Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research Report

Speech pauses in speakers with and without aphasia: A usage-based approach

Sebastian Bello-Lepe [a,b,1], *Sabrina Mahmood* [a], *Rosemary Varley* [a] and *Vitor Zimmerer* [a,*]

[a] *University College London, Department of Language and Cognition, London, UK*

[b] *Universidad de Valparaíso, Centro de Investigación del Desarrollo en Cognición y Lenguaje, Valparaiso, Chile*

ARTICLE INFO

Article history:
Received 26 April 2023
Reviewed 14 July 2023
Revised 28 September 2023
Accepted 24 June 2024
Action editor Daniel Mirman
Published online 14 July 2024

Keywords:
Pauses
Aphasia
Connected speech
Collocation strength
Usage-based approaches

ABSTRACT

Pauses in speech are indicators of cognitive effort during language production and have been examined to inform theories of lexical, grammatical and discourse processing in healthy speakers and individuals with aphasia (IWA). Studies of pauses have commonly focused on their location and duration in relation to grammatical properties such as word class or phrase complexity. However, recent studies of speech output in aphasia have revealed that utterances of IWA are characterised by stronger collocations, i.e., combinations of words that are often used together. We investigated the effects of collocation strength and lexical frequency on pause duration in comic strip narrations of IWA and non-brain-damaged (NBD) individuals with part of speech (PoS; content and function words) as covariate. Both groups showed a decrease in pause duration within more strongly collocated bigrams and before more frequent content words, with stronger effects in IWA. These results are consistent with frameworks which propose that strong collocations are more likely to be processed as holistic, perhaps even word-like, units. Usage-based approaches prove valuable in explaining patterns of preservation and impairment in aphasic language production.

1. Introduction

Pauses in speech are indicators of cognitive effort. The frequency, location, and duration of filled (e.g., *um*) or silent pauses reveal neurocognitive processes underpinning language production in both non-brain-damaged (NBD) speakers and individuals with neurocognitive pathologies, such as aphasia (Butterworth, 1979; Hird & Kirsner, 2010; Klatt, 1980; Watanabe et al., 2008). For example, pause duration can be longer before the initiation of sentences with high syntactic complexity (Ferreira, 1991; Grosjean et al., 1979) or longer noun phrases (Strangert, 1997). The grammatical role of a word or part of speech (PoS) further influences pauses in spontaneous speech: in NBD English speakers, pauses tend to

* *Corresponding author.*
E-mail address: v.zimmerer@ucl.ac.uk (V. Zimmerer).

[1] Deceased author.

https://doi.org/10.1016/j.cortex.2024.06.012

occur more before content rather than function words (Maclay & Osgood, 1959), and are longer before verbs relative to nouns (Seifart et al., 2018). Also, pauses in speech are longer before or within utterances with non-canonical syntactic structures (e.g., passive sentences) (Krivokapić, 2007; Krivokapić et al., 2022; Ruder & Jensen, 1972). In individuals with aphasia (henceforth, IWA), pauses are longer and more frequent than in NBD individuals (Angelopoulou et al., 2018; Sahraoui et al., 2015). Pauses in speech may also capture variance in populations with acquired language disorders, such as individuals with post-stroke aphasia or primary progressive aphasia (DeDe & Salis, 2020; Hird & Kirsner, 2010; Mack et al., 2015; Potagas et al., 2022).

Most studies on the effect of linguistic factors on pauses have been conducted under the framework of Generative Grammar and related theories, which determine processing difficulty by the properties of individual words and complexity of phrase structures. By contrast, usage-based theories suggest that language organization is fundamentally shaped by experience, in particular semantic and pragmatic function, as well as usage-frequency (how often language forms are encountered in everyday communication; Bybee, 2010; Bybee & Beckner, 2015; Langacker, 1987a). Lexical frequency predicts cognitive processing demands (Hasher & Zacks, 1984). NBD adults process high-frequency words more accurately and faster than low-frequency items (Balota et al., 2004; Forster & Chambers, 1973; Stemberger & MacWhinney, 1986). Higher lexical frequency has been associated with shorter gaze duration in reading (Ong & Kliegl, 2008). Similar lexical frequency effects are evident in pauses in the speech of both NBD and IWA, with longer pauses before lower frequency forms (Beattie & Butterworth, 1979; Geffen et al., 1979; Mack et al., 2015; Maclay & Osgood, 1959; Pashek & Tompkins, 2002).

While lexical frequency effects are well established, statistical properties of language also manifest in collocation strength between word combinations. Collocation strength refers to the frequency in which words occur together, weighted by the frequency of each word (Gries, 2010; Schneider, 2018). Collocation strength is not merely a function of frequency. For example, in British English, the phrase “it's lovely” has a higher collocation strength than “it's great”, despite “great” being more frequent than “lovely” (BNC, 2007). Collocation strength indicates the degree to which words are associated with another in everyday language use. Zimmerer et al. (2018) found that in semi-structured interviews, IWA (both fluent and non-fluent) produced not only more frequent words, but also more strongly collocated word combinations than non-aphasic speakers with right-hemisphere damage and NBD speakers. The authors employed a software developed for this research, the Frequency in Language Analysis Tool (FLAT; Zimmerer et al., 2016), to extract statistical language features. FLAT determines the usage frequency of every word, bigram (two-word combination) and trigram (three-word combination) from orthographic transcripts based on the spoken sub-corpus of the British National Corpus (BNC, 2007) and calculates collocation strength alongside other measures. A follow-up study replicated the results in a new sample of speakers with non-fluent aphasia (Bruns et al., 2019). Investigating use of the expression “I don't know”, it also demonstrated that IWA use strong collocations in pragmatically appropriate ways. Increased collocation strength in spontaneous speech was also found in three types of primary progressive aphasia, behavioural variant frontotemporal dementia (Zimmerer et al., 2020) and speakers with probable Alzheimer's disease (Zimmerer et al., 2016).

High collocation strength is one indicator of “formulaic language” (Schmitt et al., 2019): phrases and utterances that are processed not only as combinations of individual words, but as one holistic unit (Conklin & Schmitt, 2012; Schmitt, 2012). According to some usage-based frameworks, language formulas are easier to process because they involve words that are strongly associated with each other, and therefore more easily co-activated (Langacker, 1987a, 1987b, 2008). It is possible that some formulas are represented as single lexicalized units, and in this case, they would be easier to process because they require selection of fewer lexical representations and impose reduced combinatorial demands. Collocation strength analysis of connected speech samples of IWA (see above) suggest that such combinations would be more resilient to disruption under conditions of lexical or grammatical impairment.

