CORTEX 184 (2025) 172–193

Available online at www.sciencedirect.com

ScienceDirect

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

Research Report

People with aphasia show stable Cumulative Semantic Interference (CSI) when tested repeatedly in a web-based paradigm: A perspective for longitudinal assessment

Kirsten Stark [a,b,c,*,†], Marcus Töpel [d,e,†], Frank Regenbrecht [d,e], Cornelia van Scherpenberg [c,d], Rasha Abdel Rahman [a,b,c] and Hellmuth Obrig [c,d,e]

[a] Humboldt-Universität zu Berlin, Department of Neurocognitive Psychology, Berlin, Germany
[b] Charité - Universitätsmedizin Berlin, Einstein Center for Neurosciences, Berlin, Germany
[c] Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Berlin, Germany
[d] Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurology, Leipzig, Germany
[e] University Hospital and Faculty of Medicine Leipzig, Clinic for Cognitive Neurology, Leipzig, Germany

ARTICLE INFO

Article history:
Received 22 March 2024
Reviewed 18 July 2024
Revised 26 September 2024
Accepted 5 November 2024
Action editor Swathi Kiran
Published online 27 December 2024

Keywords:
Aphasia
Language production
Online experiments
Overt speaking
Cumulative semantic interference
Picture naming
MRI
Brain lesions

ABSTRACT

Retrieving words quickly and correctly is an important language competence. Semantic contexts, such as prior naming of categorically related objects, can induce conceptual priming but also lexical-semantic interference, the latter likely due to enhanced competition during lexical selection. In the continuous naming (CN) paradigm, such semantic interference is evident in a linear increase in naming latency with each additional member of a category out of a seemingly random sequence of pictures being named (cumulative semantic interference/CSI effect). Extensively studied in neurotypical participants, CSI studies in people with aphasia (PWA) are rare, although some lesions regularly and persistently impair word retrieval. In the present study, 20 PWA with lesions in the extended left hemispheric language network and 20 matched controls underwent a CN paradigm, naming photographs of closely related objects from 24 categories (e.g., birds) with 5 members each. The experiment was conducted web-based (Stark et al., 2022) on three days (day 1, 2, and 8). The main results are: (i) Mild-moderate aphasia does not preclude web-based testing. (ii) The CSI effect in naming latencies (~21 ms per ordinal position) did not differ significantly between groups but was more variable in the PWA; the effect was stable across days. (iii) Overall response times decreased between day 1 and day 2, but remained stable on day 8. (iv) In PWA, increased error-rates paralleled the latency-based CSI effect, suggesting stronger interference in this group. (v) Exploratory analyses suggest that lesions in a large area, including frontal, inferior parietal, pre- and post-central opercular cortices, are linked to a larger CSI effect. At a more lenient statistical threshold,

* *Corresponding author*. Neurocognitive Psychology, Department of Psychology, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099, Berlin, Germany.
E-mail address: kirsten.stark@hu-berlin.de (K. Stark).
† Kirsten Stark and Marcus Töpel share the first authorship.

https://doi.org/10.1016/j.cortex.2024.11.019

lesions in occipital and supramarginal cortices were associated with increased overall naming latencies. These results offer an initial step toward identifying the neuronal underpinnings of semantic context effects in PWA. We conclude that web-based assessment is feasible in PWA and yields a stable CSI effect over repetitive testing. While not directly clinically applicable, the findings could serve as a foundation for exploring training-interventions targeting lexical activation, interference resolution, or word selection.

1. Introduction

Fast and correct retrieval of words is a cornerstone of efficient language production. Seemingly effortless in uncompromised speakers, deficits in word retrieval are common and persisting symptoms of acquired language disorders (aphasia). Based on extensive research in neurotypical participants and people with aphasia (PWA), psycholinguistic models assume several sub-processes to word retrieval, including the activation of lexical entries (lemmas) in the mental lexicon and the subsequent selection of the correct entry from co-activated members of a cohort (e.g., "cat" from the cohort "pets"; e.g., Abdel Rahman & Melinger, 2019; Bloem et al., 2004; Dell et al., 1999; Indefrey & Levelt, 2000; Levelt, 1999; Levelt et al., 1999; Oppenheim et al., 2010; Roelofs, 2022). Research in uncompromised speakers shows that semantic context, depending on its relatedness to the target word (conceptual/lexical/post-lexical) and the presentation mode, either facilitates or inhibits word production, i.e., speeds up or hampers naming. The reported effects mostly affect naming-latencies and are usually subtle. The few studies on people with aphasia (PWA) suggest that typical aphasic errors, such as semantic paraphasia (e.g., saying "dog" when meaning cat) partly stem from similar inhibition and facilitation patterns (e.g., Piai & Knight, 2018; Schnur et al., 2009). Beyond further confirmation of psycholinguistic models, work in PWA is also warranted to devise novel theory-grounded treatment schemes. In the current study, we therefore assess the stability of the semantic interference in PWA and a matched control group in repeated tests over one week. If the interference effect remains stable across cohorts and repetitive testing, this suggests that the effect of semantic context manipulation is robust and rather short lived and that lasting changes can only be achieved through intensive training (e.g., Bruehl et al., 2023; Patra et al., 2022; Spitzer et al., 2021). Regarding a clinical perspective, assessing semantic interference in a web-based paradigm using picture naming may ease future research on whether and how modifications of the cumulative semantic interference (CSI) effect can be elicited in PWA. As a first step to evaluate the potential for interventions, we here seek to establish whether stability over repetitive testing can be assumed.

1.1. *Paradigms evoking semantic context effects in confrontational picture naming*

Picture naming is the standard way to probe controlled word retrieval. Semantic context effects have largely been evidenced by three manipulations: (i) In the *picture-word interference (PWI)* paradigm, participants name a target picture more slowly when it is presented together with a semantically related word instead of an unrelated distractor word (e.g., $cat_{word} + dog_{picture}$ slower than $cat_{word} + bus_{picture;}$ Damian & Bowers, 2003a; de Zubicaray et al., 2012; Glaser & Düngelhoff, 1984; Henseler et al., 2014; Jescheniak et al., 2005; Schriefers et al., 1990). (ii) In the *blocked cyclic naming (BCN)* paradigm, participants name pictures more slowly when presented in blocks of semantically related items when compared to blocks containing unrelated items (Belke & Stielow, 2013; Navarrete et al., 2012). (iii) In the *continuous naming* paradigm, participants name a seemingly random sequence of pictures. Notably, with each novel exemplar from a given semantic category, response-time increases, signaling the cumulative semantic interference effect (CSI; Howard et al., 2006). This interference occurs largely independent of the number of unrelated intervening items (Howard et al., 2006; Schnur, 2014) and is stronger when category members are more closely related (e.g., apes: orangutan, chimpanzee, … *vs* animals: orangutan, horse, fish, …; Rose & Abdel Rahman, 2017). Thus, across these paradigms, semantic context based on a categorical relationship (which defines lexical cohorts) increases the naming latency (Döring et al., 2022; Pino et al., 2022; Schnur et al., 2006). A model to accommodate interference effects that also includes facilitatory factors such as associative primes (e.g., bone → dog) posits differential weights of all modulating factors at the different stages of word production leading to a net-effect either facilitatory or inhibitory in nature (Abdel Rahman & Melinger, 2019; Chen & Mirman, 2012). Thus, incorrect aphasic word retrieval may be partially due to pathologically altered semantic context effects during picture naming.

In the present study, we use the continuous naming (CN) paradigm and assess the CSI effect in a group of PWA and a matched neurotypical control group. We use the CN paradigm because it is ecologically valid. This especially holds for studies involving PWA in whom sequential picture naming is regularly used for diagnostics and training (e.g., Bruehl et al., 2023; Conroy et al., 2009; Nickels, 2002; Pagnoni et al., 2021; Walker et al., 2018). Furthermore, the critical manipulation, with several semantic categories whose members are embedded in a pseudo-random sequence of items to be named, is usually not noticed by the participant (Howard et al., 2006). The CN is efficient because interleaving different categories is possible, in that for a given category, exemplars of other categories can serve as 'fillers', hereby also counteracting the statistical impact of effects such as fatigue (Döring et al., 2022; Harvey et al., 2019; Rose & Abdel Rahman, 2017; Schnur, 2014). The CSI effect is robust and substantial (in

neurotypicals 13–30 ms for each new member of a category; Dyson et al., 2021; Fuhrmeister & Bürki, 2022; Glaser & Düngelhoff, 1984; Hart, 2017; Harvey et al., 2019; Howard et al., 2006; Stark et al., 2022). Indeed, for five exemplars of a category, the overall CSI between the first and the last exemplar amounts to more than 50 ms.

Regarding the design of the present study, we only slightly modified that of our preceding study (Stark et al., 2022). In that study, young, neurotypical participants showed a CSI effect in spoken and typed responses which was comparable to previous lab-based designs. Of the two modalities tested there, we here only use the spoken response modality, since it is more ecologically valid especially for PWA and typing skills are quite variable in older and neurologically impaired populations.

1.2. *Semantic context effects impact different levels of word retrieval and production*

Most models of speech production assume a conceptual, a lexical, and a phonological-articulatory phase. In the conceptual phase, visual features are analyzed (e.g., 'has legs' *vs* 'has wings') and compared with the supramodal semantic memory (e.g., 'barks' + 'has fur' → dog). This triggers the lexical phase in which the lexical entry is activated and selected (activation of the lemma, e.g., /DOG/). To produce the intended word, the final phase involves phonological-morphological encoding and articulation (retrieval of phonological code and preparation of articulatory gestures, e.g., /dɒg/; Levelt et al., 1999). Regarding semantic interference effects, the choice of the correct lemma during the lexical phase is crucial (Dell, 1986).

Word production models largely agree that the activation of lemmas, triggered as the result of the conceptual/feature activation, includes the activation of entire lexical cohorts of related entries ("spreading activation", e.g., cat → all pets; Roelofs, 1992). While this spreading activation induces facilitatory semantic priming, it also affects the competition between the members of the cohort, resulting in slower, more error-prone or faster naming, depending on the context. Indeed, the narrower a cohort is defined (e.g., living → animals → pets → dogs), the stronger is the competition between its members (Rose et al., 2019). Yet, the loci of facilitatory and inhibitory semantic context effects are partially controversial. Facilitation is mostly assumed to be evoked by semantic relations at the conceptual level (Bloem et al., 2004; Bloem & La Heij, 2003). Conversely, interference has been taken to reflect competition for selection between cohort members due to lexical co-activation (Damian & Bowers, 2003b; Jescheniak et al., 2014; Starreveld & La Heij, 2017). The more items are active at the lexical level, the stronger the competition for selection.

In the CN task, an additional learning mechanism has been suggested. Upon naming a picture, the link between a concept and its corresponding lemma is strengthened (Howard et al., 2006; Oppenheim et al., 2010). As a result of these naming-induced adaptations of connection weights, previously named related items are activated more strongly and are therefore stronger competitors at the lexical level. Accordingly, with each new member of a semantic category, the number of previously named strong competitors increases, resulting in increasing selection difficulties—in cumulative semantic interference. For completeness it should be noted that other accounts hold for specific paradigms, including post-lexical mechanisms in the PWI task (Finkbeiner & Caramazza, 2006; Mahon et al., 2007).

Being aware of these controversies, we here proceed from the well attested fact that interference in the CN paradigm increases with each member of a given cohort being named and ask whether the CSI effect in PWA is similar to that documented in neurotypical participants. In addition, the feasibility of web-based testing and the stability of the effect across three sessions and eight days are addressed.

1.3. *Semantic context effects are relevant for people with aphasia*

In PWA with non-fluent aphasia, erroneous or unsuccessful word retrieval is a common symptom, next to the generally slower production. Word-finding problems, the core symptom of their production deficit, manifests in lexical search behavior, omissions, semantic and/or phonemic paraphasia and/or, in severe cases, neologisms. Semantic paraphasia (saying 'dog' when meaning cat) in particular support the above sketched models of lexical competition during word retrieval (Schwartz, 2014): 'Noisy access' to the mental lexicon may hinder naming due to lower activation of the target entry (Harvey & Schnur, 2015). Additionally, erroneous selection from a co-activated lexical cohort seems to play a critical role in the generation of aphasic word production errors (Nozari & Hepner, 2019). These semantic context effects are relevant at different stages of (erroneous) word production, not least when repetitive and confrontational naming is used during speech-language therapy (SLT; Abel et al., 2007). Indeed, repetitive naming 'within' the semantic cohort can lead to increased interference in some patients (Pino et al., 2022).

