**Lexical Access in Inner Speech: An Eye-Tracking Study on Word Frequency Effects in Covert Picture Naming**

Christoph Jörger (christoph.joerger@gmail.com)[1], Johannes Gerwien[1,2]

[1] Institute for German as a Foreign Language Philology, Heidelberg University

[2] Core Facility for Neurocognitive Systems Research, Heidelberg University

---

Inner speech (IS) has attracted growing interest across cognitive psychology, neuroscience, and philosophy [1], yet its study remains methodologically challenging due to its inherently covert nature. Most investigations rely on verbal interference paradigms, which risk disrupting the very process under investigation [2]. While existing psycholinguistic evidence suggests parallels between inner and overt speech at the level of phonological encoding [3–6], it remains unclear whether these extend to lexical retrieval. Here, we address this gap using eye tracking to examine word frequency effects during covert picture naming — a method that allows IS to be observed without interfering with it.

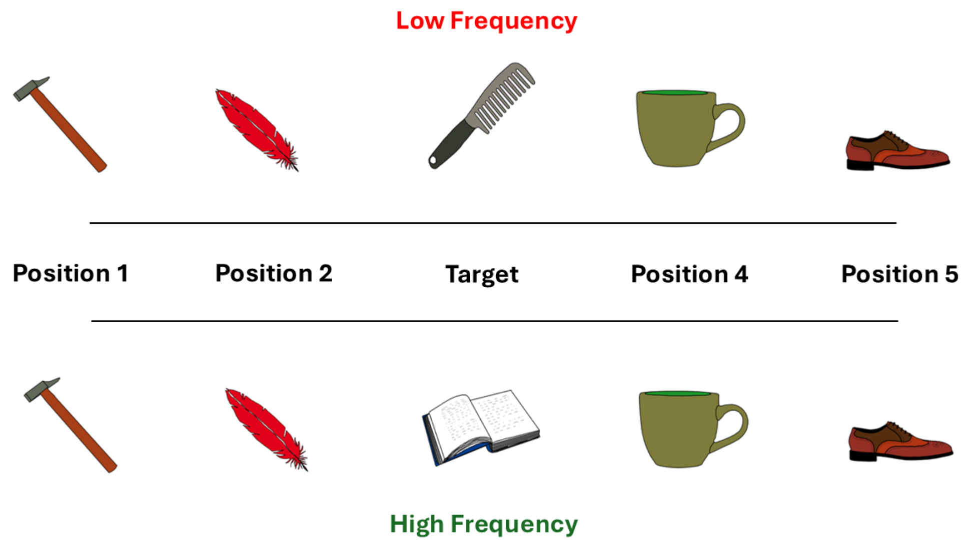
Thirty-three native speakers of German completed an audiovisual probe recognition task. On each trial, a horizontal array of five object pictures [7] was shown for 6000 ms. After the objects disappeared and following a 1500 ms delay, an auditory probe word was presented, and participants indicated whether it matched the name of one of the previously displayed objects. The critical manipulation was the lexical frequency of the name of the center picture, which was either high-frequency (HF; ≥ 29.33 pmw) or low-frequency (LF; ≤ 4.21 pmw) [8].The two sets of 30 pictures were matched for visual complexity, name agreement, and word length. Each critical object was paired with a matched filler sequence from the other frequency group (see Fig. 1), with the two matched items presented in different blocks. We analyzed first-pass gaze durations on the critical objects, predicting longer gaze durations for LF than for HF items under the hypothesis that gaze reflects covert lexical retrieval [9, 10].

Mean initial gaze durations were 572 ms for HF objects and 637 ms for LF objects. One-sided t-tests showed significant condition effects by subjects ($t(32) = 7$, $p < .001$, $d = 1.22$, 95% CI [47, 83]) and by items ($t(57.81) = 2.82$, $p = .003$, $d = 0.71$, 95% CI [21, 109]). Because the picture groups were not controlled for concept familiarity and onset complexity, we additionally fitted a linear mixed-effects model including random intercepts for subjects and items and fixed effects of word frequency (continuous), concept familiarity, onset complexity, and trial index. Word frequency still significantly improved model fit ($\chi^2(1) = 4.8$, $p = .028$).