Despite the established effects of collocation strength on language production, its effects on pauses has not yet been examined. This study, therefore, has two aims:

(1) To see whether previous results showing increased lexical frequency and collocation strength in aphasia (Bruns et al., 2019; Zimmerer et al., 2018, 2020) replicate in a new sample. We hypothesised that (H1) IWA will produce more frequent words and (H2) stronger collocations.
(2) To go beyond measuring the properties of produced linguistic forms and investigate how the variables of frequency and collocation strength relate to pauses in speech and, therefore, to effort in online language processing. Three hypotheses emerged from this aim: (H3) Pause duration will be longer in IWA than in NBD individuals; (H4) Pauses will be shorter and fewer before words with higher lexical frequency, and this effect will be greater in IWA. (H5) Pauses will be shorter and fewer within combinations with greater collocation strength, and this effect will be greater in the IWA.

The novelty of our study relies specially on the last two hypotheses, in which we predicted that increased cognitive demands in the production of less frequent lexical items and weaker collocations would be reflected in the duration of pauses.

We investigated the effects of lexical frequency and collocation strength (as determined by FLAT) on pauses in spontaneous connected speech in IWA and NBD individuals. The participants narrated a (mostly) wordless cartoon. We measured the duration of silent pauses, filled pauses, or combinations of both before each word, and correlated these with the word frequency of the following word and the collocation strength of bigrams. We entered part-of-speech as a covariate.

2. Methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/

exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Pre-registration and post-hoc tests

We pre-registered our analysis on the Open Science Framework website (https://osf.io/8nwe2/) after data collection and transcription, but before pause annotation and data analysis. We report results according to the pre-registered procedure, but because of properties of distributions which we did not consider at the time of pre-registration (zero-inflation in particular), we added a post-hoc test better suited for these (see section 3.1.6.). We also explored the correlation between proxies for aphasia severity and their influence on pause duration (see section 3.1.7.).

2.2. Participants

Data were collected as part of a previous study on grammatical processing in aphasia (Mahmood et al., 2016). The NBD group was recruited from a university register of research volunteers. Ethics approval was granted by the relevant institutional Ethics Committee and all the participants provided informed consent to take part in the study (LC/2013/05). Inclusion criteria were English as the native language (all our participants were British English speakers), presence of aphasia in the IWA group and absence of aphasia in the NBD group. Exclusion criteria were a diagnosis of developmental or other cognitive disorders. The sample consisted of 20 NBD individuals (3 male, 17 female) and 20 IWA (16 male, 4 female). The sample size is consistent with previous work on pauses in aphasia (e.g., Angelopoulou et al., 2018). Table 1 shows sociodemographic information and language assessment results.

Beyond production in discourse (see 2.3 below), we measured production of single words and comprehension of single words and sentences to further profile our speakers. We selected the Boston Naming Test (BNT; Goodglass et al., 2001) picture naming task to test lexical production, and the Comprehensive Aphasia Test (CAT; Swinburn et al., 2008) word-picture-matching and sentence-picture matching subtests for testing comprehension. Because we did not use the entire CAT, this study does not have a single standardised measure of aphasia severity. Thus, to examine the relationship between discourse and other measures, our correlation tests used BNT scores and a composite (mean) of CAT subtests.

2.3. Test procedure

Participants met with a researcher in a quiet room. Speech data were elicited using "The dinner party" cartoon (Fletcher & Birt, 1983). In this task, participants described an 8-picture story that contained no dialogue or narration. The instructions were: "Look at these pictures. Together, they make a story. Could you tell me in your own words everything you see going on in the pictures". Speech was digitally recorded and orthographically transcribed using F4transkript (Jones & German, 2016).

2.4. Annotation procedure

The audio of each sample was loaded into ELAN Linguistic Annotator V 6.0 (Max Planck Institute for Psycholinguistics, 2020). We counted words of each sample using the R package psych V.1.9 (Revelle, 2022). After segmenting the words in the audio sample, the transcriptions were aligned to their respective segment in an annotation tier. Pauses were defined as any speech hesitation, which could be silent, filled with interjections, or containing both. Duration and location of pauses were identified based on visual inspection of the spectogram and on acoustic inspection, first at 100% and subsequently at 30% playback rates. We measured word onsets and offsets and defined pauses as the time (measured in msec; ms) between one word's offset and the following word's onset. We combined filled and silent pauses in our analysis, as both can reflect processing demands in speech.

We created an additional tier categorising the pauses: 1. "No pause", where the distance between words was between 0 and 250 ms. This rationale takes into account the observations of Goldman-Eisler (1968) that hesitations of this duration are related to articulatory processing; 2. "Silent pause", annotated as a hesitation greater than 250 ms with no acoustic signal between words; 3. "Filled pause", annotated as hesitation longer than 250 ms between words marked by filler words (interjections such as "uhm" or "uh"); 4. "Filled/Silent pause",

Table 1 – NBD individuals and IWA sociodemographic data and IWA language assessment.

	NBD individuals (SD)	IWA (SD)	
Sociodemographic variables			Mean differences
Age	67.45 (8.02)	58.85 (9.93)*	$t(38) = 2.64, p = .01$
Years of education	16.8 (2.75)	15.55 (3.03)	$t(38) = 1.42, p = .17$
Stroke post-onset time		5.33 years (2.36)	
Language assessment variables			
BNT (out of 60)	38.5 (16.01)		
CAT comprehension of spoken words (out of 30)	27.2 (1.88)		
CAT comprehension of written words (out of 30)	27.9 (2.29)		
CAT comprehension of spoken sentences (out of 32)	26.75 (5.55)		
CAT comprehension of written sentences (out of 32)	26.65 (4.63)		
CAT composite score (mean of all subtests)	27.41 (2.57)		

Group comparisons were carried out using t-tests (* = significantly different at $p < .05$). BNT = Boston Naming Test, used to assess lexical retrieval; CAT = Comprehensive Aphasia Test, subtests used to determine extent of lexical and grammatical impairment. Language assessment showed that IWA group had mild to moderate language impairment.

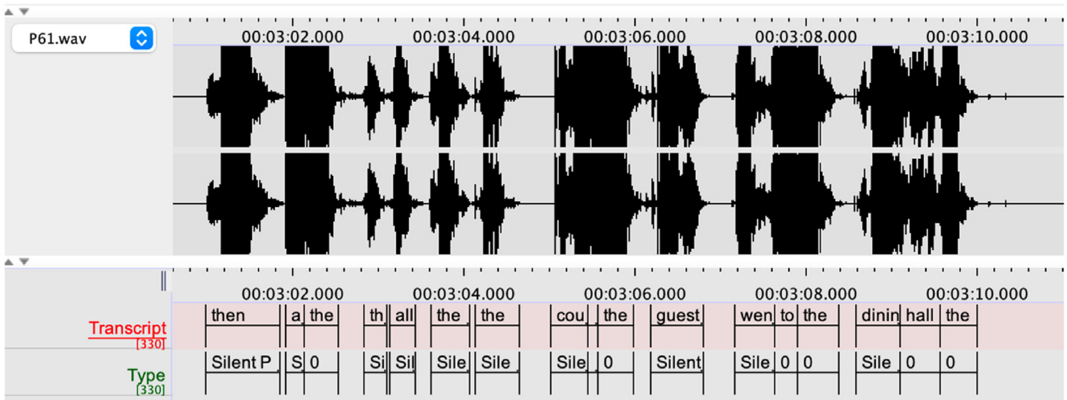

Fig. 1 – Example of annotation in the ELAN software, using two tiers: "Type" classifies the type of pause (silent pause, filled pause, filled-silent/silent-filled pause or no pause), and "Transcript" displays the transcription of each narration.

annotated as a filled pause followed by a silent pause or vice versa. As silent and filled pauses both reflect processing demands in language production, we considered both in the statistical analysis. Fig. 1 shows the ELAN annotator interface.