Different therapy schemes have been proposed for the treatment of word finding difficulties. The widely used principle of "vanishing cues" trains naming of an item repetitively with all available cues (written word form, auditory presentation, etc.) to then gradually reduced cues, with the aim to support "errorless learning" (Haslam et al., 2010; Middleton & Schwartz, 2012; Nunn et al., 2023). Of note to our present study, another therapeutic principle, the so-called "Semantic Feature Analysis" capitalizes on the facilitatory effect of conceptual/featural analyses for lexical retrieval. Indeed, describing predefined features of the object to be named eased lexical retrieval in patients with anomia (Efstratiadou et al., 2018). Interestingly, another novel approach suggests that both semantic word processing and executive control can be trained by using established semantic-interference paradigms as a training intervention (Bruehl et al., 2023). After a four-week training program, naming improved in conditions with and without interference, partly generalizing to untrained items. This was mirrored by activation changes in specific brain regions. The latter two intervention schemes highlight the clinical relevance of a better understanding of semantic context effects in PWA to advance the development of individually tailored, theoretically grounded principles of anomia

treatment, which is one pillar of SLT (Brady et al., 2016; Breitenstein et al., 2017).

1.4. Studies in PWA show variable semantic context effects

Studies on semantic context effects yielded variable results in people with aphasia (PWA). Most studies have focussed on the picture word interference (PWI; Piai & Knight, 2018; Pino et al., 2022; Ries et al., 2019) and the blocked cyclic naming (BCN) tasks (Schnur et al., 2006) while continuous naming (CN) studies are rare: A study on 15 PWA used data from an unstructured large corpus of pictures. It suggests a persistent interference effect in error rates, which increased for every novel member of a given category. Since the effect was driven by semantic naming errors (saying "cat" for the picture of a dog), the authors conclude that PWA, like neurotypicals, show relatively permanent, experience-driven changes in connection strengths between semantic and word-form representations, often called incremental learning (Harvey et al., 2019). In a review on the differences between BCN and CN, Belke and Stielow (2013) propose that exaggerated context effects in PWA for the BCN task may stem from a lack in top-down bias, especially in Broca-type aphasia. Although this has not been studied explicitly, this mechanism should not affect the CN task, since it does not involve repetition of the same items. In the same vein, increased cumulative semantic interference across blocks of a BCN task was reported in a patient with Broca's aphasia and interpreted as a problem of lexical control (Scott & Wilshire, 2010). In another case study, two non-fluent PWA showed exaggerated blocking effects that increased with repetition, while a fluent person with aphasia did not show such an effect (Biegler et al., 2008). Combing lesion and functional imaging data, Schnur and colleagues suggest the left IFG to bias lexical selection when competition between candidates arises (Schnur et al., 2009). To our knowledge, as yet, a formal comparison between BCN and CN in a larger cohort of PWA with variable lesion/deficit profiles is missing. Notably, however, in a recent study using both a CN and a BCN task, 6 out of 20 PWA showed 'paradoxical' *decreased* response times with ordinal position in the CN task, while other PWA showed *increased* response times (i.e., larger CSI) when compared to the control group (Nappo et al., 2022). This may indicate that PWA will show different behaviour, depending on whether activation of the lemma or selection from the categorical competitors is impaired. PWA with activation issues may 'profit' from the semantic context while those with a selection-problem will show an exaggerated CSI (Nappo et al., 2022).

In the current study, we primarily target the questions (i) whether a controlled CN-task elicits a CSI effect in PWA, similar to that in matched neurotypical controls, and (ii) whether repeated testing will change the size of the CSI. Given the variability of semantic context-effects in PWA previously reported, CSI effects in PWA may be increased, decreased, or similar when compared to neurotypical controls. Since the effect of a lesion in PWA has been suggested to depend on site/size and the patholinguistic profile, we do not have a specific prediction regarding the direction of change. To further explore whether and how lesion site affects interference effects (Harvey & Schnur, 2015; Piai et al., 2016; Piai & Knight, 2018; Schnur et al., 2009), we perform an exploratory lesion-behaviour analysis correlating overall naming latency and size of the CSI effect with lesion site and size.

1.5. Web-based assessment of semantic context effects may ease follow up of therapy-induced changes in PWA

In post-stroke clinical populations, morbidity-related reduced mobility is a challenge (Jørgensen et al., 2002; Langhorne et al., 2009). Web-based assessment could ease their repetitive testing. Moreover, infection-prevention during the pandemic has shown that remote testing may be mandatory for other medical reasons. However, linguistic effects in the range of milliseconds, such as the CSI, require reliable and accurate response time assessment. Testing patients web-based at their homes entails differences in hardware and software equipment and variably stable internet connection, and requires the participant to handle the browser and the experimental platform (Vogt et al., 2022). Driven by the pandemic, several studies have shown that reliable display times and response recordings are possible even when data are recorded via the participants' web-browsers (e.g., Anwyl-Irvine et al., 2021), especially if within-participant comparisons are intended (Rodd, 2024; Sauter et al., 2020; Vogt et al., 2022). Numerous lab-based experiments have been successfully replicated web-based (e.g., Fairs & Strijkers, 2021; Hilbig, 2016; Sauter et al., 2022; Stark et al., 2022; Stoehr et al., 2023; Vogt et al., 2022). The results show acceptable data quality and smaller than expected cross-setup variability (often <10 ms; e.g., Anwyl-Irvine et al., 2021; Bridges et al., 2020; de Leeuw & Motz, 2016; Mathôt & March, 2021; Reimers & Stewart, 2015). Although these data are encouraging, the exact degree of imprecision remains unknown and is likely to depend both on the technical setup and on the study population: In a recent study, Bürki and Vasishth (2024) compared students in a lab-based speech production experiment to participants from a more general online pool in a web-based version of the same paradigm. They found good within-participant consistency, but smaller effect sizes in the web-based data. This was due to higher between-participant variation, possibly from the more diverse participant pool and home distractions. While aware of these limitations, web-based testing offers significant opportunities for longitudinal studies and tracking therapy-induced changes in language production.

While different outcome measures of language production have been studied web-based in both neurotypicals and PWA (e.g., Bevivino et al., 2023; Kandel et al., 2022; Lee et al., 2023; Mcconnell et al., 2024; Payne, 2020; Python et al., 2023; Schwarz et al., 2022), we are not aware of any studies assessing time-critical measures of language production in PWA yet. However, regarding unimpaired language production, web-based assessment of voice onset times has been demonstrated to be feasible in young, neurotypical populations (Bürki & Vasishth, 2024; Corps & Meyer, 2023; Fairs & Strijkers, 2021; Gavard & Ziegler, 2024; He et al., 2021, 2024; Li et al., 2022; Mooijman et al., 2023; Stoehr et al., 2023; Vogt et al., 2022), including in the continuous naming paradigm (Stark et al., 2022). The latter study demonstrated a relatively large and stable CSI effect with an increase in naming latency of

~31 ms with each additional category member named. The effect was comparable in spoken and written responses and in the range of effect sizes reported from data collected in the lab. Moreover, internet-based therapy in PWA has been advanced recently (Asghar et al., 2021; Cherney et al., 2021; Kiran et al., 2014), including online naming tasks providing automated feedback (Pompili et al., 2011). Therefore, web-based assessment and interventions must be considered powerful tools for improving therapeutic care and long-term treatment, also for people with acquired language disorders. Besides such general advantages, diversity of study participants may be increased by recruiting participants in rural or other 'hard-to-reach' environments (Henrich et al., 2010). The current study aimed to establish the feasibility and validity of web-based response time assessment in PWA, with the potential to open new avenues for research and training assessment. Using a paradigm and stimuli evoking reliable interference in web-based settings in young participants, we sought to investigate the specific mechanisms and potential pitfalls in PWA.

1.6. *Aims of the present study*

Based on the above considerations the present study targets web-based testing in people with aphasia (PWA) for a semantic interference paradigm. The ecologically valid CN paradigm assesses semantic context effects in picture naming, one of the pillars of diagnostics and therapy in PWA. We address the questions whether (i) the CSI effect can be demonstrated in a web-based continuous naming task in people with mild or moderate aphasia, and (ii) whether such assessment can be performed repeatedly, with stable effects.

We hypothesized that web-based assessment should be feasible in people with mild-to-moderate aphasia. We tested the cohort of PWA and control participants using the same design as previously tested in young neurotypical individuals (Stark et al., 2022). Because we expected PWA to need more support especially for the initial familiarization with the set-up, remote support via telephone was provided throughout all sessions.

Regarding the repeated testing, we expected adherence to the protocol in all participants who completed the first day. Hypotheses on potential modulations of the CSI effect upon repeated testing are less straight-forward, since, to our knowledge, repeated testing in the CN-paradigm has not been formally addressed before. Three results are possible: (i) A stable effect over repeated testing would indicate that the interference between co-activated cohort members is not modulated until the next experimental session. Indeed, head priming effects in compound production have been reported stable across sessions in repetitive testing (within one day; Lorenz et al., 2022). (ii) Alternatively, an increase in CSI over successive testing would speak for a long-term strengthening of the within-cohort connections, leading to increased spreading activation and thereby to an increased CSI effect. (iii) A decrease in the magnitude of the CSI would suggest that selection of the correct lemma from the activated lexical cohort becomes more efficient upon test repetition. Although not a main target of the study we expected an overall increase in naming speed. This hypothesis was based on other schemes of repetitive testing. Moreover, we expected slower naming and more variable CSI effects in the PWA when compared to the neurotypical control group. In exploratory follow-up analyses, we controlled for the effect of aphasic syndrome severity and chronicity.

To gain a first idea on the neural correlates of the CSI, we additionally correlated lesion size and site based on brain images with overall naming skills, the size of the CSI effect, and changes over repetitive testing. This exploratory analysis used a multivariate approach. Candidate regions relevant for confrontational naming are the anterior and posterior hubs of the neuronal network supporting language. The anterior part (including the inferior frontal gyrus/IFG) may be of special relevance for selecting the correct member from an activated lexical cohort (Anders et al., 2019; Badre et al., 2005; Pino et al., 2022). Posterior temporal and inferior parietal regions seem to be involved in the activation of the lexical cohort, generating lexical competition (Harvey & Schnur, 2015; Piai & Knight, 2018; Pisoni et al., 2012). Both lesions in anterior and in posterior temporal and inferior parietal regions were taken into consideration for exploratory lesion-behavior analyses.

The hypotheses and analyses were preregistered on the Open Science Framework: https://osf.io/gkqy9.

2. Methods

2.1. *Participants*

Twenty people with aphasia (PWA; 3 female; M_{age} = 53.4, SD = 5.43, range: 39–62) and 20 sex-, education- and age-matched control participants (Control; 3 female: M_{age} = 53.2, SD = 5.72, range: 38–62) who successfully completed all test sessions and embedded attention checks within the preregistered time frame were recruited for the experiment (Table 1). Sample size was determined by an a-priori power analysis using the R package simr (v1.0.5; Green & MacLeod, 2016) and was based on estimates from a previous patient and a previous online study (Lorenz et al., 2021; Stark et al., 2022). To account for the expected greater noise in the online setting, we increased the estimated sample size for power of >80% by 25% to 20 valid data sets per group. Four additional PWA were tested but excluded because audio files were not available (n = 2) or because the test-schedule deviated substantially from the protocol (n = 2).

All PWA had a chronic, acquired left hemispheric lesion (over three months after onset). Participants were included only if their spontaneous speech showed impairment according to the German Aachen Aphasia test (AAT) dimension "Semantics", as assessed based on a semi-structured interview (i.e., scores of ≤4 for interview with 6 rating dimensions, including "Semantics". All dimensions are assessed on 5-point rating scales with 5 indicating "unimpaired"; Huber et al., 1984). The overall test profiles ("profile height") of the included participants showed mild to moderate aphasia. According to the classifier of the AAT, the following aphasia

Table 1 – Epidemiological and patholinguistic essentials of final PWA cohort and matched controls ($n = 20$ each).

	age [*yrs.*]	sex [*m/f*]	edu [*yrs.*]	MPO	stroke	hand. [*R/L*]	LQ [*%*]	AAT			LEMO	Matched Control		
								TT [*% rank*]	NAM [*% rank*]	SEM	SYNON [*% corr.*]	age [*yrs.*]	sex [*m/f*]	edu [*yrs.*]
1	51	m	19	19	isch	R	82	97	98	3	85	50	m	16
2	53	m	19	15	isch	R	100	99	99	4	90	53	m	18
3	60	m	12	7	isch	L	−80	95	97	4	83	61	m	13
4	54	m	21	121	ICH	R	100	39	42	3	75	52	m	22
5	51	m	19	93	ICH	R	90	79	91	4	88	51	m	12
6	51	m	13	93	isch	R	90	86	94	3	83	51	m	22
7	57	m	16	30	ICH	R	100	97	93	4	88	57	m	20
8	55	m	16	23	ICH	R	70	93	99	4	85	54	m	21
9	62	f	22	68	ICH	R	100	58	52	3	83	62	f	18
10	51	m	13	23	ICH	R	82	97	100	4	98	51	m	13
11	55	m	12	120	isch	R	100	99	96	4	85	57	m	17
12	55	m	12	4	isch	R	100	15	28	3	93	52	m	13
13	58	m	13	42	isch	L	−100	99	100	4	95	59	m	12
14	44	m	11	32	isch	R	67	86	97	3	90	43	m	9
15	59	m	15	40	isch	R	100	83	100	4	95	56	m	12
16	61	m	13	65	isch	L	−100	83	100	3	85	61	m	20
17	39	m	19	17	ICH	R	100	99	100	4	88	38	m	20
18	55	m	16	82	ICH	R	90	99	100	4	90	52	m	15
19	53	f	17	32	isch	R	82	74	83	4	98	52	f	12
20	54	f	12	259	ICH	R	*n.a.*	21	75	3	73	52	f	12
av	**53.4**	**17/3**	**15.5**	**59.3**	**11/9**	**17/3**	**61.7**	**79.9**	**87.2**	**3.6**	**87.3**	**53.2**	**17/3**	**15.8**
SD	*5.43*		*3.35*	*57.92*			*68.0*	*25.56*	*20.85*	*.49*	*6.47*	*5.72*		*4.02*

