that Age did not predict SCT change. By contrast, Severity explained a small amount (13%) of variance in SCT change. However, these results were borderline significant, possibly due to low statistical power. Accordingly, these findings should be taken as preliminary. Nonetheless, this pattern does parallel a larger study ($N =$

107) exploring predictors of outcomes of naming therapy, where Kristinsson et al. (2023) report that aphasia severity (and to a lesser degree age) predicted therapy response.

In combination with our companion article (Bruns et al., 2025) our results suggest that usage-based Construction Grammar can provide useful insights into motivating sentence therapies for aphasia. In addition to UTILISE therapy being directed at high-frequency sentences that are used to navigate everyday communicative interactions, the approach allows the methodological refinement of assessing therapy outcomes via sets of matched treated and untreated constructions in a SCT to assess generalisation effects. Furthermore, a SCT may represent a viable clinical assessment tool in that it can be scored relatively quickly and, in comparison to measures of spontaneous speech, is designed to elicit specific structures and does not require extensive training in skills such as tagging and analysis. Overall, the SCT not only complements the broader assessment battery of spontaneous speech samples, but it also has the potential to serve as a sensitive and ecologically meaningful proximal measure for detecting therapy outcomes at the sentence level.

Notes

1. The t-scores measure the degree of association between the words of each trigram (three-word combination), where higher t-scores correspond to more strongly associated/collocated words, relative to how often they would appear together if words occurred randomly in the corpus (see Gries, 2010; Zimmerer et al., 2018, for an explanation of t-scores).

Acknowledgements

We would like to thank the participants for volunteering to take part in the research. We are grateful to Ellen Harrington O'Gorman, Rosie Lucas and Tae Horsfield for their assistance with data collection and transcription.

Disclosure statement

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

Funding

The study was funded by the Stroke Association grant [TSA SA PG 18\100054]. The views expressed are those of the authors and not necessarily those of the funder.

ORCID

Victoria Fleming http://orcid.org/0000-0003-2582-4914
Rosemary Varley http://orcid.org/0000-0002-1278-0601

Data availability statement

Anonymised data (subject to participant consent to data sharing) are available from RV in response to reasonable request.

Trial registration

Prospectively registered on 13/09/2019 at ISRCTN14466044 https://doi.org/10.1186/ISRCTN14466044

References

Bailey, D. J., Bunker, L., & Wambaugh, J. L. (2024). Feasibility of an abstract verb naming treatment for aphasia. *Aphasiology*, *38*(7), 1303–1329. https://doi.org/10.1080/02687038.2023.2279180

Barde, L. H. F., Schwartz, M. F., & Boronat, C. B. (2006). Semantic weight and verb retrieval in aphasia. *Brain and Language*, *97*(3), 266–278. https://doi.org/10.1016/j.bandl.2005.11.002

Bastiaanse, R., & Edwards, S. (2004). Word order and finiteness in Dutch and English Broca's and Wernicke's aphasia. *Brain and Language*, *89*(1), 91–107. https://doi.org/10.1016/S0093-934X(03)00306-7

Bishop, D. V. M. (2003). *The test for Reception of Grammar: TROG-2*. Pearson Assessment.

Blomert, L., Koster, C., & Kean, M. L. (1995). *Amsterdam-Nijmegen Test voor Alledaagse Taalvaardigheid: Handleiding*. Swets & Zeitlinger.

BNC Consortium. (2007), *The British National corpus, XML edition*. http://www.natcorp.ox.ac.uk/

Bruns, C., Rodgers, F., Dathan, K., Dean, M., Warren, J., Fleming, V., Javadi, A.-H., & Varley, R. (2025). Reconstructing sentence processing in aphasia: A randomised control trial of a usage-based intervention. *Aphasiology*, 1–26. https://doi.org/10.1080/02687038.2025.2530572

Bruns, C., Varley, R., Zimmerer, V. C., Carragher, M., Brekelmans, G., & Beeke, S. (2019). "I don't know": A usage-based approach to familiar collocations in non-fluent aphasia. *Aphasiology*, *33*(2), 140–162. https://doi.org/10.1080/02687038.2018.1535692

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

Caplan, D., & Hanna, J. E. (1998). Sentence production by aphasic patients in a constrained task. *Brain and Language*, *63*(2), 184–218. https://doi.org/10.1006/brln.1998.1930

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern assessment of verbs and sentences (NAVS). *Aphasiology*, *26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Diessel, H. (2015). Frequency shapes syntactic structure. *Journal of Child Language*, *42*(2), 278–281. https://doi.org/10.1017/S0305000914000622

Doogan, C., Dignam, J., Copland, D., & Leff, A. (2018). Aphasia recovery: When, how and who to treat? *Current Neurology and Neuroscience Reports*, *18*(12), 90. https://doi.org/10.1007/s11910-018-0891-x

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

Ellis, C., & Urban, S. (2016). Age and aphasia: A review of presence, type, recovery and clinical outcomes. *Topics in Stroke Rehabilitation*, *23*(6), 430–439. https://doi.org/10.1080/10749357.2016.1150412

Fink, R. B., Schwartz, M. F., Rochon, E., Myers, J. L., Socolof, G. S., & Bluestone, R. (1995). Syntax stimulation revisited. *American Journal of Speech-Language Pathology*, *4*(4), 99–104. https://doi.org/10.1044/1058-0360.0404.99

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language*, *33*(3), 217–222. https://doi.org/10.1055/s-0032-1320041

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

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*(C), 451–471. https://doi.org/10.1016/S0093-934X(75)80083-6

Goldberg, A. E. (1995). *A construction Grammar approach to argument structure*. The University of Chicago Press.