Lexical frequency and collocation strength were determined using the FLAT software (Zimmerer et al., 2016, 2018). Lexical frequency was measured as the frequency of occurrence of a word in the spoken sub-corpus of the BNC (2007), in occurrences per million words. Collocation strength of each bigram was measured using t-scores (Gablasova et al., 2017; Gries, 2010), which are based on the raw frequency of each bigram and the frequency of its individual words. The formula is:

$$tscore_{ab} = \frac{frequency_{ab} - expected\ frequency_{ab}}{\sqrt{frequency_{ab}}}$$

where $frequency_{ab}$ is the observed frequency of the bigram in the spoken BNC and expected $frequency_{ab}$ is the frequency of the bigram if the order of the words was random.

Following the procedure outlined in previous studies using FLAT (Bruns et al., 2019; Zimmerer et al., 2018, 2020), bigrams were excluded from the collocation strength analysis if they contained proper nouns or pseudowords. Lexical repetitions were removed so that the analysis only considered one instance of the repeated word, unless raters considered repetitions intentional. Ungrammatical combinations (e.g., "man think") were also excluded, since these would be classified as rare not because of greater language capacity, but because of failed combinatorial operations. In bigrams with a filled pause (e.g., uhm or uh), the interjection was removed, and the collocation strength between the words preceding and following the pause was measured (note that these interjections were considered for pause duration analysis). FLAT also automatically excludes combinations which cross sentence boundaries (marked with sentence-final punctuation) or utterance boundaries (annotated via line break). The total number of bigrams removed was 159 (4.3% of all bigrams); of those 159 bigrams removed, 124 were produced by IWA and 35 by the NBD group. These numbers primarily reflect that IWA produced more grammatical errors and less connected speech.

We annotated PoS for each word using the R package "Spacyr" V.1.2.1 (Benoit & Matsuo, 2022), a wrapper to the Python "spaCy" library. Spacyr carries out a morphosyntactic analysis of each word in a transcript. Categories were grouped into two-factor levels: content words (nouns, verbs, adjectives, and adverbs) and function words (pronouns, prepositions, conjunctions, determiners, and interrogatives). Table 2 shows an excerpt from raw data.

3. Results

3.1. *Statistical analysis procedure*

3.1.1. *Outlier detection and log-transformation*

To identify outliers for lexical frequency and collocation strength values, we used Grubbs' test. This analysis was applied to the data after the removal of ungrammatical bigrams. There was no evidence of outliers for lexical frequency in the NBD group, $G(1.75) = .99$, $p = 1$, and IWA groups, $G(1.56) = .99$, $p = 1$. For collocation strength, the NBD group showed outliers with a t-score below −43.79 ($N = 18$ out 2068; .87%) and above 125.87 ($N = 58$ out 2068; 2.81%), $G(6.59) = .97$, $p < .001$. In the IWA group, outliers were found below −55.64

Table 2 – Example of raw data for the utterance "they have a fish", generated from ELAN annotation (pause type and pause duration), Spacyr (PoS) and FLAT (lexical frequency, bigram and collocation strength) analysis.

Participant ID	Group	Word	PoS	Lexical frequency (per million words)	Pause type	Pause duration	Bigram	Collocation strength
1	NBD individuals	they	Function word	9656	No pause	0 ms	and they	62.8
1	NBD individuals	have	Content word	7727	No pause	0 ms	they have	41.6
1	NBD individuals	a	Function word	20620	Silent pause	100 ms	have a	81.1
1	NBD individuals	fish	Content word	749	No pause	0 ms	a fish	5.83

($N = 9$ out 1475; .61%) and above 157.29 ($N = 13$ out of 1475; .88%), $G(5.02) = .97$, $p < .001$. These values demonstrate greater variance of collocation strength within the NBD group. Outlier values were removed from the statistical analysis.

Since we entered every word and bigram that met our selection criteria into our models, we included pause duration values which were 0 ms (i.e. no pause). We subjected pause duration values to log transformation, with a bin size of .1 log units (Hird & Kirsner, 2010). After log transformation, the values for pause duration ranged from zero to five units. After transformation, the distribution was zero-inflated; it had a mode at zero, followed by a normal distribution curve.

We refitted lexical frequency values into the same pause duration log-transformed range to analyse and converge the data on comparable scales. The dataset for lexical frequency was rescaled with minimum and maximum (min = 0, max = 5). As collocation strength values converged in narrowed scales, we did not apply refitting on this dataset.

3.1.2. *Group comparisons*

Table 3 displays the results for each group and comparisons between groups. We carried out comparisons of pause duration, word count, lexical frequency and collocation strength using Wilcoxon signed-rank tests. We used Chi-square tests for proportions of content vs. function words. While all participants made filled and silent pauses, the proportion of filled pauses was lower than that of silent pauses. The IWA group produced significantly fewer words, more pauses, and longer pauses, with higher proportions of silent and filled + silent pauses. The lexical frequency of content, function words and collocation strength values were higher for the IWA group. There were no significant differences between proportions of content and function words between groups.

3.1.3. *Linear mixed models*

To examine the effects of lexical frequency and collocation strength on pauses, we fitted linear mixed models (LMM) using the lme4 package (Bates et al., 2015) for R (R Core Team, 2020). Independent models were considered for Lexical Frequency and Collocation Strength analyses since Lexical Frequency and Collocation Strength are strongly correlated variables. For these models, we excluded words which were the first in an utterance to avoid confounding effects of utterance planning. Within each set, we determined the best model by starting with the simplest and adding interactions only when they resulted in a significantly better model, as determined by a likelihood ratio test.

Table 4 summarizes all LMMs. In the reference models of each set, the predictors were Group (NBD or IWA) and PoS, while Lexical Frequency and Collocation Strength (respectively) were placed as random factors. In Model 1, the predictors of Pause Duration were Lexical Frequency (or Collocation Strength), Group and PoS. Model 2 tested the interaction between Lexical Frequency or Collocation Strength with Group. Model 3 tested the interaction between PoS and Group. Finally, Model 4 examined the three-way interaction between all predictors. Note that Models 1–4 had a simpler random effect structure as we were worried about convergence. As the result, estimates are anti-conservative.