Note. **Edu** = total education years; **MPO** = months post-onset; **hand.** = pre-morbid handedness according to self-report; **LQ** = laterality quotient in % (Oldfield, 1971); **AAT** = Aachen Aphasia Test (Huber et al., 1984); **TT** = percentile rank of Token Test; **NAM** = overall naming test in AAT in percentile rank; **SEM** = rating of semi-standardized spontaneous speech in the AAT regarding 'semantic structure', with 0 being the worst performance and 5 indicating uncompromised speech [0–5]; **LEMO** = German test battery based on the Logogen model (Stadie et al., 2013); **SYNON** = % correct of test for synonym judgement subtest of the LEMO test battery; ***yrs.*** = years; ***m/f*** = male/female; ***R/L*** = right/left; ***% rank*** = percentile rank; ***% corr.*** = percentage of correct responses; **isch** = ischemic stroke; **ICH** = intracerebral hemorrhage; **av** = average across each group; ***SD*** = standard deviation within each group.

subtypes were diagnosed: amnestic ($n = 6$), Broca ($n = 2$), Wernicke ($n = 1$), conduction ($n = 1$), transcortical-motor ($n = 1$), and residual ($n = 6$). Three participants showed "non-classifiable" aphasia.[1] Participants with severe or progressive cognitive impairments, apraxia, or depressive symptoms were excluded. Table 1 provides ratings for the dimension "Semantics" (SEM_{AAT}), additional measures of the AAT (Token Test [TT_{AAT}] and the subtest "Naming" [NAM_{AAT}]), and the LMEO test battery (Stadie et al., 2013). In all PWA, imaging for lesion delineation was available: 18 PWA had a high-resolution MRI from the in-house 3T-scanners (T1 mprage 1 mm^3 iso-voxel, FLAIR, 1*1*5 mm). For the other two PWA, a clinical MRI or CT with a lower resolution was available. The lesion overlay is shown in Fig. 3A.

Control participants did not report any neurological or psychiatric disorders that could affect cognitive or linguistic abilities. All participants were native German speakers and indicated (corrected-to-)normal vision. They gave their informed consent and were financially rewarded after each session. The experimental procedures were approved by the Institutional Review Board of the University of Leipzig, Germany (ethical approval 144/18-ek), in accordance with the Declaration of Helsinki.

2.2. Material and design

The material, design, and procedure were largely identical to a previous web-based experiment with young (18–35 years) neurotypical participants (Stark et al., 2022). To respect the needs of the PWA (i) two self-paced pauses were allowed mid-task (after 60 trials each), (ii) the stimulus presentation duration was increased from 2000 ms to 3000 ms, and (iii) the instructions were linguistically simplified to avoid patient frustration and misunderstanding. In the previous study, neurotypical young participants showed robust cumulative semantic interference in a single session testing, comparable to lab-based effects (Stark et al., 2022).

The continuous naming task included 120 colored target photographs of everyday objects from 24 narrow semantic categories, along with 40 semantically unrelated fillers (see supplementary material for complete list). Each semantic category thus consisted of 5 members (e.g., 'duck', 'owl', 'swan', 'ostrich', 'pigeon' for the category *birds*). The stimulus order was pseudo-randomized so that (i) members of the same semantic category were separated by two to eight items, either from other semantic categories or fillers, (ii) the order within each category was randomized across participants and sessions, and (iii) categories sharing the same superordinate category (e.g., food: vegetables and fruits) were spaced as far apart as possible (program MIX; Van Casteren & Davis, 2006; see also Stark et al., 2022). This setup allowed the CSI to be assessed over blocks of 20 items, minimizing the influence of linguistic features (e.g., word frequency) and fatigue.

2.3. Procedure

The PWA and the control participants completed the experiment three times within one week (days: d1, d2, and d8). While the procedures and stimuli were identical across sessions, the stimulus order for a given participant differed across days.[2] The experiment was programmed in SoSciSurvey (Leiner, 2019) and audio-recording was implemented using RecordRTC.js (Khan, 2020). Participants completed the experiment on a private computer while being in contact with the experimenter via telephone throughout the session. After welcoming the participant, the experimenter provided support whenever necessary during the experiment and remained quiet otherwise. Support was provided, for example, by helping the participant open the experiment link, by reading aloud the experimental instructions, or by clarifying the instructions during familiarization. PWA mostly needed assistance during the first session only. However, the experimenter remained available throughout all sessions and for all participants. When participants were still treated at the clinic or did not possess a private computer or laptop ($n = 12$ in the PWA group), the scenario was modeled, by using a separate room in the clinic, in which the participant entered the web-based experiment. The majority of participants (37/40) used the same setup across all sessions, with only few exceptions involving minor changes in screen size ($n = 1$), browser ($n = 1$), or device ($n = 1$).

During the experiment, participants were first familiarized with the general procedure and the materials. The familiarization was followed by an attention check consisting of two familiarized and two novel items that had to be judged regarding their novelty. This was used for later assessment of data quality. After four practice trials, the main task started. Each trial started with a fixation cross presented for 500 ms, followed by the target picture presented for 3 s, after which the next trial started automatically. Participants were instructed to name each picture as fast and as accurately as possible. The experiment ended after 160 trials (120 targets + 40 fillers) with another attention check and the debriefing. The attention checks were implemented to evaluate data quality, and performance in these checks or in the familiarization did not impact the test procedure. One experimental session lasted about 35 min on average for the PWA group and about 23 min for the control group.

2.4. Data processing

Voice onset times (VOTs) were defined as the start of each (target) word, excluding stuttering and articles (e.g., when participants uttered 'eh, eh, eh, tiger' or 'the tiger', the /t/ sound of tiger was used for VOT onset). VOTs were automatically detected from the trial-based audio files using Chronset (Roux et al., 2017). The correctness of the detected VOTs was then checked and, if necessary, manually

[1] The AAT yields 'non-classifiable' when the classifier does not arrive at any of the four standard syndromes (global, Broca, Wernicke, amnestic) or the two non-standard syndromes (conduction, transcortical). 'Residual aphasia' is diagnosed when there are clear deficits in one of the dimensions, while the overall profile yields no/residual aphasia.

[2] Due to a technical error, one participant had the same stimulus order on day 1 and 2 and had his third session on day 9; another participant was tested on day 3 instead of day 2. The results remained identical when these participants were excluded from data analyses.

Table 2 – Response classification in the PWA and control group (CG): 5,518 (PWA) *vs* 6,937 (CG) responses considered as correct were used for VOT analyses and 1,541 (PWA) *vs* 167 (CG) responses were classified as errors.

Session	Correct				Incorrect							technical/invalid (thereof corr.)
	corr.	altern.	phon. [<25%]	w/arti.	NR	phon. [>25%]	sem.	sup.	unrel.	neo.	partial	
example:	*'couch'*	*'sofa'*	*'coch'*	*'the couch'*	-	*'cautch'*	*'chair'*	*'furniture'*	*'bug'*	*'sitty'*	*'c'*	
People with Aphasia												
d1	**1,443**	**182**	**31**	**67**	339	20	79	29	28	12	102	68 (54)
d2	**1,648**	**162**	**36**	**18**	265	20	73	41	18	10	79	30 (28)
d8	**1,665**	**158**	**40**	**68**	216	16	58	32	26	7	71	43 (37)
Control group												
d1	**1,917**	**291**	**2**	**66**	14	–	28	23	2	–	10	47 (8)
d2	**2,095**	**209**	**6**	**2**	10	1	28	6	2	3	7	31 (4)
d8	**2,133**	**205**	**9**	**2**	7	–	16	6	3	–	1	18 (3)

Note. **Corr.** = correct; **altern.** = accepted alternative responses (see Appendix); **phon.** = phonematic paraphasia. Answers with <25% of the word being affected were considered as correct, answers with >25% of the word being affected were considered as incorrect (phon. [>25%]); **w/arti.** = article was articulated before word, but word onset was used for VOT detection; **NR** = no response was given within 3 s; **sem.** = semantic paraphasia.; **sup.** = superordinate word or category named; **unrel.** = unrelated word; **neo.** = neologism; **partial** = partial naming, i.e., correct beginning of the target name was uttered, but the recording ended before 25% of the word were pronounced; **technical** = technical errors such as missing recordings; **thereof corr.** = thereof correct, i.e., invalid VOT detection, but correct response. Only correct responses were considered for VOT analyses. For error analyses, correct and incorrect responses plus technical errors with correct responses were considered. **d1, d2, d8** = day 1, day2, day 8.

corrected and errors were classified using Praat software (Boersma & Weenink, 2020; van Scherpenberg et al., 2020). The error classification was based on Best et al. (2005) and is summarized in Table 2. A detailed overview of error type by ordinal position is provided in the (Appendix Table B1). Of the overall 14,400 experimental trials, 1,708 were excluded due to errors on the participant side, 134 due to invalid VOT detection (delayed naming of the previous item obscuring the response), and 103 due to technical errors. Thus, statistical analyses of VOTs were based on 12,455 responses considered correct (5,518 for PWA only). Exploratory analyses of error rates were based on 14,297 correct and erroneous observations (7,178 for PWA only).

2.5. *Statistical analyses*

2.5.1. *Behavioral analyses*

The CSI effect is defined as the increase in response time (VOT) or errors with each additional member of a semantic category presented (referred to as 'ordinal position' 1–5 in this paper). Beyond a replication of the CSI effect in the web-based setting (main effect of ordinal position), we were interested in whether the factors session (day1 *vs* day2 *vs* day8) and group (Control group *vs* PWA) influenced the CSI (i.e., interaction effect with ordinal position), independent of session- or group-specific increases or decreases in VOTs or error rates (main effects).

We conducted separate (generalized) linear mixed model (LMM) analyses for VOTs and errors using the package lme4 (REML criterion using optimizer "bobyqa"; v1.1-35.1; Bates et al., 2015) in R (v4.3.1; R Core Team, 2020). VOTs were log-transformed as suggested by a boxcox-test to account for their non-normal distribution.[3] The predictor ordinal position, indicating the order of presentation of an item within its semantic category, was entered to the model as a continuous, mean-centered predictor ranging from −2 to 2 (i.e., 5 levels). Accordingly, a positive β_1 estimate indicates a linear increase in VOTs by ordinal position. The predictors session and group were coded as factors with centered, orthogonal contrasts such that the intercept is the grand mean (i.e., the estimate β_0 indicates the mean VOT/error rate across all conditions), and day 2 and 8 are both compared to the session on day 1 (estimates $\beta_2 = \mu_{d2}\text{-}\mu_{d1}$ and $\beta_3 = \mu_{d8}\text{-}\mu_{d1}$, d = day; positive estimates thus reflect longer latencies/more errors on day 2 and 8), and the PWA group is compared to the control group (estimate $\beta_4 = \mu_{PWA}\text{-}\mu_{Control}$; positive estimates reflect longer latencies/more errors in the PWA; see Schad et al., 2020).

Initially, a structure with fully crossed random effects was adopted (Baayen et al., 2008; Barr et al., 2013), which was then gradually reduced by dropping random effects with variances close to zero. Previously, the number of optimizer iterations was increased to $2 \bullet 10^5$, as suggested by Brauer and Curtin (2018). The VOT-model converged with a random intercept and random slopes for ordinal position and session by subject and a random intercept and slopes for ordinal position and group by category. The same procedure was taken for exploratory analyses of error rates using a GLMM with a Binomial distribution. The error model converged with a random intercept and slope for ordinal position by subject and a random intercept and slope for group by category.

Since participants in the control group made few errors, we performed an additional analysis with error rates of the PWA group only. This model converged with a random intercept and slope for ordinal position by subject and a random intercept by category. *P*-values were estimated using the Satterthwaite approximation for VOT analyses (lmerTest package, v3.1-3; Kuznetsova et al., 2017) and calculated using the Wald *Z*-statistics for error rates. A family-wise error rate of $\alpha = .05$ was adopted for all tests. Anonymized data and scripts are provided on the OSF: https://osf.io/gbnvw/.

[3] Because standard errors were extremely small when a generalized linear mixed model with a gamma distribution was fitted, suggesting overfitting, we deviated from our pre-registered analyses by fitting a linear mixed model and transforming the VOT data toward a normal distribution.

