



Effects of Lexical Frequency and Collocation
Strength of Word Combinations on Speech Pause
Duration of Individuals with and Without Aphasia

Sebastian Bello-Lepe, Sabrina Mahmood, Rosemary Varley and
Vitor Zimmerer

EasyChair preprints are intended for rapid
dissemination of research results and are
integrated with the rest of EasyChair.

August 26, 2021

Effects of lexical frequency and collocation strength of word combinations on speech pause duration of individuals with and without aphasia

Sebastian Bello-Lepe^{1*}, Sabrina Mahmood¹, Rosemary Varley¹ and Vitor Zimmerer¹

¹ Department of Language and Cognition, University College London, London, United Kingdom

*corresponding author, sebastian.bello@ucl.ac.uk

Introduction

In aphasia, an increase in number and length of pauses between words contributes to communication difficulties. Pauses (PS) in speech reveal neurocognitive processes underpinning language production (Butterworth, 1979). Previous studies have found that PS was lower before words with higher frequency (Beattie & Butterworth, 1979; Goral et al., 2010). However, frequency also manifests as collocation strength between words. Stronger collocations may be processed more holistically, reducing processing effort, and speakers with aphasia tend to produce more strongly collocated combinations (Bruns et al., 2019, Zimmerer et al., 2018). In this study, we investigated the effects of Lexical Frequency and Collocation Strength on PS in narrations of individuals with aphasia (IWA) and neurotypical controls (NC). We predicted pauses would be shorter before words of higher frequency, or within stronger collocations.

Methods

20 NC and 20 IWA narrated the “Dinner Party” comic (Fletcher & Birt, 1983). Aphasic participants presented with a range of impairments and severities, including both fluent and non-fluent profiles. Transcriptions were annotated using ELAN (Max-Planck-Institute for Psycholinguistics, 2020). We set no minimum duration for pauses, and values could be zero (no pause before a word). Lexical Frequency and Collocation Strength (measured as t-scores) were analysed using the Frequency in Language Analysis Tool (FLAT; Zimmerer et al., 2018). We further determined word category using the R package “Spacyr” (Benoit & Matsuo, 2018).

Results

Table 1. Descriptive statistics for Word Count, Pause Duration, Lexical frequency and Collocation Strength (Bigram T-Scores) in neurotypical controls and individuals.			
Neurotypical Controls			
Word Count	Pause Duration Mean (SD)	Lexical Frequency Mean (SD)	Collocation Strength Mean (SD)
294.4 (180.7)	157ms (421)	9322 per million (13154)	19.46 (38)
Individuals with aphasia			

Word count Mean (SD)	Pause Duration Mean (SD)	Lexical Frequency Mean (SD)	Collocation Strength Mean (SD)
227.2 (142.2)	776ms (1822)	11237 per million (14237)	20.3 (37.9)

Linear mixed effect models revealed that IWA showed longer PSs ($p < .001$): IWA produced longer pauses before function words ($p < .001$). If word category was considered, Lexical Frequency effects were not significant ($p = .518$). However, the effect of collocation was, and PSs were shorter within stronger collocations ($p < .001$). There was an interaction between Group and Collocation Strength, with greater effects of Collocation Strength in IWA ($p < .001$). An effect size analysis showed greater Coefficient standard on IWA group for Collocation Strength (NC= $\beta(-0.15)$; IWA = $\beta(-0.23)$; $p < .001$) and Lexical Frequency (NC= $\beta(0.01)$; IWA = $\beta(-0.05)$, $p < .001$).

Conclusions

PS was influenced by Collocation Strength, supporting the view that strong collocations reduce processing demands. Collocation Strength had a bigger effect on pauses in aphasia, suggesting that as analytic capacities decrease, statistical properties, such as Collocation Strength, exhibit a greater influence on language production. Frequency-based approaches may be valuable in explaining patterns of preservation and impairment in aphasic language production.

References

- Beattie, G., & Butterworth, B. (1979). Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and Speech*, 22(3), 201-211.
- Benoit, K., & Matsuo, A. (2018). R Package: 'spacyr.' <https://mran.microsoft.com/snapshot/2018-0330/web/packages/spacyr/spacyr.pdf>
- Bruns, C., Varley, R., Zimmerer, V., 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.
- Butterworth, B. (1979). Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain & Language*, 5, 133-161.

ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>

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

Goral, M., Levy, E., Swann-Sternberg, T., & Opler, L. (2010). Frequency and Word-Length Factors and Lexical Retrieval in Sentence Production in Aphasia. *Procedia, Social and Behavioral Sciences*, 6, 107-108.

Zimmerer, VC., Newman, L., Thomson, R., Coleman, M., & Varley, R. (2018). Automated Analysis of Language Production in Aphasia and Right Hemisphere Damage: Frequency and Collocation Strength. *Aphasiology*, 32 (11) p1267-1283.

Acknowledgments

This research is supported by the Chilean National Agency for Research and Development (ANID).