3.1.4. *Effects of lexical frequency on pause duration*

Full results for each Lexical Frequency model are reported in Supplementary file. The Reference model revealed a significant effect of Group ($\beta = .46$, $t = 5.01$, $p < .001$), as IWA made longer pauses, but no significant effect of PoS ($\beta = -.01$, $t = -.43$, $p = .67$). Model 1 was not significantly better ($X^2(1) = 0$, $p = 1$), but did show a significant effect of Lexical Frequency ($\beta = -.02$, $t = -2.12$, $p = .034$), in addition to the significant effect of Group ($\beta = .49$, $t = 5.20$, $p < .001$) (Table 5). The

Table 3 – Means for pause duration, type of pauses, PoS distributions, lexical frequency and collocation strength values by group.

Variable	NBD individuals			IWA			Mean differences
	Mean	SD	Median (IQR)	Mean	SD	Median (IQR)	
Pauses							
Pause Duration between words (ms)	213.2	88.0	<.1 (203)	1110.5	983.7	75.0 (907.5)	$Z = -13.66$, $p < .001$, $r = -.24$
Silent Pauses (%)	71.5	30.2	26.0 (53.7)	79.6	53.6	57.0 (36.7)	$Z = -13.67$, $p < .001$, $r = .83$
Filled Pauses (%)	1.4	2.6	0 (1.2)	1.9	3.3	0 (1)	$Z = -.66$, $p = .25$, $r = -.10$
Filled + Silent Pause (%)	3.7	4.6	4 (16.2)	14.7	17.8	17 (26.7)	$Z = -2.76$, $p = .002$, $r = -.43$
No pauses between words (%)	23.35	65.4	6 (3.2)	3.80	53.1	2 (10)	$Z = -1.99$, $p = .02$, $r = -.31$
PoS							
Word count	2093	78.50	95 (110)	1587	71.1	75 (96.5)	$Z = -7.45$, $p < .001$, $r = .64$
Content words (%)	13.86	7.80	13 (109)	13.49	35.4	8 (31)	X^2 (1, N = 40) = .10, $p = .07$
Function words (%)	86.14	37.4	78.5 (109.5)	86.51	45.6	61 (60.7)	
Lexical Frequency and Collocation Strength							
Mean Lexical Frequency (per million words)	16594	1839	10510 (22535)	18168	3167	17454 (21247)	$Z = -4.97$, $p < .001$, $r = .48$
Lexical Frequency (content words)	2653	6578	749 (161)	3181	7213	1011 (104)	$Z = -6.20$, $p < .001$, $r = .32$
Lexical Frequency (function words)	18390	14055	1754 (2132)	20642	14222	2336 (1930)	$Z = -5.01$, $p = .001$, $r = .09$
Collocation Strength	22.97	4.50	3.6 (32.2)	25.35	10.11	2.1 (40.8)	$Z = -1.70$, $p = .02$, $r = .18$

The section on pauses contains proportions of filent pauses, filled pauses, filled + silent pauses and no pauses between words. The median pause duration between words for NBD individuals is close to zero because most times, they made no pause between words, resulting in zero-inflated distributions (see 2.1 and 3.1.6). IQR = Interquartile range. Pauses, PoS and lexical frequency were calculated before removal of ungrammatical bigrams and lexical frequency and collocation strength outliers. Mean and median lexical frequency was calculated after removal of outlier values. Mean and median collocation strength (t-scores) was calculated after removal of ungrammatical bigrams and outlier values.

Table 4 – LMMs investigating predictors for pause duration. Because of different dataset sizes and strong correlations between predictor variables, different analyses were applied to Lexical Frequency and Collocation Strength models.

	Lexical Frequency	Collocation Strength
Models without interactions		
Reference model	Pause Duration ~ Group + PoS + (1 + Lexical Frequency| Speaker)	Pause Duration ~ Group + PoS + (1 + Collocation Strength| Speaker)
Model 1	Pause Duration ~ Lexical Frequency + Group + PoS + (1 | Speaker)	Pause Duration ~ Collocation Strength + Group + PoS + (1 | Speaker)
Models with interactions		
Model 2	Pause Duration ~ Lexical Frequency * Group + PoS + (1 | Speaker)	Pause Duration ~ Collocation Strength * Group + PoS + (1 | Speaker)
Model 3	Pause Duration ~ Lexical Frequency + Group * PoS + (1 | Speaker)	Pause Duration ~ Collocation Strength + Group * PoS + (1 | Speaker)
Model 4	Pause Duration ~ Lexical Frequency*Group*PoS + (1 | Speaker)	Pause Duration ~ Collocation Strength*Group*PoS + (1 | Speaker)

strongest model was Model 4, which included all two-way interactions and the three-way interaction, and was significantly stronger than the best model with a single two-way interaction ($X^2(3) = 14.20$, $p = .003$). That model showed significant effects of Group, Lexical Frequency, and significant interactions between Lexical Frequency and PoS ($\beta = .26$, $t = 2.36$, $p = .019$), as frequency effects were stronger for content words, and Group and PoS ($\beta = .16$, $t = 2.01$, $p = .045$), since pauses were longer before content words, and that effect was stronger for PwA (Table 5). With regards to variance explained, models were weak; in Model 5, fixed effects explained 5% of the variance.

3.1.5. *Effects of collocation strength on pause duration*

We report full results for each Collocation Strength model in Supplementary file. The Reference model revealed a significant effect of Group ($\beta = .45$, $t = 5.51$, $p < .001$), as IWA had longer pauses than the NBD group. Model 1 was not significantly better than the Reference model ($X^2(1) = 0$, $p = 1$). Once interactions were added, the main effect of Group remained significant, and interactions between Group and other predictors were also significant (Group: $\beta = .56$, $t = 5.80$, $p < .001$; Group*Collocation Strength: $\beta = -.003$, $t = -2.22$, $p = .001$). Analysis using a likelihood ratio test found that Model 2, which contains an interaction between Collocation Strength and Group, had significantly greater explanatory power than simpler models ($X^2(1) = 10.26$, $p = .001$; Table 6). In Model 2, the interaction between Collocation Strength and Group was statistically significant ($R^2 = .14$, $\beta = -.003$, $t = -2.22$, $p = .001$), as the effect of Collocation strength on pauses was stronger in IWA (Fig. 2). More complex models did not significantly improve explanatory power.

3.1.6. *Hurdle models*

Because of the zero-inflated distribution of pause durations, we followed the pre-registered analysis with Bayesian hurdle-Gaussian models (Heilbron, 1994) to corroborate LMM results. Hurdle-Gaussian models consist of two steps: the first aims to explain the zero values, and the second the positive values. We used hurdle models to further scrutinise LMM results, in particular Model 2 for Collocation Strength. See Supplementary file for results.