2.5.2. Exploratory covariate analyses

To assess whether our results were influenced by severity and/or chronicity of the aphasic syndrome, we conducted exploratory analyses on the relationship of clinical outcome measures with ordinal position and session on VOTs and errors in the PWA group. These clinical outcome measures included (1) the Token-Test (TT^{AAT}, z-transformed percentile ranks), (2) the overall naming (NAM^{AAT}, z-transformed percentile ranks), and (3) the rating of semantic structure in the spontaneous-speech (SEM^{AAT}, score [0–5]). These measures are all subscores from the Aachen Aphasia Test (AAT; Huber et al., 1984). (4) From the LEMO battery (Stadie et al., 2013) the subtest requiring synonym judgment on word pairs with a semantic distractor ($SYNON^{LEMO}$, sum score) was included. Finally, (5) the months since onset of the lesion (MPO) and (6) the lesion size ($SIZE^{LES}$; see below) were used. To determine which of these measures explained significant variance beyond ordinal position and session while avoiding convergence issues, we adopted a stepwise procedure. We first included each measure individually as a continuous, mean-centered predictor to base models predicting VOTs and errors in the PWA group. The base VOT-model was log(VOT) ~ ordinal position*session + (1+ordinal position + session | subject) + (1 | category) and the base error-model was Error ~ ordinal position*session + (1+ordinal position | subject) + (1 | category). We then performed likelihood ratio tests comparing each covariate model to its respective base model to determine which measures significantly improved the model fit (see Appendix, Table D1 for a summary of comparisons). Finally, all covariates significantly improving the model fit were included in comprehensive models alongside ordinal position and session for VOTs and errors separately. The random structure was determined for the base models using the same procedure as for the main analyses and was kept consistent across covariate and comprehensive models.

2.5.3. Lesion behavior analysis

The lesion-behavior analysis was performed based on MRI ($n = 19$) and CT ($n = 1$) images of the PWA group. Using MRIcron (Rorden & Brett, 2000), lesions were manually delineated on each slice of the T1-images (FLAIR as reference) and the CT-images. The resulting lesion masks were then projected into the standard stereotactic space (MNI) by the 'clinical toolbox' (www.nitrc.org/projects/clinicaltbx/) in SPM12 (fil.ion.ucl.ac.uk/spm). The unified segmentation approach was applied (Ashburner & Friston, 2005) and estimation of normalization parameters was restricted to healthy tissue (Brett et al., 2001). Correlation analysis between behavior and lesion pattern was performed using a multivariate approach with support-vector-regression (Zhang et al., 2014; as implemented in the SVR- LSM toolbox, DeMarco & Turkeltaub, 2018). The multivariate approach takes inter-voxel-correlation into account and estimates the lesion-symptom map at all voxels simultaneously in a single model. To that end, all voxels significant at $p < .005$ were identified. These were then entered in a permutation-based-corrected cluster-level analysis. We assessed the stability of the approach by using the software vlsm2 (Bates et al., 2003), which showed qualitatively identical results. In all analyses only voxels lesioned in > 10% of participants (i.e., > 3 participants) were included in the analyses. Lesion size, age, and MPO were entered as covariates to factor out these unspecific effects. For lesion size, we used the diameter of a sphere corresponding in volume to the lesion ('DiaS').[4] Since plasticity related changes are more likely in the early chronic stage and become less likely with increasing chronicity, MPO were log-transformed. The analyses were performed for three parameters: (i) mean VOT across all trials ($_{log10}$VOT), (ii) mean cumulative semantic interference effect across all sessions (CSI effect), and (iii) changes in mean VOT from day 1 to day 2. Since lesions may increase but potentially also decrease the CSI effect, analyses were performed in both directions ('high values good' and 'high values bad' in the terminology of the SVR-LSM toolbox).

3. Results

3.1. Voice onset time (VOT)

The mean VOTs (response times) for the 5 ordinal positions and three days are shown in Fig. 1 for PWA (1A) and the control group (1B). Both groups exhibited a linear VOT increase by ordinal position, with an average increase of 21.08 ms ($SE = 5.20$) and plateaus at ordinal positions 2 or 4 (main effect of ordinal position $p < .001$; see Table 3 for statistical results).

Without accounting for dependencies between data points, the effect appeared larger in the PWA group. However, after controlling for these dependencies (summarySEwithin(.)-function; Morey, 2008), the effect was similar in both groups, although the PWA showed greater variance (PWA: $M_{day1} = 16.77$ ms, $SE_{day1} = 23.70$, $M_{day2} = 23.61$ ms, $SE_{day2} = 19.07$, $M_{day8} = 22.19$ ms, $SE_{day8} = 18.30$; Controls: $M_{day1} = 20.18$ ms, $SE_{day1} = 14.15$, $M_{day2} = 20.34$ ms, $SE_{day2} = 11.85$, $M_{day8} = 22.63$ ms, $SE_{day8} = 10.93$). Hence, the ordinal position effect did not differ significantly between groups (interaction ordinal position × group $p < .134$).

VOTs were shorter on days 2 and 8 compared to day 1 (main effects session $ps < .001$) and PWA, overall, were about 173 ms slower than controls (PWA: $M = 1352.93$; $SE = 7.16$; Controls: $M = 1179.71$, $SE = 4.24$; main effect group $p = .013$). The difference in VOT reduction between sessions did not differ significantly between groups (PWA group: day 2 *vs* day1 = 121.23 ms, day 8 *vs* day 1 = 136.20 ms; control group: day2 *vs* day1 = 86.21 ms and day8 *vs* day1 = 105.4 ms; interactions session × group $ps > .246$). Moreover, the linear increase by ordinal position did not differ significantly over time (interactions ordinal position × session (x group) > .090).

[4] Since damage to a specific cortical area and underlying white matter are common, especially in vascular lesions, the effect of the lesion volume is not linearly additive. Therefore, we use the diameter of a sphere corresponding to the volume. Moreover, the more commonly used total volume size in cm^3 is strongly skewed across the 20 participants, whereas 'DiaS' shows normal distribution.

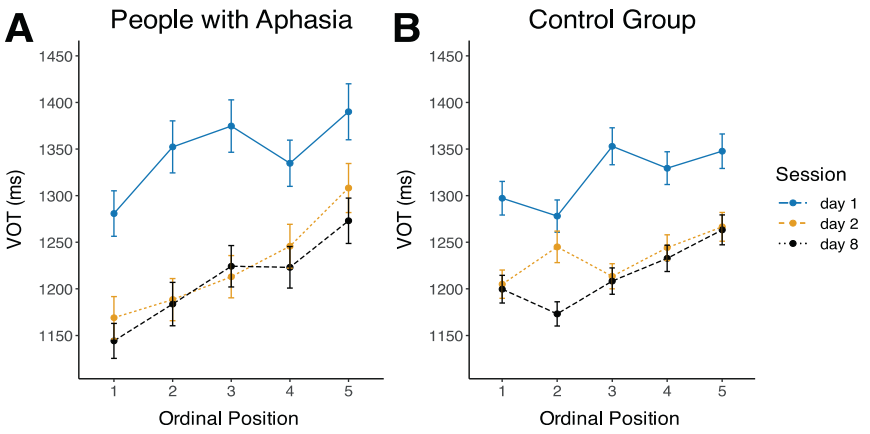

Fig. 1 – Mean naming latencies (VOTs) in milliseconds (ms) as a function of ordinal position and session for people with aphasia (PWA; A) and the control group (B). Participants were tested on days 1, 2, and 8 with the same experimental stimuli, presented in different orders. Mean VOTs were calculated across participants and semantic categories. Error bars show standard errors of the mean and are adjusted for within-participant designs (method by Morey, 2008; as implemented in summarySEwithin( ) function in R package Rmisc, v1.5.1, Hope, 2013).

Table 3 – Ordinal position (ord.pos.), session, and group have independent influences on log-transformed voice onset times (VOTs), as shown by a linear mixed effects regression model (LMM).

Model: log(VOT) ~ ord.pos.*session*group + (ord.pos.+session | subject) + (ord.pos.+group | category)					
Effect	Estimate	SE	t	p	
Intercept	7.12	.04	199.9	<.001	
Fixed main effect					
Ord.pos. (continuous)	.02	<.01	6.80	**<.001**	
Session (d2-d1)	−.08	.01	−7.01	**<.001**	
Session (d8-d1)	−.09	.01	−7.72	**<.001**	
Group (PWA-Control)	.16	.06	2.62	**.013**	
Fixed interaction					
Ord.pos. × session (d2-d1)	<.01	<.01	1.18	.238	
Ord.pos. × session (d8-d1)	<.01	<.01	.94	.349	
Ord.pos. × group	.01	<.01	1.53	.134	
Session (d2-d1) × group	−.03	.02	−1.18	.246	
Session (d8-d1) × group	−.02	.02	−.87	.391	
Ord.pos × session (d2-d1) × group	.01	.01	1.70	.090	
Ord.pos × session (d8-d1) × group	<.01	.01	.40	.691	
Random effects					
Effect	Variance	SD	Correlation		
By-subject random effects					
Intercept	.04	.19			
Ord.pos.	<.01	.01	.18		
Session (d2-d1)	<.01	.06	−.13	.24	
Session (d3-d1)	<.01	.06	−.49	.26	.64
By-category random effects					
Intercept	.01	.09			
Ord.pos.	<.01	<.01	.07		
Group (PWA-control)	<.01	.03	.08	.16	
Residual	.06	.25			
Model fit					
R^2	Marginal		Conditional		
	.07		.47		

Note. Number of participants = 40; number of categories = 24; total N = 12,455; *SE*: standard error. Ordinal position (ord.pos.) is entered to the model as a mean-centered continuous predictor; factors session and group are contrast coded such that days 2 and 8 are each compared to day 1 (d2-d1 and d8-d1) and that the PWA group is compared to the control group, the intercept being the grand mean. *p* values are estimated using the Satterthwaite approximation. Relevant *p*-values <.05 are shown in bold. R^2 was retrieved with the R-package *performance* (v0.10.4; Lüdecke et al., 2021).

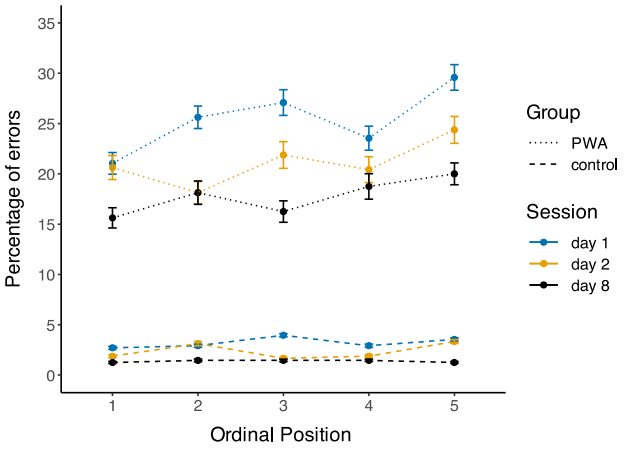

Fig. 2 – Mean error rates in percent as a function of ordinal position and session for people with aphasia (PWA) and the control group. Participants were tested on days 1, 2, and 8 with the same experimental stimuli, presented in different orders. Error rates describe all production errors, including paraphasia, replacements, neologisms, and null reactions, but no technical errors. Mean error rates were calculated across participants and semantic categories. Error bars show standard errors of the mean, adjusted for within-participant designs (method by Morey, 2008; as implemented in summarySEwithin() function in R package Rmisc, v1.5.1, Hope, 2013).

Data for each participant and each category are provided in the Appendix.

3.2. *Error rates*

We found a similar pattern in our exploratory analyses of error rates. As shown in Fig. 2, participants from both groups made fewer errors on days 2 and 8 when compared to day 1 (log-odds = −.36, $Z = -3.68$, $p < .001$ and log-odds = −.80, $Z = -7.17$, $p < .001$), PWA made more errors than controls (log-odds = 2.88, $Z = 6.47$, $p < .001$), and the linear increase in errors by ordinal position was a trend (log-odds = .07, $Z = 1.60$, $p = .109$). Again, there was no interaction between ordinal position, session, and group (all *p*s > .362).

Since error rates were rare in the control group (2.32 % across sessions), we additionally performed an analysis selectively on the PWA data (Table 4). In this analysis the ordinal position effect reached significance, indicating that PWA made more errors on the last than on the first ordinal positions (1–2% more errors per ordinal position, *SE* = 1.2–1.4 %, depending on the session; $p = .049$). Overall error rate decreased from day 1 to days 2 and 8 (day 2 *vs* day 1: decrease of errors by 4.61 %; day 8 *vs* day 1: decrease of errors by 7.85 %; *p*s < .001).

3.3. *Effects of adding covariates to the model*

When clinical outcome measures were added separately as covariates and compared to the base VOT model in PWA, only the token test and overall naming (TT^{AAT} and NAM^{AAT})

Table 4 – Ordinal position (ord.pos.) and session have independent influences on error rates in the PWA group, as shown by a generalized linear mixed effects regression model (GLMM) with a Binomial distribution.