Goodglass, H., Gleason, J. B., Bernholtz, N. A., & 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

Gries, S. T. (2010). Useful Statistics for corpus Linguistics. In A. S. Pérez & M. A. Sánchez (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269–291). Peter Lang.

Helm-Estabrooks, N., & Ramsberger, G. (1986). Treatment of agrammatism in long-term Broca's aphasia. *International Journal of Language & Communication Disorders*, *21*(1), 39–45. https://doi.org/10.3109/13682828609018542

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Michie, S., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*(mar07 3), g1687. https://doi.org/10.1136/bmj.g1687

IBM Corp. (2025). *IBM SPSS Statistics for windows* (Version 31.0.0) [Computer software].

JASP Team. (2025). *JASP* (Version 0.95) [Computer software].

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston naming test* (2nd ed.). Lippincott Williams & Wilkins.

Kay, J., Coltheart, M., & Lesser, R. (1992). *Psycholinguistic assessments of Language processing in Aphasia*. Lawrence Erlbaum Associates.

Khoshhal Mollasaraei, Z., den Ouden, D., Fridriksson, J., Hickok, G., & Matchin, W. (2025). A review of sentence elicitation techniques for assessing grammatical deficits in aphasia. *Aphasiology*, 1–25. https://doi.org/10.1080/02687038.2025.2570803

Kristinsson, S., Basilakos, A., den Ouden, D. B., Cassarly, C., Spell, L. A., Bonilha, L., Rorden, C., Hillis, A. E., Hickock, G., Johnson, L., Busby, N., Walker, G. M., McLain, A., & Fridriksson, J. (2023). Age and severity impact naming therapy. Bigger sample size: Predicting outcomes of Language Rehabilitation: Prognostic factors for immediate and long-term outcomes after aphasia therapy. *Journal of Speech, Language, and Hearing Research*, *66*(3), 1068–1084. https://doi.org/10.1044/2022_JSLHR-22-00347

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Langsdorf, L. E., Kübler, S., & Schubert, T. (2022). Investigation of reward effects in overlapping dual-task situations. *Acta Psychologica*, *222*, Article 103465. https://doi.org/10.1016/j.actpsy.2021.103465

Noguchi, K., Gel, Y. R., Brunner, E., & Konietschke, F. (2012). nparLD: An R software package for the nonparametric analysis of longitudinal data in factorial experiments. *Journal of Statistical Software*, *50*(12), 1–23. https://doi.org/10.18637/jss.v050.i12

Office for National Statistics. (2020). *Standard occupational classification (SOC) 2020*. https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/soc2020/soc2020volume1structureanddescriptionsofunitgroups

Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., Latimer, N., Julious, S., Cross, E., Alshreef, A., Harrison, M., McKevitt, C., Farrin, A., & Chater, T. (2019). Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (big CACTUS): A multicentre, single-blinded, randomised controlled trial. *Lancet Neurology*, *18*(9), 821–833. https://doi.org/10.1016/S1474-4422(19)30192-9

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

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, *37*(3), 440–479. https://doi.org/10.1016/0093-934X(89)90030-8

Swinburn, K., Porter, G., & Howard, D. (2004). *The Comprehensive aphasia test*. Psychology Press.

Thompson, C. K., Choy, J. J., Holland, A., & Cole, R. (2010). Sentactics®: Computer-automated treatment of underlying forms. *Aphasiology*, *24*(10), 1242–1266. https://doi.org/10.1080/02687030903474255

Tooley, K. M. (2020). Contrasting mechanistic accounts of the lexical boost. *Memory & Cognition*, *48*(5), 815–838. https://doi.org/10.3758/s13421-020-01019-3

van Boxtel, W. S., Rainey, K. M., Ferreira, V., Martin, N., Bauman, E., & Lee, J. (2025). Structural priming treatment in aphasia: The role of lexical and abstract syntactic representations. *American Journal of Speech-Language Pathology*, *34*(6), 3318–3342. https://doi.org/10.1044/2025_AJSLP-25-00095

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

Veale, J. F. (2014). Edinburgh Handedness inventory - short form: A revised version based on confirmatory factor analysis. *Laterality*, *19*(2), 164–177. https://doi.org/10.1080/1357650X.2013.783045

Wechsler, D. (2011). *Wechsler abbreviated scale of intelligence - 2nd. Ed. (WASI-II)*. Psychological Corporation.

Whiteside, S. P., Inglis, A. L., Dyson, L., Roper, A., Harbottle, A., Ryder, J., Cowell, P. E., & 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

Wilshire, C. E., Lukkien, C. C., & Burmester, B. R. (2014). The sentence production test for aphasia. *Aphasiology*, *28*(6), 658–691. https://doi.org/10.1080/02687038.2014.893555

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Zatorski, R. J., & Lesser, R. (1981). The lexicon and sentence generation in aphasia. *Brain and Language*, *13*(1), 185–190. https://doi.org/10.1016/0093-934X(81)90138-3

Zimmerer, V. C., Hardy, C. J. D., Eastman, J., Dutta, S., Varnet, L., Bond, R. L., Russell, L., Rohrer, J. D., Warren, J. D., & Varley, R. A. (2020). Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex*, *133*, 103–119. https://doi.org/10.1016/j.cortex.2020.08.027

Zimmerer, V. C., Newman, L., Thomson, R., Coleman, M., & Varley, R. A. (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