Results from hurdle models investigating lexical frequency only partially supported linear model results. Model 1, which included all fixed variables but no interactions, showed a significant effects of Group ($\beta = .46$, $p < .001$) and Lexical Frequency ($\beta = -.02$, $p = .046$) on pauses. However, a hurdle model with all interactions (Model 5) did not converge well and only determined significant main effects for Group and Lexical Frequency, while interactions were not significant. Significant effects only explained the distribution of non-zero values.

For collocation strength, results were fully corroborated by hurdle models. There was moderate evidence of a Group effect, which was statistically significant ($\beta = .56$, $p < .001$). The two-way interaction effect between Collocation Strength and Group was moderate and statistically significant ($\beta = -.004$, $p < .001$). Pauses had a moderate probability of being shorter within bigrams with higher Collocation Strength, and the effect was stronger in IWA. Effects were only significant for explaining non-zero values.

Table 5 – Best model for Pause Duration including Lexical Frequency. Model 5 is characterized by significant main effects of Lexical Frequency, Group, and significant interactions involving PoS.

Model 5	Pause Duration ~ Lexical Frequency * Group * PoS + (1 | Speaker)			
Conditional R^2	.13			
Marginal R^2	.06			
Random effects	Variance		SD	
Speaker (intercept)	.07		.92	
Residuals	.27		.96	
Fixed effects	β coefficient (estimate)	t-value	95%CI	Significance (p-value)
Intercept	.88	13.01	.75, 1.01	<.001
Lexical Frequency	−0.26	−2.33	−.48, −.04	.02
Group	0.48	4.85	.29, 0.68	<.001
PoS	−0.06	−1.30	−.16, .03	.19
Lexical Frequency*Group	−0.08	−.45	−.44, .27	.65
Lexical Frequency*PoS	0.26	2.36	.04, .48	.019
Group*PoS	0.16	2.01	−003, .32	.045
Lexical Frequency*Group*PoS	0.02	.12	−.34, .38	.91

Table 6 – Best model for Pause Duration including Collocation Strength. Model 2 is characterized by significant interactions between Group and Collocation Strength, and the main effect of Group.

	Model 2 structure: Pause Duration ~ Collocation Strength * Group + PoS + (1 | Speaker)		
Conditional R^2	.14		
Marginal R^2	.08		
Random effects	Variance	SD	
Speaker(intercept)	.07	.26	
Residuals	.96	.98	
Fixed effects	β coefficient (estimate)	t-value 95%CI	p-value
Intercept	.25	1.56−.06, .56	.11
Collocation Strength	.0002	−.15−.0037, −.0031	.88
Group	.56	5.7 .37, .75	<.001
PoS	−.06	1.04−.05, .17	.29
Group*Collocation Strength	−.003	−3.19. −.005, −.001	.001

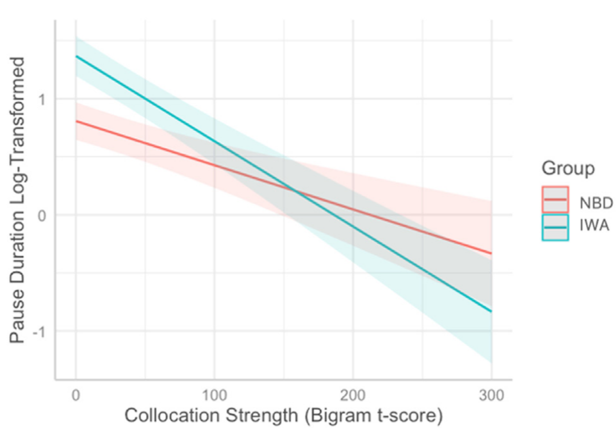

Fig. 2 – Predicted values of the interaction between Collocation Strength and Pause Duration for the best fitting model (Model 2). For both groups, Pause Duration was shorter within bigrams with high Collocation Strength values. The effect is greater for IWA.

3.1.7. *Effects of aphasia severity on pause duration*

The duration of pauses in speech may vary with the severity of aphasia (Goodglass et al., 1964). We used performance on the BNT and a composite score based on CAT comprehension subtests as proxies for aphasia severity (see 2.2.). Both were significantly and strongly correlated, $r(18) = .71$, $p < .001$. We applied separate LMMs to explore the influence of BNT and CAT comprehension composite scores on pause duration.

For lexical frequency, we fitted the baseline model Pause Duration ~ Lexical Frequency * PoS + (1|Speaker). The baseline model showed a significant positive effect of PoS ($\beta = .01$, $t = 1.97$, $p = .02$). A likelihood ratio test revealed that adding BNT scores (X^2 (1) = .07, $p = .7$) or CAT composite scores (X^2 (1) = .15, $p = .6$) did not increase the explanatory power of the baseline model.

For collocation strength, we fitted the baseline model Pause Duration ~ Collocation Strength * PoS + (1|Speaker). The baseline model showed a significant negative effect of collocation strength ($\beta = -.01$, $t = -5.28$, $p < .001$) and a significant positive interaction between collocation strength and PoS ($\beta = .004$, $t = 2.38$, $p = .01$). A likelihood ratio test revealed that adding BNT scores (X^2 (1) = .28, $p = .5$) or CAT composite scores (X^2 (1) = .12, $p = .7$) did not increase the explanatory power.

4. Discussion

Prior studies have shown that IWA produce words with higher lexical frequency and word combinations which are more strongly collocated (Bruns et al., 2019; Kittredge et al., 2008; Zimmerer et al., 2018, 2020). While these studies suggested more familiar language is produced with less effort, our study of speech pause data as indicators of effort provide clear evidence that this assumption is likely correct.

Pauses in speech were extracted from “The dinner party” comic strip narrations produced by a group of individuals with mild to moderate aphasia and NBD controls. Samples of speech from IWA contained more and longer pauses, replicating previous findings (Angelopoulou et al., 2018; Perkins, 1995). Both groups produced more silent than filled pauses, but IWA showed a higher proportion of silences. Types of pauses may indicate efforts at different processing stages. The elicitation task requires constructing a plausible narration from a picture sequence and involves integrative processing across language and other cognitive networks, such as event processing. Interpretation of complex visual events may be dissociated from the linguistic description of such events (Brown et al., 2020; Fedorenko & Blank, 2020; Fedorenko & Varley, 2016; Ivanova et al., 2021). Our findings may reflect this dissociation, since it is assumed that filled pauses demonstrate higher level picture/event conceptualisation, whereas silent pauses indicate the cognitive demands of lexical searching or phrase construction (Angelopoulou et al., 2018; Butterworth, 1976, 1979). The higher proportion and duration of pauses in IWA may reflect the greater challenge of language production, demonstrating increased cognitive and linguistic demands to construct utterances.