*Model: Error ~ ord.pos.*session + (ord.pos. | subject) + (1 | category)*				
Effect	Log-Odds	*SE*	*Z*	*p*
Intercept	−1.86	.36	−5.19	<.001
Fixed main effect				
Ord.pos. (continuous)	.08	.04	1.97	**.049**
Session (d2-d1)	−.39	.08	−4.64	**< .001**
Session (d8-d1)	−.71	.09	−8.06	**< .001**
Fixed interaction				
Ord.pos. × session (d2-d1)	−.05	.06	−.78	.436
Ord.pos. × session (d8-d1)	−.04	.06	−.67	.501
Random effects				
Effect	Variance	*SD*	Correlation	
By-subject random effects				
Intercept	2.25	.59		
Ord.pos.	.02	1.50	.34	
By-category random effects				
Intercept	.35	.59		
Model fit				
R^2	Marginal		Conditional	
	.02		.45	

Note. Number of participants = 20; number of categories = 24; total *N* = 7,178; *SE*: standard error. Ordinal position (ord.pos.) is entered to the model as a mean-centered continuous predictor; the factor session is contrast coded such that days 2 and 8 are each compared to day 1 (d2-d1 and d8-d1), the intercept being the grand mean. *P*-values are estimated using the Wald *Z*-values. Relevant *p*-values < .05 are shown in bold. R^2 was retrieved with the R-package *performance* (v0.10.4; Lüdecke et al., 2021).

explained additional variance in the data, as indicated by significant likelihood ratio tests, $X_{TT\_AAT}(6) = 13.00$, $p = .043$ and $X_{NAM\_AAT}(6) = 13.61$, $p = .034$. The other measures, including SEM^{AAT} ($p = .068$), $SYNON^{LEMO}$ ($p = .096$), MPO ($p = .415$), and $SIZE^{LES}$ ($p = .522$), did not explain additional variance. To assess the combined impact, TT^{AAT}, NAM^{AAT}, and their interaction were then included in one comprehensive model, along with the predictors ordinal position and session and the random structure from the base PWA model (see Appendix D for an overview of all models and comparisons). In this comprehensive model, the main effects of ordinal position and session remained unaffected (all *p*s < .003). Unsurprisingly, better overall naming skills in the NAM^{AAT} were associated with faster VOTs in the experimental paradigm ($b = -.20$, $SE = .09$, $p = .044$). Other main effects/interactions were not significant, suggesting that the observed increase in VOTs by ordinal position and the decrease across sessions were independent from symptom severity as measured by the test scores. However, caution is warranted when interpreting these results, since 85 % of our PWA group had NAM^{AAT} scores in the upper 25 percentiles (see Table 1).

For error rates, models that included the token test (TT^{AAT}), the overall naming (NAM^{AAT}), the semantic structure of the spontaneous speech (SEM^{AAT}), or the lesion size ($SIZE^{LES}$) explained significantly more variance than the base model without covariates, as indicated by significant likelihood ratio tests, $X_{TT\_AAT}(6) = 22.21$, $p = .001$; $X_{NAM\_AAT}(6) = 30.17$, $p < .001$; $X_{SEM\_AAT}(6) = 12.75$, $p = .047$, $X_{SIZE\_LES}(6) = 14.75$, $p = .022$, but models including synonym judgement ($SYNON^{LEMO}$; $p = .226$) and MPO ($p = .331$) did not. The comprehensive model including TT^{AAT}, the NAM^{AAT}, SEM^{AAT}, and $SIZE^{LES}$ along with ordinal position and session (but without interaction between the covariates to avoid convergence problems) replicated the overall results (see Appendix D for summaries of all comparisons and models): Error rates increased with ordinal position and decreased from day 1 to days 2 and 8. Moreover, PWA with higher NAM^{AAT} scores made less errors than PWA with lower scores (Log-Odds = −1.13, $SE = .40$, $p = .005$). Notably, PWA with better spontaneous speech in the SEM^{AAT} exhibited a stronger CSI effect in errors, while those with poorer spontaneous speech had consistently high error rates across all ordinal positions (SEM^{AAT} × ordinal position: Log-Odds = .09, $SE = .05$, $p = .048$; see Figure D1 in Appendix for visualizations of significant interactions). Additionally, the ordinal position effect differed between PWA with lower versus higher scores in the Naming test only between days 1 and 8 (interactions ordinal position × session(day8-1) × NAM^{AAT}: Log-Odds = .29, $SE = .11$, $p = .008$), but not between days 2 and 1 ($p = .160$). Specifically, participants with lower Naming test scores (and overall, more errors) showed a particularly strong CSI effect on day 1, which decreased across sessions until, on day 8, it did not differ any more to that of PWA with higher scores. Finally, the CSI varied with performance in the TT^{AAT} only on day 8 (*vs* day 1; interaction ordinal position × session(day8-1) × TT^{AAT}: Log-Odds = −.22, $SE = .11$, $p = .046$). Specifically, on day 8, participants with lower TT^{AAT} scores exhibited a stronger CSI effect compared to those with higher scores. Again, due to ceiling effects in the test scores (see Table 1), these results need to be interpreted with caution.

3.4. *Exploratory lesion-behavior analysis*

The multivariate lesion-behaviour analysis yielded *separate* clusters for the overall naming latency and the size of the CSI effect. This underlines that magnitude of the CSI and overall naming speed are not caused solely by overall impairment and symptom severity in the PWA group. Results of the analyses are provided in Table 5, including peak and centre location, and the corresponding anatomical regions according to the Harvard Oxford atlas (https://nilearn.github.io/dev/modules/description/harvard_oxford.html). Correlating log-transformed mean voice onset time ($_{log10}$mVOT) with lesion site suggests that lesions in the middle occipital gyrus, parts of the supramarginal gyrus (SMG), and neighbouring parts of the inferior parietal lobe (IPL) increase naming latency (red volumes in Fig. 3B; log(VOT) ↑). The cluster in the middle occipital gyrus was just above the typical threshold ($p = .051$) while the more anterior cluster (SMG and IPL) only reached the level of a trend ($p = .093$). A second analysis used the variability in size of the CSI. This analysis suggests that lesions in a large cluster extending from the middle frontal gyrus to inferior parietal areas including parts of the pre- and post-central gyrus and the opercular cortex led to a larger CSI effect (blue volume in Fig. 3B). It means that participants with lesions in these areas showed a larger increase in VOTs with ordinal position. The cluster was significant with $p < .001$ after correction at the

Table 5 – Statistical results from the lesion-behavior analyses at the cluster level.

	cluster		peak				center			
	p	size/cm^2	x	y	z	region	x	y	z	region
VOT ↑ (a)	.051	4.5	−36	−68	11	lat.occip.^{inf}	−33	−80	12	lat.occip.^{sup}
VOT ↑ (b)	.093	2.3	−60	−37	48	SMG^{ant}	−50	−43	40	SMG^{post}
CSI ↑	<.001	22.0	−49	−8	17	^{cent}operc.	−34	−25	40	post.cent.

Note. The statistics and the peak as well as centre are provided in MNI space. "Peak" refers to the location with the maximal statistical power, while "centre" is the geometrical centre of the overall cluster. **x,y,z, region** = MNI-coordinates. "region" is provided according to the Harvard Oxford atlas (https://nilearn.github.io/dev/modules/description/harvard_oxford.html). **VOT ↑ (a)/(b)** = denote the two clusters in which a lesion correlates with an increase in voice onset time; **CSI ↑** = denotes the large cluster in which a lesion correlates with an increase in CSI effect; **lat.occip.^{inf}** and **lat.occip.^{sup}** = lateral occipital gyrus inferior/superior; **SMG^{ant}** and **SMG^{post}** = anterior/posterior supramarginal gyrus as a part of the inferior parietal lobe (IPL); **^{cent}operc.** = central opercular region; **post.cent.** = posterior central gyrus.

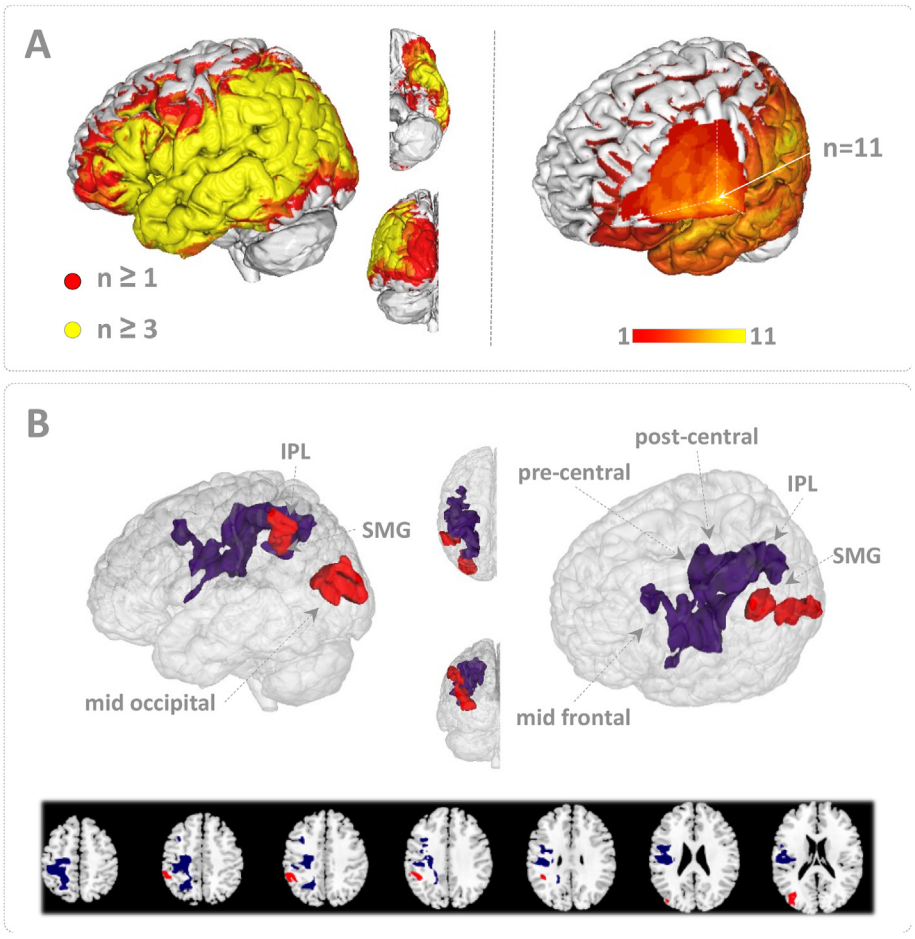

Fig. 3 – Lesion analysis. Panel A shows the overall lesion distribution of all 20 participants. *Left*: Colored areas denote that at least one participant had a lesion in the respective voxel. Yellow indicates that lesions of ≥ 3 participants overlapped in the area; this is the area which was analyzed. Red and orange colors indicate areas in which 1 or 2 participants had lesions. *Right*: Depicts the same data, but on a different scale. Here, the color code indicates lesion overlap in 1 (red) to 11 (yellow) participants. The maximal overlap ($n = 11$) was in the insular region. Panel B presents the results of the lesion-behavior analysis. For better visibility, two different angles and a tomographic representation of the same data are provided. Red volumes: lesions in this volume correlated with slower overall reaction times ($_{log10}$VOT in ms, $p < .005$ at voxel-level, $p < .1$ at cluster-correction-level). Blue volume: participants who were lesioned in this volume showed a larger CSI effect ($p < .005$ at voxel-level, $p < .05$ at cluster-correction-level). IPL = inferior parietal lobe; SMG = supramarginal gyrus.

cluster level. No clusters correlated with smaller CSI effects. Neither did the decrease in VOT from day one to two correlate with any specific lesion site.

4. Discussion

Investigating the cumulative semantic interference (CSI) effect over repetitive testing in a web-based continuous naming experiment, we show that this paradigm is feasible in people with an acquired lesion in the language network, leading to mild-moderate aphasia (PWA). This is remarkable since the effect rests not only on error-counting but also on subtle reaction time differences. With initial remote support, participants showed sufficient skills to manage the web-based setting and adhered well to the study protocol with tests on three days over one week. PWA showed a robust CSI effect in naming latencies, which did not statistically differ from the matched neurotypical controls. The CSI effect in naming latencies was of similar size in both groups and compared to previous studies, suggesting overall similar inhibitory effects during lexical selection. In the PWA group, latency effects were more variable and were mirrored by an increase in error rates, demonstrating relatively strong interference for this group that not only resulted in slower naming of correct target names, but also caused more errors during lexical selection. In line with the majority of continuous naming studies with neurotypical adults, we found no effects of ordinal position in error rates of the control group (e.g., Howard et al., 2006; Rose & Abdel Rahman, 2017; Schnur, 2014). Because the increase in naming latencies and error rates does not indicate a speed-accuracy tradeoff, we take this pattern to be due to an enhanced vulnerability of lexical-semantic processing in PWA compared to the control group.

Moreover, the CSI effect was unaffected by repeated exposure to the task and material, suggesting that the CSI effect dissipates until the next experimental session. The overall decrease in naming latencies especially from day 1 to day 2 did not alter the CSI effect. This overall decrease in naming latency likely indicates unspecific and overall task familiarization effects. Confirming the versatility of the web-based procedure to monitor semantic context effects, an exploratory lesion-behavior-analysis indicated that overall naming latencies and the magnitude of the CSI effect correlated with lesions in different brain regions. In line with previous reports in lesion-site dependent variable semantic context effects in PWA (Nappo et al., 2022; Pino et al., 2022), the results highlight the relevance of subtle diagnostic tools to find optimal training strategies. To our knowledge, this is the first study to investigate the CSI effect under repetitive confrontation with the same paradigm and material.