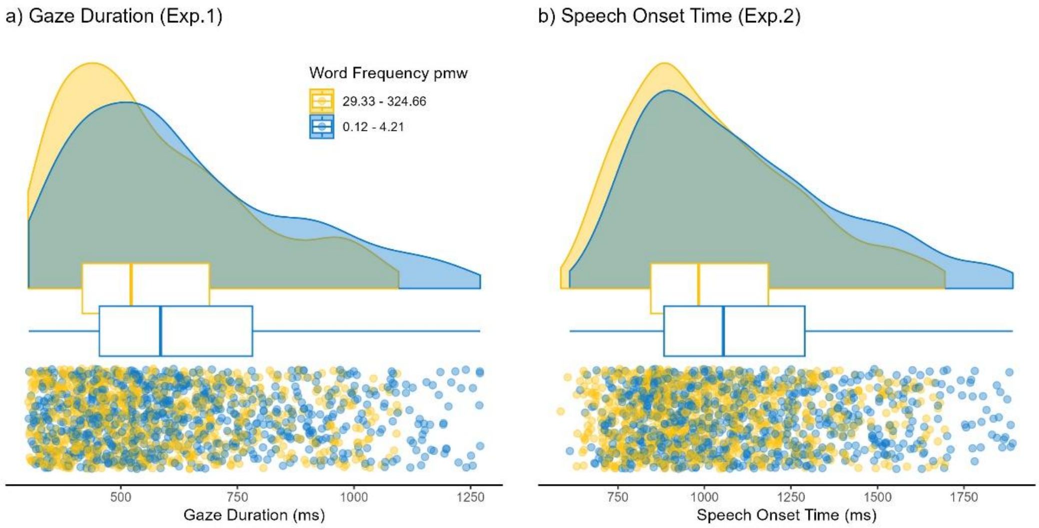
To obtain independent evidence for word frequency effects in overt production using the same materials, 29 different native speakers of German completed a picture-naming task. Mean naming latencies were 1037 ms for HF items and 1124 ms for LF items. One-sided t-tests showed significant condition effects by subjects ($t(28) = 11.57$, $p < .001$, $d = 2.15$, 95% CI [72, 102]) and by items ($t(53.12) = 3.7$, $p < .001$, $d = 0.95$, 95% CI [39, 135]). In a parallel linear mixed-effects model, word frequency (continuous) marginally improved model fit ($\chi^2(1) = 3.61$, $p = .057$). The results of both experiments are summarized in Figure 2.

These findings extend the evidence for procedural overlap between inner and overt speech – previously limited to phonological encoding – to lexical retrieval. They also demonstrate the utility of gaze duration as a temporally fine-grained, ecologically valid index of IS use. In addition, its temporal resolution may prove useful in adjudicating ongoing debates about whether the word frequency effect originates at the lemma or lexeme level [11].

## Figure 1. Item Design




## Figure 2. Results




## References


[1] Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. Psychological Bulletin, 141(5), 931–965.

[2] Nedergaard, J. S. K., Wallentin, M., & Lupyan, G. (2023). Verbal interference paradigms: A systematic review investigating the role of language in cognition. *Psychonomic Bulletin & Review*, *30*(2), 464–488.

[3] Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the Time Course of Phonological Encoding. *Journal of Memory and Language*, *34*(3), 311–334.

[4] Zelinsky, G.J. & Murphy, G.L. (2000). Synchronizing visual and language processing: an effect of object name length on eye movements. *Psychological Science, 11*(2), 125-131.

[5] Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–1160.

[6] Corley, M., Brocklehurst, P. H. & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 162–175.

[7] Duñabeitia, J.A., Crepaldi, D., Meyer, A.S., New, B., Pliatsikas, C., Smolka, E. & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816.

[8] Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A.M., Bölte, J. & Böhl, A. (2011). The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424.

[9] Meyer, A.S., Sleiderink, A.M. & Levelt, W.J.M. (1998). Viewing and naming objects: eye movements during noun phrase production. *Cognition, 66*(2), B25–B33.

[10] Griffin, Z.M. (2001) Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*(1), B1–B14.

[11] Corps, R.E. & Meyer, A.S. (2023). Word frequency has similar effects in picture naming and gender decision: A failure to replicate Jescheniak and Levelt (1994). Acta Psychologica, 241, 104073.