IWA produced content words with higher lexical frequency, which is in concordance with previous research (Beattie & Butterworth, 1979; Geffen et al., 1979; Zimmerer et al., 2018). Furthermore, pauses were shorter before more frequent words, replicating previous findings (Goral et al., 2010). Note that LMM analysis, but not the post-hoc hurdle models, found that the effect of frequency on pauses was greater for IWA than NBD speakers, meaning that this particular interaction was not robust. Previously, high-frequency words are proposed to be easier to access than low-frequency words due to a higher resting state of underpinning neural networks (McClelland & Elman, 1986), but more recent studies using model simulations have proposed that more frequent words are more strongly weighted in the lexical network (Nozari et al., 2010). Thus, high-frequency words are more resilient to damage to lexical networks. Future research can integrate other variables that relate to frequency, such as neighbourhood density, age of acquisition, or grammatical category (Baayen et al., 2016), exploring how they interact, and which factors are stronger determinants of pause duration.

In accordance with the results of Zimmerer et al. (2018), IWA produced bigrams with a higher collocation strength than NBD controls. A novel finding was the effect of bigram collocation strength on pauses in speech, with shorter pauses within stronger collocated bigrams. The effect was greater in IWA. These results may reflect how residual language in IWA is expressed in stronger collocated word combinations, which help decrease processing demands in connected speech, as evidenced by shorter pauses. According to usage-based theories (Bybee & Beckner, 2015; Christiansen & Chater, 2018; Goldberg, 2003; Langacker, 1987a), strongly collocated word combinations may be processed holistically, either via strengthened connections between individual words or as single morpheme-like units (Siyanova-Chanturia et al., 2017; Tremblay & Baayen, 2010). These theories predict greater ease of processing stronger collocations and are, therefore, supported by our study.

However, many neurolinguistic models are based on the distinction between lexicon and grammar. For instance, Friederici (2011) claimed that syntactic processes are strongly supported by frontal left-hemisphere areas, whereas processing of lexical-semantic information is based on fronto-temporal areas, assuming distinctively different neural and cognitive processing for both functions. Hagoort (2013) makes a distinction between left frontal areas which support combination of linguistic units (e.g., words), and left temporal areas which support storage of these units. We need to consider whether usage properties such as collocation strength can be integrated into these models, or whether they call for different models entirely. For example, Van Lancker Sidtis (2012) observed a decrease in familiar language use in individuals with right hemisphere damage or Parkinson's disease, suggesting that the representation of familiar language can be supported by the right hemisphere as well as subcortical nuclei. This view is supported by some other evidence: Skipper et al. (2022) asked NBD individuals to repeat a number of sentences over 15 days. Subsequently, the participants were asked to listen to the learned sentences and novel ones during fMRI scanning. The repeated sentences, compared to novel sentences, elicited stronger activation of the bilateral sensorimotor areas and the right hemisphere frontal gyrus. Further evidence has been provided by studies employing event-related potentials (ERP). Siyanova-Chanturia et al. (2017) compared strong collocations (e.g., “knife and fork”) with rarer combinations of semantically related nouns (e.g., “spoon and fork”). A larger right-lateralised P300, followed by a smaller N400, was elicited in response to the usual collocations. Here, the larger P300 effect is based on high predictability, such as within multiword expressions, whereas the smaller N400 effect reflects easier semantic integration (see also Vespignani et al., 2010). We believe that these observed differences in neurological processing of familiar language are related to our behavioural finding that stronger collocations involve fewer and shorter speech pauses.

While results based on LMMs (which we pre-registered) and post-hoc hurdle-Gaussian models were largely similar, we did observe discrepancies. In LMMs, the interaction between lexical frequency and group revealed a significant effect on pauses. In hurdle models, it was not significant. Furthermore, the interaction between collocation strength and group was strong in LMMs, but moderate in hurdle models. While LMMs can be robust even when distributions are not parametric, hurdle models are better suited. Nevertheless, use of both model types in one study is not elegant. Future studies with similar designs should provide clarity, and we advise focusing on hurdle models.

To our surprise, aphasia severity did not have a significant effect when added to models which only included IWA. However, we would advise against the conclusion that severity is not associated with pause duration or the effects of lexical frequency or collocation strength. As there was an effect of aphasia in group comparisons, it would be reasonable to expect an effect of severity. We note that our sample of IWA represented the mild-to-moderate part of the aphasia severity scale, and it is possible that a more heterogeneous sample could detect such an effect.

Finally, and related to clinical practice, many aphasia assessments and therapy protocols focus on single-word properties (Bruehl et al., 2023). Clinical research could consider word collocations both in assessment and therapy. For instance, Melodic Intonation Therapy employs high frequency formulas, based on the observation of production of song lyrics even in severe aphasia, with this resilience stemming from their likely storage in a holistic manner (Stahl & Kotz, 2014). Unification Therapy Integrating Lexicon and Sentences (UTILISE) includes high-frequency constructions as early training items and introduces variations later (Varley et al., 2020). For example, the construction *I made it* (PERSON made THING) can be systematically loosened and lengthened with new lexical items inserted into the PERSON (*He made it*), THING (*I made coffee*) slots, or an adjunct added (*She made coffee today*). Finally, the availability of new technologies allows clinicians to efficiently and accurately obtain the statistical properties of language and/or duration of pauses. In general, aphasia assessment includes measures that might reflect fluency in language production (e.g., mean length of utterance) but do not allow naturalistic labelling of fluency, such as pause duration. Pauses in speech, given their likely sensitivity to collocation strength, could be considered an outcome measure in interventions.

5. Limitations

Analysis of combinations was restricted to bigrams. However, collocation strength within larger units may contribute further insights. Another limitation is that we did not address the impact of other linguistic phenomena on pauses. Our model does not consider the type of syntactic structure, specific parts of speech beyond the dichotomy of content and function words, or other usage variables such as age of acquisition of individual words and phrases. The effect of age of acquisition has been established at the lexical level but is underexplored at the level of word combinations (Arnon et al., 2017). Also, while it is common to consider frequency effects at the word level, frequency and co-occurrence effects may manifest at other levels, such as phonemes and morphemes. Further studies might compare the collocation strength of sub-lexical units to bigrams or multiword utterances and their influence on cognitive effort in speech. Future research could add such variables; however, more complex models require larger data sets and face the great challenge of frequency-based variables often being strongly correlated. A further limitation concerns the combination of filled and silent pauses, which could reflect different cognitive processes. In our study, we combined these to not further complicate models, and after analysis, we understand that our data would not be suitable for such a comparison because of the relatively small number of filled pauses. However, pause type may interact with our predictors.

When registering the study, we were worried about convergence and chose a less complex structure for our models' random effects. While this is a valid decision, it made models less conservative. Because we already added hurdle models post-hoc, we chose not to complicate our report by exploring model variations with different random effect structures. While we assume that the stronger effects reported here would survive more conservative models, replication using different models would help further examination.