4.1. *Robust CSI can be demonstrated web-based in PWA when technical support is provided*

When PWA named different members of close semantic categories (e.g., category 'fish': 'eel'/'dolphin'/'goldfish'/'shark'/'ray'), their naming latencies increased by about 21 ms on average with each additional member being named. Similar to previous work, this effect was elicited with category members being separated by two to eight unrelated items (see also Schnur, 2014). The effect was paralleled in a linear increase in error rates by about 6–7 errors per ordinal position. Most errors were omissions, which leaves open the question whether strongly increased response times may have led to these time-out errors. There was, however, an increase across error types, including semantic errors. The findings complement previous research that established interference effects in PWA (e.g., Belke & Stielow, 2013; Biegler et al., 2008; Piai & Knight, 2018; Pino et al., 2022; Riès et al., 2015; Schnur et al., 2006; Scott & Wilshire, 2010). Albeit evidence for the CN-paradigm is rare, the few available studies report roughly similar effect sizes of 16–21 ms in latencies (but 69 ms in preprint by Lorenz et al., 2021; Nappo et al., 2022; Riès et al., 2015) and .5–8% in errors (Harvey et al., 2019; Nappo et al., 2022). Like in the present data, the size of the effect varied greatly between participants (Nappo et al., 2022).

These data, to our knowledge, present the first web-based reaction time sensitive investigation in PWA (Stark et al., 2022; Vogt et al., 2022 for evidence on young, healthy participants), presenting unique challenges, such as higher frustration susceptibility, which might negatively affect response times and study adherence. Through careful pre-selection of PWA and controls, natural to clinical studies, and telephone support to address task- and technique-related frustration, we achieved high adherence to the task and possibly reduced the risk of distraction in the web-based design (Bürki & Vasishth, 2024). This approach may also help to mitigate issues of social isolation common in PWA (Cruice et al., 2006, 2021; Kurland et al., 2018). One key takeaway from the current study is that, with minor adaptations to established protocols (Stark et al., 2022) and a slightly increased involvement of the experimenter, web-based settings can help bridge the research gap on outpatient PWA. However, due to the large variability in PWA data, laboratory-based studies may be needed for identifying relevant subgroups.

4.2. *The CSI effect is stable over repetitive testing*

Although overall naming latencies and error rates decreased from day 1 to day 2 (101 ms) and from day 1 to day 8 (119 ms), the CSI effect in both groups remained largely unaffected by the repeated testing (on average: 21 ms). The overall increase in naming speed may result from several factors including general familiarization with the experimental procedure and the material, including picture recognition. The stability across the 2nd and 3rd test session suggests such unspecific effects.

Regarding the stability of the CSI effect over repetitive testing, we demonstrate that the mechanism underlying the effect does not outlast retesting on the consecutive day or after a week. This speaks for a 'decay' of the effect, as has been demonstrated for long lags between members of a category (> 8 intervening items; Schnur, 2014) and was also suggested by Damian and Als (2005). Our design does not allow for more concise predictions regarding the time course of the decay. However, our results support the claim that models of the CSI must integrate more complex mechanisms of fading connection strengths in addition to persistent (Schnur, 2014), experience driven adjustments of the connection strength (incremental learning: Belke & Stielow, 2013; Howard et al., 2006; Oppenheim et al., 2010).

Future studies may further investigate whether and how training interventions alter the CSI effect. In this vein, an interesting approach targets executive control of linguistic production (Bruehl et al., 2023). The demonstration of a stable CSI over consecutive testing, which can be assessed web-based may be of relevance for further research of such intervention schemes. The paradigm may not become a routine clinical tool, but similarly to the blocked cyclic naming task, it might be *"a useful diagnostic tool in identifying and specifying disorders of lexical access."* (Belke & Stielow, 2013, p. 2154).

4.3. *PWA and control participants show similar interference effects*

Despite their presumed 'noisy access' (Harvey & Schnur, 2015) to the mental lexicon, the here described average interference effect in naming latencies in PWA was remarkably similar to the control participants. The fact that the cumulative semantic interference effect was statistically robust in PWA although response times and error rates clearly decreased across sessions implies that the CSI effect does not correlate with the overall difficulties in word retrieval. Regarding the large variability between PWA, 'noisy access' will not only affect target words but also the categorical competitors leading to varying degrees of coactivation within the PWA group. Hence, semantic paraphasia may result from (pathologically) weak activation of the target and/or from (pathologically) strong co-activation within the cohort. The net CSI effect must be considered as the summation of different factors for each PWA participant, including increased and decreased co-activation. In fact, a seemingly paradoxical decrease in interference has been demonstrated for lesions in posterior

temporal areas in a picture-word interference paradigm (Pino et al., 2022) and the continuous naming paradigm (Nappo et al., 2022). The latter study reports 'paradoxical' facilitation (decrease in VOT with ordinal position) in some participants. In the data presented here, the variance in the CSI trajectories was considerably larger in the PWA compared to controls (standard errors of ~20 vs 12 ms per session for raw data; see figure in Appendix C). The variance is evidenced in the random effects structure of the model. While some PWA exhibited a particularly strong CSI effect, a few showed inconsistent trajectories with a net CSI effect close to zero ($n = 2$). The larger variability suggests that the CSI magnitude in PWA results from an interplay between altered activation or coactivation of the target lemma and its competitors, and/or an altered inhibition-/selection-process.

Besides the surprisingly homogeneous latency effects in the different cohorts, error rates speak for partial interrelations between lexical retrieval difficulties and the CSI: (i) PWA showed a steep increase in error rates by ordinal position, which was not seen in control participants, who had very low error rates. Because null reactions significantly increased with exposure of additional category members in PWA only, the error-based CSI effect cannot be interpreted fully independently from the latency-based CSI effect. If we assume that at least some of the trials in which no response was recorded within 3 s actually indicate extremely long response times, the differences in the latency effects may be underestimated. Indeed, some recordings contained only partial naming responses and were hence considered as errors. However, since only 2.6 % of PWA responses contained partial recordings (i.e., participants started to name the target, but the recording ended before they finished), extremely long RTs alone are unlikely to account for the CSI effect in errors. (ii) The variability of naming latencies was larger in the PWA than in the control group, as indicated by larger standard errors and more diverse individual trajectories across sessions (Appendix C). Thus, particularly strong or weak co-activation of some cohort members may have led to strong interference or facilitation effects, respectively (Abdel Rahman & Melinger, 2019; Chen & Mirman, 2012; Nappo et al., 2022; Pino et al., 2022). (iii) Covariate analyses additionally suggest that PWA with more severe deficits (as judged by the AAT including the Token Test) showed larger error-based CSI effects. Thus, rather than unaffected incremental learning, our data suggest a more diverse effect in the PWA group that *on average* led to similar effects in both groups.

4.4. *Exploratory lesion behavior analyses suggests that the interference effect and naming latencies correlate with lesions in different brain regions*

Proceeding from the finding of a larger variability in susceptibility to semantic interference in PWA, we performed an exploratory analysis correlating overall naming skills and the size of the CSI with lesion site. We found that overall naming skills and size of the CSI also dissociated regarding the area in which lesions modulated the respective parameter. Patients with lesions in a large cluster extending from the middle frontal gyrus to inferior parietal areas and including parts of the pre- and post–central gyrus and the opercular cortex showed a larger CSI effect. Patients with lesions in the middle occipital gyrus, parts of the supramarginal gyrus (SMG) and parts of the inferior parietal lobe (IPL) showed higher mean VOT. While these results are exploratory in nature, they complement previous research that identified different lesion patterns for overall naming effects, interference effects in the picture-word interference task, and associative facilitation (Pino et al., 2022), for interference in naming and comprehension (Harvey & Schnur, 2015), and for naming latencies as well as error-proneness (Schnur et al., 2006). It is further in line with research in uncompromised speakers, suggesting that the CSI effect relies on functional connectivity in a general cognitive control network (Canini et al., 2016) rather than on the left IFG alone (Britt et al., 2016).

4.5. *Implications for a framework of how semantic context effects are altered in people with aphasia*

Our results may be of note to the more general question of how semantic context effects are altered in people with aphasia. Previous studies in PWA using the paradigms of blocked cyclic naming (Nappo et al., 2022; Python et al., 2023; Schnur et al., 2006), picture word interference (Pino et al., 2022; Python et al., 2018; van Scherpenberg et al., 2021), and continuous naming (Harvey et al., 2019; Nappo et al., 2022) converge on alterations due to both, altered activation of the lemma in the mental lexicon and altered threshold setting for the selection of competing lemmas (Anders et al., 2017). Although studies agree that the lexicon is more sensitive to (posterior) temporal lesions while threshold alterations are mostly caused by lesions in the frontal language areas, both processes interact. Therefore, the net-effect of an increase or decrease in interference will depend on a complex interplay between lesion-site, compensation strategy, and specific task requirements. For the current study using a CN-paradigm, we propose a framework sketched in Fig. 4. Besides lemma-activation and -selection, the framework posits that alterations in the incremental learning process (Oppenheim et al., 2010) may be affected by lesions in the language network. We may highlight that the high variability of CSI effects in our PWA group and the dissociation of lesion correlates for the CSI effect on the one hand and overall naming abilities on the other hand should be considered but a first exploratory step towards a better understanding of the correlations with overall patholinguistic and lesion profile.

4.6. *Perspectives and limitations*

We show that traces of incremental co-activation of a semantic category dissipate relatively fast (within one day) and are not re-activated by repetitive exposure to the same stimulus material. By contrast, overall naming latency decreases with familiarization with the paradigm and material, especially in patients with aphasia.

When drawing conclusions from between session and between group effects in particular, possible confounds must be considered, especially in web-based settings where technical setups, study populations, and distractions may differ from the typical highly controlled lab environment (e.g., Anwyl-Irvine et al., 2021; Bridges et al., 2020; Bürki & Vasishth, 2024;

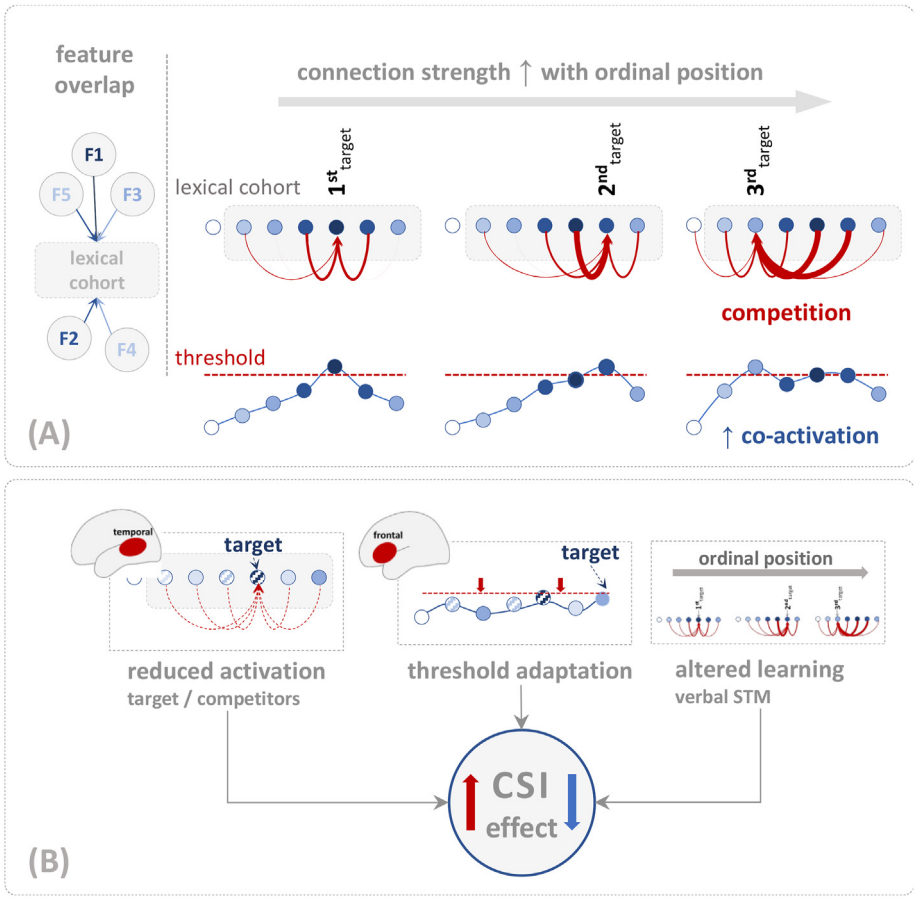

Fig. 4 – (A) illustrates the assumed cognitive processes underlying the cumulative semantic interference (CSI) effect. *left:* Featural similarities (F1–F5, e.g., 'animal', 'has 4 legs', 'furry' …) within semantic categories (e.g., 'pets') lead to coactivation of the corresponding lexical representations (lemmas); e.g., when lemma 'dog' is activated, 'cat' and 'horse' become co-activated (features: 'animal', 'has 4 legs'). *right:* coactivation eases lexical access but also increases competition for selection of correct lemma (gray box represents lexical cohort). Each naming of a category member strengthens the connection between conceptual representation ↔ lemma, yielding stronger competition for the next category member to be named with each new member (ordinal position). This makes selection more difficult, that is slower and more prone to error. (B) sketches possible alterations of these processes due to acquired lesions in the language network. The *left box* illustrates that 'noisy access' to the lemmas reduces overall naming speed and may decrease the activation difference between target and non-target cohort members within the mental lexicon (affected by temporo-parietal lesions). Whether this increases or decreases competition depends on the specific deficit profile. The *center box* illustrates that besides effects within the mental lexicon, (dys)functional alterations of the threshold at which a lexical candidate's activation is sufficient to be selected impacts on naming speed and accuracy. This process is assumed to depend on inferior frontal brain areas. The *right box* highlights that the change in connection strength—essential for the buildup of the CSI—can also be altered. Verbal short-term and working memory capacities may be relevant for the efficiency of the 'incremental learning' mechanism. Note that all three factors sketched can lead to increases or decreases in the net CSI effect.