The final limitation concerns our removal of bigrams which do not occur in our reference corpus or are ungrammatical (or both). The removal of ungrammatical utterances, while necessary when using our frequentist methods in natural speech, might distort evaluation of formulaic output in language production. While ungrammatical combinations are less likely to be formulaic and may not even involve higher-level speech planning in NBD (Ramanarayanan et al., 2009), it is possible that ungrammatical formulas are developed in IWA as compensation for production difficulties. Repetition or pragmatically specific use of an ungrammatical expression could result in its holistic representation. This possibility is currently underexplored and could be addressed by future research.

While important questions about pausing behaviour in language production remain, our study demonstrates that the inclusion of usage-frequency variables, including collocation strength at a multiword level, can help understand which aspects of language forms affect processing demands.

Author note

We mourn the tragic death of our dear colleague Sebastian Bello-Lepe, first author of this article, on December 25th, 2023, following a senseless act of violence. He was a dedicated researcher and speech and language therapist, and a beloved colleague to many. We will miss his energy, optimism, and humor. Our deepest condolences go out to his family, his friends, and his partner.

Sebastian Bello-Lepe died after this manuscript was accepted.

Funding

The Chilean National Agency for Research and Development (ANID) funded the Ph.D studies of Sebastian Bello-Lepe (72190125).

Data/code availability

The conditions of our ethics approval, particularly those concerning patient confidentiality, do not permit public archiving of the original audio data. Other data and scripts for statistical analysis can be found along with the pre-registration of this study: https://osf.io/8nwe2/. Access will be granted with no conditions. For any questions about the datasets and material, please contact the corresponding author. ELAN is available at https://archive.mpi.nl/tla/elan. The FLAT is available here: https://osf.io/v8mg9/. Legal copyright restrictions prevent public availability of BNT and TROG test materials.

Open practices

The study in this article has earned Preregistered badges for transparent practices. The preregistered studies are available at: https://osf.io/v8mg9/.

CRediT authorship contribution statement

Sebastian Bello-Lepe: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Sabrina Mahmood:** Investigation, Data curation. **Rosemary Varley:** Writing – review & editing, Supervision, Project administration. **Vitor Zimmerer:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Acknowledgments

We thank all the participants for their willingness to take part in the study. We also thank the reviewers for their valuable contributions to this report.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2024.06.012.

REFERENCES

Angelopoulou, G., Kasselimis, D., Makrydakis, G., Varkanitsa, M., Roussos, P., Goutsos, D., Evdokimidis, I., & Potagas, C. (2018). Silent pauses in aphasia. *Neuropsychologia, 114*, 41–49. https://doi.org/10.1016/j.neuropsychologia.2018.04.006

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

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

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech, 22*(3), 201–211. https://doi.org/10.1177/002383097902200301

Benoit, K., & Matsuo, A. (2022). *Spacyr: Wrapper to the 'spaCy' 'NLP' Library. R package version 1.2.1*. https://spacyr.quanteda.io.

BNC Consortium. (2007). *British national corpus, XML edition, oxford text archive*. http://hdl.handle.net/20.500.12024/2554.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. https://doi.org/10.48550/ARXIV.2005.14165.

Bruehl, S., Willmes, K., & Binkofski, F. (2023). Interfered-naming therapy for aphasia (INTA): A neuroscience-based approach to improve linguistic-executive processing. *Aphasiology, 37*(2), 205–226. https://doi.org/10.1080/02687038.2021.1994917

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

Butterworth, B. (1976). *Semantic planning, lexical choice and syntactic organization in spontaneous speech*.

Butterworth, B. (1979). Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain and Language, 8*(2), 133–161. https://doi.org/10.1016/0093-934X(79)90046-4

Bybee, J. (2010). *Language, usage and cognition* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511750526

Bybee, J. L., & Beckner, C. (2015). Usage-based theory. In B. Heine, H. Narrog, B. Heine, & H. Narrog (Eds.), *The oxford handbook of linguistic analysis*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199677078.013.0032.

Christiansen, M. H., & Chater, N. (2018). *Creating language integrating evolution, acquisition, and processing*. The MIT Press.

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

DeDe, G., & Salis, C. (2020). Temporal and episodic analyses of the story of cinderella in latent aphasia. *American Journal of Speech-Language Pathology, 29*(1S), 449–462. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0210

Fedorenko, E., & Blank, I. A. (2020). Broca's area is not a natural kind. *Trends in Cognitive Sciences, 24*(4), 270–284. https://doi.org/10.1016/j.tics.2020.01.001

Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients: Language versus thought. *Annals of the New York Academy of Sciences, 1369*(1), 132–153. https://doi.org/10.1111/nyas.13046

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210–233. https://doi.org/10.1016/0749-596X(91)90004-4

Fletcher, M. H. W., & Birt, D. (1983). *Storylines—picture sequences for language practice*. Longman.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological*

Reviews, 91(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence: Collocations in corpus-based language learning research. *Language Learning, 67*(S1), 155–179. https://doi.org/10.1111/lang.12225

Geffen, G., Stierman, I., & Tildesley, P. (1979). The effect of word length and frequency on articulation and pausing during delayed auditory feedback. *Language and Speech, 22*(2), 191–199. https://doi.org/10.1177/002383097902200208

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

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech *(2. Print)*. Acad. Press.

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

Goodglass, H., Quadfasel, F. A., & Timberlake, W. H. (1964). Phrase length and the type and severity of aphasia. *Cortex, 1*(2), 133–153. https://doi.org/10.1016/S0010-9452(64)80018-6

Goral, M., Levy, E., Swann-Sternberg, T., & Obler, L. (2010). Frequency and word-length factors and lexical retrieval in sentence production in aphasia. *Procedia - Social and Behavioral Sciences, 6*, 107–108. https://doi.org/10.1016/j.sbspro.2010.08.054

Gries, S. T. (2010). Useful statistics for corpus linguistics. In *A mosaic of corpus linguistics: Selected approaches* (3631587899th ed., pp. 269–291) (Peter Lang AG).

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11*(1), 58–81. https://doi.org/10.1016/0010-0285(79)90004-5

Hagoort, P. (2013). MUC (Memory, Unification, control) and beyond. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00416

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal, 36*(5), 531–547. https://doi.org/10.1002/bimj.4710360505

Hird, K., & Kirsner, K. (2010). Objective measurement of fluency in natural language production: A dynamic systems approach. *Journal of Neurolinguistics, 23*(5), 518–530. https://doi.org/10.1016/j.jneuroling.2010.03.001

Ivanova, A. A., Mineroff, Z., Zimmerer, V., Kanwisher, N., Varley, R., & Fedorenko, E. (2021). The Language network is recruited but not required for nonverbal event semantics. *Neurobiology of Language, 2*(2), 176–201. https://doi.org/10.1162/nol_a_00030

Jones, C., & German, A. (2016). F4transkript, a simple interface for efficient annotation. *Language Documentation And Conservation, 10*, 347–355, 10.125/24701/1/jones.

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

Klatt, H. (1980). Pauses as indicators of cognitive functioning in aphasia. In H. W. Dechert, & M. Raupach (Eds.), *Temporal variables in speech* (pp. 113–120). DE GRUYTER MOUTON. https://doi.org/10.1515/9783110816570.113.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics, 35*(2), 162–179. https://doi.org/10.1016/j.wocn.2006.04.001

Krivokapić, J., Styler, W., & Byrd, D. (2022). The role of speech planning in the articulation of pauses. *The Journal of the Acoustical Society of America, 151*(1), 402–413. https://doi.org/10.1121/10.0009279

Langacker, R. W. (1987a). *Foundations of cognitive grammar. Vol. 1: Theoretical prerequisites (Nachdr.* Stanford Univ. Press, 1.

Langacker, R. W. (1987b). *Foundations of cognitive grammar: Volume II: Descriptive application*. Stanford University Press.

Langacker, R. (2008). *Cognitive grammar: A basic introduction*. York: Oxford University PressNew. https://doi.org/10.1093/acprof:oso/9780195331967.001.0001

Mack, J. E., Chandler, S. D., Meltzer-Asscher, A., Rogalski, E., Weintraub, S., Mesulam, M.-M., & Thompson, C. K. (2015). What do pauses in narrative production reveal about the nature of word retrieval deficits in PPA? *Neuropsychologia, 77*, 211–222. https://doi.org/10.1016/j.neuropsychologia.2015.08.019

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *WORD, 15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

Mahmood, S., Zimmerer, V. C., & Varley, R. A. (2016, September 25-30). *Grammaticality Judgements in Aphasia and Parkinson's disease* [Poster presentation]. In *17th International Science of aphasia Conference*. Venice, Italy https://sstp.nl/article/view/22696.

Max Planck Institute for Psycholinguistics. (2020). ELAN (Version 6.0) the language archive. https://archive.mpi.nl/tla/elan.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language, 63*(4), 541–559. https://doi.org/10.1016/j.jml.2010.08.001

Ong, J. K. Y., & Kliegl, R. (2008). Conditional co-occurrence probability acts like frequency in predicting fixation durations. *Journal of Eye Movement Research, 2*(1). https://doi.org/10.16910/jemr.2.1.3

Pashek, G. V., & Tompkins, C. A. (2002). Context and word class influences on lexical retrieval in aphasia. *Aphasiology, 16*(3), 261–286. https://doi.org/10.1080/02687040143000573

Perkins, L. (1995). An exploration of the impact of psycholinguistic impairments on conversational ability in aphasia. *International Journal of Psycholinguistics, 11*, 167–188.

Potagas, C., Nikitopoulou, Z., Angelopoulou, G., Kasselimis, D., Laskaris, N., Kourtidou, E., Constantinides, V. C., Bougea, A., Paraskevas, G. P., Papageorgiou, G., Tsolakopoulos, D., Papageorgiou, S. G., & Kapaki, E. (2022). Silent pauses and speech indices as biomarkers for primary progressive aphasia. *Medicina, 58*(10), 1352. https://doi.org/10.3390/medicina58101352

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Ramanarayanan, V., Bresch, E., Byrd, D., Goldstein, L., & Narayanan, S. S. (2009). Analysis of pausing behavior in spontaneous speech using real-time magnetic resonance imaging of articulation. *The Journal of the Acoustical Society of America, 126*(5), EL160–EL165. https://doi.org/10.1121/1.3213452

Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University version 2.2.9. https://CRAN.R-project.org/package=psych.

Ruder, K. F., & Jensen, P. J. (1972). Fluent and hesitation pauses as a function of syntactic complexity. *Journal of Speech and Hearing Research, 15*(1), 49–60. https://doi.org/10.1044/jshr.1501.49

Sahraoui, H., Mauclair, J., Baqué, L., & Nespoulous, J. L. (2015). What do pause patterns in non-fluent aphasia tell us about monitoring speech? A study of morpho-syntactic complexity, accuracy and fluency in agrammatic sentence and connected

discourse production. *Frontiers in Psychology, 6*. https://doi.org/10.3389/conf.fpsyg.2015.65.00060
Schmitt, N. (2012). Formulaic Language and collocation. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (p. wbeal0433). Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0433.
Schmitt, N., Sonbul, S., Vilkaitė-Lozdienė, L., & Macis, M. (2019). Formulaic Language and collocation. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (1st ed., pp. 1–10). Wiley. https://doi.org/10.1002/9781405198431.wbeal0433.pub2.
Schneider, U. (2018). ΔP as a measure of collocation strength: Considerations based on analyses of hesitation placement in spontaneous speech. *Corpus Linguistics and Linguistic Theory, 0*(0). https://doi.org/10.1515/cllt-2017-0036
Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., de Jong, N. H., & Bickel, B. (2018). Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences, 115*(22), 5720–5725. https://doi.org/10.1073/pnas.1800708115
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
Skipper, J. I., Aliko, S., Brown, S., Jo, Y. J., Lo, S., Molimpakis, E., & Lametti, D. R. (2022). Reorganization of the neurobiology of language after sentence overlearning. *Cerebral Cortex, 32*(11), 2447–2468. https://doi.org/10.1093/cercor/bhab354
Stahl, B., & Kotz, S. A. (2014). Facing the music: Three issues in current research on singing and aphasia. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01033
Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition, 14*(1), 17–26. https://doi.org/10.3758/BF03209225
Strangert, E. (1997). Relating prosody to syntax: Boundary signalling in Swedish. *5th European conference on speech communication and technology (Eurospeech 1997)*. https://doi.org/10.21437/Eurospeech.1997-87
Swinburn, K., Porter, G., & Howard, D. (2008). Comprehensive aphasia test *(Reprint)*. Psychology press.
Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral 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). The Continuum International Publishing Group.
Van Lancker Sidtis, D. (2012). Formulaic language and language disorders. *Annual Review of Applied Linguistics, 32*, 62–80. https://doi.org/10.1017/s0267190512000104
Varley, R., Heilemann, C., Warren, J., Dąbrowska, E., & Javadi, A.-H. (2020). Computer therapy combined with non-invasive brain stimulation for sentence processing difficulties in post-stroke aphasia: A randomised control trial (the UTILISE study). *Open Science Framework*. https://doi.org/10.31219/osf.io/fduqh [Preprint].
Watanabe, M., Hirose, K., Den, Y., & Minematsu, N. (2008). Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication, 50*(2), 81–94. https://doi.org/10.1016/j.specom.2007.06.002
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; a Journal Devoted To the Study of the Nervous System and Behavior, 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
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