Mathôt & March, 2021). In our study, the study population and home distractions should be relatively controlled given that participants were pre-selected and supervised via telephone. In previous studies, technical setups usually biased response time not more than 10 ms (e.g., Anwyl-Irvine et al., 2021), but are confounded with the group differences here. However, given that 50 % of the PWA were tested in the clinic on similar or identical setups,[5] but the variance in the data was still smaller in the control group, it is unlikely that technical differences alone contribute to group differences. Moreover, exploratory analyses showed that test location (home or clinic) did not affect ordinal position or session effects in the PWA although the variance was somewhat reduced when tested in the clinic. Although the large variability in our data likely reflects real variations beyond technical artifacts, small changes may have gone undetected.

While the observed interference effects were robust over repetitive testing and between groups, we may point out that participants suffered from mild to moderate aphasia, as indicated by close-to-ceiling effects in some of the clinical outcome measures. More severe aphasia may strongly increase the

[5] An overview of the hardware/software setups is provided as supplementary material on the OSF.

difficulties to partake in this or similar experimental paradigms (Hart, 2017), especially in web-based setups. Regarding the exploratory lesion analyses, the major limitation is the small number of participants. Moreover, larger studies addressing this issue may need to apply tools regarding the network-connections rather than single hubs of the language-network.

Interestingly, naming latencies did not increase (or even decreased) for the fourth member of a category (ordinal position 4). This as yet unexplained effect observed in both groups, especially on day 1 is similar to reports in neurotypical individuals at ordinal positions 4 or 5 (Costa et al., 2009; Kuhlen & Abdel Rahman, 2017; Rose & Abdel Rahman, 2016; Stark et al., 2022). Tentatively, such a 'saturation-effect' might stem from the fact that close semantic (ad-hoc) categories, such as 'seating furniture', typically have four to five members. This would imply that after the 4th or 5th category–exemplar the ad-hoc category is updated to a somewhat broader category.

5. Conclusion

People with lesions to the left-lateralized language-network regularly experience deficits of lexical retrieval resulting in slow and erroneous naming. Therefore, repetitive confrontational naming is a common therapeutic focus in the attempt to alleviate aphasic symptoms. The continuous naming (CN) paradigm highlights that besides facilitatory effects of semantic context, competition between cohort members increases latency and errors, supplying insight into the structure of the mental lexicon. To the best of our knowledge, the current study is the first to investigate the stability of cumulative semantic interference over repetitive testing over one week. Moreover, we show that PWA can be tested in a web-based setting repetitively and show a stable cumulative interference effect. Despite the higher variability between individuals in the PWA-group, the individual size of the effect was also largely stable across sessions. Our results open the perspective to test changes in the mental lexicon in response to an intervention, using the CSI as an indicator of enhanced or decreased lexical interference. Web-based testing additionally allows for long-term follow-up of patients without the logistical effort often involved in transporting PWA with co-morbid physical limitations to a research center or clinic. Extending more psycholinguistic research to people with acquired language disorders can enrich both current theories and therapeutic interventions.

TOP statement

We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. Anonymized data and analyses code can be found here: https://osf.io/gbnvw/. The conditions of our ethics approval do not permit public archiving of individual MRI data, as they cannot be fully anonymized. Readers seeking access to the data should contact the last author (HO). Access will be granted to named individuals in accordance with ethical procedures governing the reuse of sensitive data. Specifically, requestors must meet the following conditions to obtain the data: completion of a formal data sharing agreement; use of the data only for non-commercial research purposes. The procedure, hypotheses, and analyses were preregistered on the Open Science Framework: https://osf.io/gkqy9. Legal copyright restrictions do not permit us to publicly archive the full set of photograph stimuli used in this experiment, but a complete list of the stimuli used is provided in the Appendix.

CRediT authorship contribution statement

Kirsten Stark: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Formal analysis, Data curation. **Marcus Töpel:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Frank Regenbrecht:** Writing – review & editing, Resources, Investigation, Conceptualization. **Cornelia van Scherpenberg:** Writing – review & editing, Resources, Methodology. **Rasha Abdel Rahman:** Writing – review & editing, Supervision, Conceptualization. **Hellmuth Obrig:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Funding acquisition, Formal analysis, Conceptualization.

Author note

We have no conflicts of interest to disclose. Anonymized data, source code, and supplemental material can be found here: https://osf.io/gbnvw/.

The procedure, hypotheses, and analyses were preregistered on the OSF: https://osf.io/gkqy9.

Open practices section

The study in this article has earned Preregistered badge for transparent practices. The preregistered studies are available at: https://osf.io/gkqy9.

Funding

No funding was received for conducting the study. During the preparation of this paper, KS was funded by the Charité - Universitätsmedizin Berlin, Einstein Center for Neurosciences, Berlin, Germany. The authors have no relevant financial or non-financial interests to disclose.

Standard STS statement

This is among the last of the manuscripts which is passing through as the initial checklist was in place before STS and STR was introduced.

Appendix/Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2024.11.019.

REFERENCES

Abdel Rahman, R., & Melinger, A. (2019). Semantic processing during language production: An update of the swinging lexical network. *Language, Cognition and Neuroscience, 34*(9), 1176–1192. https://doi.org/10.1080/23273798.2019.1599970

Abel, S., Willmes, K., & Huber, W. (2007). Model-oriented naming therapy: Testing predictions of a connectionist model. *Aphasiology, 21*(5), 411–447. https://doi.org/10.1080/02687030701192687

Anders, R., Llorens, A., Dubarry, A.-S., Trébuchon, A., Liégeois-Chauvel, C., & Alario, F.-X. (2019). Cortical dynamics of semantic priming and interference during word production: An intracerebral study. *Journal of Cognitive Neuroscience, 31*(7), 978–1001. https://doi.org/10.1162/jocn_a_01406

Anders, R., Riès, S., Van Maanen, L., & Alario, F.-X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology, 34*(1–2), 1–20. https://doi.org/10.1080/02643294.2017.1282447

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods, 53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Asghar, I., Egaji, O. A., & Griffiths, M. (2021). An overview of the digital solutions for helping people with aphasia through bibliometric analysis. *ENeurologicalSci, 22*, Article 100311. https://doi.org/10.1016/J.ENSCI.2021.100311

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *Neuroimage, 26*(3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018

Bürki, A., & Vasishth, S. (2024). Assessing effect sizes, variability, and power in the web-based study of language production. https://doi.org/10.21203/rs.3.rs-4520462/v1.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron, 47*(6), 907–918. https://doi.org/10.1016/j.neuron.2005.07.023

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronker, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience, 6*(5), 448–450. https://doi.org/10.1038/nn1050

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *The Quarterly Journal of Experimental Psychology: QJEP, 66*(11), 2135–2160. https://doi.org/10.1080/17470218.2013.775318

Best, W., Nickels, L., & Williamson, L. (2005). *A coding system for picture naming responses with a focus on semantic errors.* In Unpublished manuscript.

Bevivino, D., Hemforth, B., & Turco, G. (2023). Data from online production experiments and challenges for collecting good-quality recordings for proosodic analyses. https://shs.hal.science/halshs-04337979.

Biegler, K. A., Crowther, J. E., & Martin, R. C. (2008). Consequences of an inhibition deficit for word production and comprehension: Evidence from the semantic blocking paradigm. *Cognitive Neuropsychology, 25*(4), 493–527. https://doi.org/10.1080/02643290701862316

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48*(3), 468–488. https://doi.org/10.1016/S0749-596X(02)00503-X

Bloem, I., van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language, 51*(2), 307–323. https://doi.org/10.1016/j.jml.2004.05.001

Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer (6.1.16)*. https://www.fon.hum.uva.nl/praat/.

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews, 2016*(6). https://doi.org/10.1002/14651858.CD000425.pub4

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods, 23*(3), 389–411. https://doi.org/10.1037/met0000159

Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., Huber, W., Willmes, K., Ringelstein, E. B., Haeusler, K. G., Abel, S., Glindemann, R., Domahs, F., Regenbrecht, F., Schlenck, K. J., Thomas, M., Obrig, H., de Langen, E., Rocker, R., … Bamborschke, S. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: A randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *Lancet (London, England), 389*(10078), 1528–1538. https://doi.org/10.1016/S0140-6736(17)30067-3

Brett, M., Leff, A. P., Rorden, C., & Ashburner, J. (2001). Spatial normalization of brain images with focal lesions using cost function masking. *Neuroimage, 14*(2), 486–500. https://doi.org/10.1006/nimg.2001.0845

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, 1–29. https://doi.org/10.7717/peerj.9414

Britt, A. E., Ferrara, C., & Mirman, D. (2016). Distinct effects of lexical and semantic competition during picture naming in younger adults, older adults, and people with aphasia. *Frontiers in Psychology, 7*(JUN). https://doi.org/10.3389/fpsyg.2016.00813

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

Canini, M., Della Rosa, P. A., Catricalà, E., Strijkers, K., Branzi, F. M., Costa, A., & Abutalebi, J. (2016). Semantic interference and its control: A functional neuroimaging and connectivity study. *Human Brain Mapping, 37*(11), 4179–4196. https://doi.org/10.1002/hbm.23304

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of

facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417–430. https://doi.org/10.1037/a0027175

Cherney, L. R., Lee, J. B., Kim, K.-Y. A., & van Vuuren, S. (2021). Web-based oral reading for language in aphasia (web ORLA®): A pilot randomized control trial. *Clinical Rehabilitation, 35*(7), 976–987. https://doi.org/10.1177/0269215520988475

Conroy, P., Sage, K., & Ralph, M. L. (2009). Improved vocabulary production after naming therapy in aphasia can gains in picture naming generalise to connected speech. *International Journal of Language and Communication Disorders, 44*(6), 1036–1062. https://doi.org/10.3109/13682820802585975

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*. https://doi.org/10.1016/j.actpsy.2023.104073

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences of the United States of America, 106*(50), Article 21442. https://doi.org/10.1073/PNAS.0908921106

Cruice, M., Woolf, C., Caute, A., Monnelly, K., Wilson, S., & Marshall, J. (2021). Preliminary outcomes from a pilot study of personalised online supported conversation for participation intervention for people with aphasia. *Aphasiology, 35*(10), 1293–1317. https://doi.org/10.1080/02687038.2020.1795076

Cruice, M., Worrall, L., & Hickson, L. (2006). Quantifying aphasic people's social lives in the context of non-aphasic peers. *Aphasiology, 20*(12), 1210–1225. https://doi.org/10.1080/02687030600790136

Döring, A.-L., Abdel Rahman, R., Zwitserlood, P., & Lorenz, A. (2022). Cumulative semantic interference is blind to morphological complexity and originates at the conceptual level. *Plos One, 17*(6), Article e0268915. https://doi.org/10.1371/journal.pone.0268915

Damian, M. F., & Als, L. C. (2005). Long-Lasting Semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(6), 1372–1384. https://doi.org/10.1037/0278-7393.31.6.1372

Damian, M. F., & Bowers, J. S. (2003a). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review, 10*(1), 111–117. https://doi.org/10.3758/BF03196474

Damian, M. F., & Bowers, J. S. (2003b). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review, 10*(1), 111–117. https://doi.org/10.3758/BF03196474

de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a web browser? Comparing response times collected with JavaScript and psychophysics toolbox in a visual search task. *Behavior Research Methods, 48*(1), 1–12. https://doi.org/10.3758/s13428-015-0567-2

de Zubicaray, G. I., Miozzo, M., Johnson, K., Schiller, N. O., & McMahon, K. L. (2012). Independent distractor frequency and age-of-acquisition effects in picture–word interference: fMRI evidence for post-lexical and lexical accounts according to distractor type. *Journal of Cognitive Neuroscience, 24*(2), 482–495. https://doi.org/10.1162/jocn_a_00141

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science, 23*(4), 517–542. https://doi.org/10.1207/s15516709cog2304_6

DeMarco, A. T., & Turkeltaub, P. E. (2018). A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. *Human Brain Mapping, 39*(11), 4169–4182. https://doi.org/10.1002/hbm.24289

Dyson, L., Morgan, J., & Herbert, R. (2021). Semantic processing in aphasia: Evidence from semantic priming and semantic interference. *Language, Cognition and Neuroscience, 36*(4), 491–516. https://doi.org/10.1080/23273798.2020.1849749

Efstratiadou, E. A., Papathanasiou, I., Holland, R., Archonti, A., & Hilari, K. (2018). A systematic review of semantic feature analysis therapy studies for aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR, 61*(5), 1261–1278. https://doi.org/10.1044/2018_JSLHR-L-16-0330

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *Plos One, 16*(10), Article e0258908. https://doi.org/10.1371/journal.pone.0258908

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a stroop-like task. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 42*(6), 790–796. https://doi.org/10.1016/S0010-9452(08)70419-2

Fuhrmeister, P., & Bürki, A. (2022). Distributional properties of semantic interference in picture naming: Bayesian meta-analyses. *Psychonomic Bulletin & Review, 29*(2), 635–647. https://doi.org/10.3758/S13423-021-02016-6

Gavard, E., & Ziegler, J. C. (2024). Semantic and syntactic predictions in reading aloud: Are good predictors good statistical learners? *Journal of Cognition, 7*(1). https://doi.org/10.5334/joc.363

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology. Human Perception and Performance, 10*(5), 640–654. https://doi.org/10.1037//0096-1523.10.5.640

Green, P., & MacLeod, C. J. (2016). Simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Hart, P. (2017). *Cumulative semantic interference in people with aphasia*.

Harvey, D. Y., & Schnur, T. T. (2015). Distinct loci of lexical and semantic access deficits in aphasia: Evidence from voxel-based lesion-symptom mapping and diffusion tensor imaging. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 67*, 37–58. https://doi.org/10.1016/j.cortex.2015.03.004

Harvey, D. Y., Traut, H. J., & Middleton, E. L. (2019). Semantic interference in speech error production in a randomized continuous naming task: Evidence from aphasia. *Language, Cognition and Neuroscience, 34*(1), 69. https://doi.org/10.1080/23273798.2018.1501500

Haslam, C., Moss, Z., & Hodder, K. (2010). Are two methods better than one? Evaluating the effectiveness of combining errorless learning with vanishing cues. *Journal of Clinical and Experimental Neuropsychology, 32*(9), 973–985. https://doi.org/10.1080/13803391003662686

He, J., Frances, C., Creemers, A., & Brehm, L. (2024). Effects of irrelevant unintelligible and intelligible background speech on spoken language production. *The Quarterly Journal of Experimental Psychology: QJEP*. https://doi.org/10.1177/17470218231219971

He, J., Meyer, A. S., Creemers, A., & Brehm, L. (2021). Conducting language production research online: A web-based study of semantic context and name agreement effects in multi-word production. *Collabra: Psychology, 7*(1). https://doi.org/10.1525/collabra.29935

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henseler, I., Mädebach, A., Kotz, S. A., & Jescheniak, J. D. (2014). Modulating brain mechanisms resolving lexico-semantic interference during word production: A transcranial direct current stimulation study. *Journal of Cognitive Neuroscience, 26*(7), 1403–1417. https://doi.org/10.1162/jocn_a_00572

Hilbig, B. E. (2016). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods, 48*(4), 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hope, R. M. (2013). *Rmisc: Ryan miscellaneous*.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*(3), 464–482. https://doi.org/10.1016/J.COGNITION.2005.02.006

Huber, W., Poeck, K., & Willmes, K. (1984). The aachen aphasia test. *Advances in Neurology, 42*, 291–303. https://pubmed.ncbi.nlm.nih.gov/6209953/.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865).

Jørgensen, L., Engstad, T., & Jacobsen, B. K. (2002). Higher incidence of falls in long-term stroke survivors than in population controls: Depressive symptoms predict falls after stroke. *Stroke; a Journal of Cerebral Circulation, 33*(2), 542–547. https://doi.org/10.1161/HS0202.102375

Jescheniak, J. D., Hantsch, A., & Schriefers, H. (2005). Context effects on lexical choice and lexical activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 905–920. https://doi.org/10.1037/0278-7393.31.5.905

Jescheniak, J. D., Matushanskaya, A., Mädebach, A., & Müller, M. M. (2014). Semantic interference from distractor pictures in single-picture naming: Evidence for competitive lexical selection. *Psychonomic Bulletin & Review, 21*(5), 1294–1300. https://doi.org/10.3758/s13423-014-0606-5

Kandel, M., Wyatt, C. R., & Phillips, C. (2022). Agreement attraction error and timing profiles in continuous speech. *Glossa Psycholinguistics, 1*(1). https://doi.org/10.5070/G601157

Khan, M. (2020). RecordRTC. https://github.com/muaz-khan/RecordRTC.

Kiran, S., Roches, C. Des, Balachandran, I., & Ascenso, E. (2014). Development of an impairment-based individualized treatment workflow using an iPad-based software platform. *Seminars in Speech and Language, 35*(1), 38–50. https://doi.org/10.1055/s-0033-1362995

Kuhlen, A. K., & Abdel Rahman, R. (2017). Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition, 166*, 94–106. https://doi.org/10.1016/j.cognition.2017.05.024

Kurland, J., Liu, A., & Stokes, P. (2018). Effects of a tablet-based home practice program with telepractice on treatment outcomes in chronic aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR, 61*(5), 1140–1156. https://doi.org/10.1044/2018_JSLHR-L-17-0277

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Langhorne, P., Coupar, F., & Pollock, A. (2009). Motor recovery after stroke: A systematic review. *The Lancet. Neurology, 8*(8), 741–754. https://doi.org/10.1016/S1474-4422(09)70150-4

Lee, J., Keen, A. D., Farr, E., & Christ, S. (2023). TelePriming sentence production in aphasia. *Frontiers in Human Neuroscience, 17*. https://doi.org/10.3389/fnhum.2023.1274620

Leiner, D. J. (2019). *SoSci survey* (Version 3.1.06) https://www.soscisurvey.de.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3*, 223–232.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Li, K. K., Schwarz, J., Sim, J. H., Zhang, Y., Buchanan-Worster, E., Post, B., & McDougall, K. (2022). Recording and timing vocal responses in online experimentation. In *Proceedings of the Annual Conference of the International speech Communication association* (pp. 4053–4057). INTERSPEECH. https://doi.org/10.21437/Interspeech.2022-10697. 2022-September.

Lorenz, A., Döring, A., Scherpenberg, C. Van, Pino, D., Abdel Rahman, R., & Obrig, H. (2021). On the relation of semantic context effects in picture naming and semantic categorization: Evidence from aphasia. *EasyChair Preprint, 6649*.

Lorenz, A., Pino, D., Jescheniak, J. D., & Obrig, H. (2022). On the lexical representation of compound nouns: Evidence from a picture-naming task with compound targets and gender-marked determiner primes in aphasia. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior, 146*, 116–140. https://doi.org/10.1016/J.CORTEX.2021.09.019

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Mathôt, S., & March, J. (2021). *Conducting linguistic experiments online with OpenSesame and OSWeb*. https://doi.org/10.31234/osf.io/wnryc

Mcconnell, K., Hintz, F., & Meyer, A. S. (2024). Individual differences in online research: Comparing lab-based and online administration of a psycholinguistic battery of linguistic and domain-general skills. *PsyArXiV Preprints*. https://doi.org/10.31234/osf.io/fubn5

Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation, 22*(2), 138–168. https://doi.org/10.1080/09602011.2011.639619

Mooijman, S., Schoonen, R., Ruiter, M. B., & Roelofs, A. (2023). Voluntary and cued language switching in late bilingual speakers. *Bilingualism*. https://doi.org/10.1017/S1366728923000755

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Nappo, R., Galati, G., Bureca, I., & Romani, C. (2022). Semantic interference and facilitation in picture naming: The effects of type of impairment and compensatory strategies. *Cognitive Neuropsychology, 39*(5–8), 325–355. https://doi.org/10.1080/02643294.2023.2189004

Navarrete, E., Del Prato, P., & Mahon, B. Z. (2012). Factors determining semantic facilitation and interference in the cyclic naming paradigm. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00038

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology, 16*(10–11), 935–979. https://doi.org/10.1080/02687030244000563

Nozari, N., & Hepner, C. R. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology, 36*(5–6), 193–207. https://doi.org/10.1080/02643294.2018.1476335

Nunn, K., Vallila-Rohter, S., & Middleton, E. L. (2023). Errorless, errorful, and retrieval practice for naming treatment in aphasia: A scoping review of learning mechanisms and treatment ingredients. *Journal of Speech, Language, and Hearing Research: JSLHR, 66*(2), 668–687. https://doi.org/10.1044/2022_JSLHR-22-00251

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97–113.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*(2), 227–252. https://doi.org/10.1016/J.COGNITION.2009.09.007

Pagnoni, I., Gobbi, E., Premi, E., Borroni, B., Binetti, G., Cotelli, M., & Manenti, R. (2021). Language training for oral and written naming impairment in primary progressive aphasia: A review. *Translational Neurodegeneration, 10*(1). https://doi.org/10.1186/s40035-021-00248-z. BioMed Central Ltd.

Patra, A., Traut, H. J., Stabile, M., & Middleton, E. L. (2022). Effortful retrieval practice effects in lexical access: A role for semantic competition. *Language, Cognition and Neuroscience, 37*(8), 948–963. https://doi.org/10.1080/23273798.2022.2027991

Payne, J. S. (2020). Developing L2 productive language skills online and the strategic use of instructional tools. *Foreign Language Annals, 53*(2), 243–249. https://doi.org/10.1111/flan.12457

Piai, V., & Knight, R. T. (2018). Lexical selection with competing distractors: Evidence from left temporal lobe lesions. *Psychonomic Bulletin & Review, 25*(2), 710–717. https://doi.org/10.3758/s13423-017-1301-0

Piai, V., Riès, S. K., & Swick, D. (2016). Lesions to lateral prefrontal cortex impair lexical interference control in word production. *Frontiers in Human Neuroscience, 9*(JAN2016). https://doi.org/10.3389/fnhum.2015.00721

Pino, D., Mädebach, A., Jescheniak, J. D., Regenbrecht, F., & Obrig, H. (2022). BONEs not CATs attract DOGs: Semantic context effects for picture naming in the lesioned language network. *Neuroimage, 246*, Article 118767. https://doi.org/10.1016/j.neuroimage.2021.118767

Pisoni, A., Papagno, C., & Cattaneo, Z. (2012). Neural correlates of the semantic interference effect: New evidence from transcranial direct current stimulation. *Neuroscience, 223*, 56–67. https://doi.org/10.1016/j.neuroscience.2012.07.046

Pompili, A., Abad, A., Trancoso, I., Fonseca, J., Martins, I. P., Leal, G., & Farrajota, L. (2011). An on-line system for remote treatment of aphasia. In *Proceedings of the second Workshop on speech and language processing for assistive Technologies* (pp. 1–10). https://aclanthology.org/W11-2301.pdf.

Python, G., Demierre, C., Bourqui, M., Bourbon, A., Chardenon, E., Trouville, R., Laganaro, M., & Fougeron, C. (2023). Comparison of in-person and online recordings in the clinical teleassessment of speech production: A pilot study. *Brain Sciences, 13*(2), 342. https://doi.org/10.3390/brainsci13020342

Python, G., Glize, B., & Laganaro, M. (2018). The involvement of left inferior frontal and middle temporal cortices in word production unveiled by greater facilitation effects following brain damage. *Neuropsychologia, 121*, 122–134. https://doi.org/10.1016/j.neuropsychologia.2018.10.026

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reimers, S., & Stewart, N. (2015). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript Web experiments. *Behavior Research Methods, 47*(2), 309–327. https://doi.org/10.3758/s13428-014-0471-1

Riès, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T., & Dronkers, N. F. (2015). Specifying the role of the left prefrontal cortex in word selection. *Brain and Language, 149*, 135–147. https://doi.org/10.1016/J.BANDL.2015.07.007

Ries, S. K., Piai, V., Perry, D., Griffin, S., Jordan, K., Henry, R., Knight, R. T., & Berger, M. S. (2019). Roles of ventral versus dorsal pathways in language production: An awake language mapping study. *Brain and Language, 191*, 17–27. https://doi.org/10.1016/j.bandl.2019.01.001

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language, 134*, 104472. https://doi.org/10.1016/j.jml.2023.104472

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*(1–3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Roelofs, A. (2022). A neurocognitive computational account of word production, comprehension, and repetition in primary progressive aphasia. *Brain and Language, 227*, Article 105094. https://doi.org/10.1016/j.bandl.2022.105094

Rorden, C., & Brett, M. (2000). Lesion analysis. *Behavioural Neurology, 12*(0953–4180), 191–200.

Rose, S. B., & Abdel Rahman, R. (2016). Cumulative semantic interference for associative relations in language production. *Cognition, 152*, 20–31. https://doi.org/10.1016/j.cognition.2016.03.013

Rose, S. B., & Abdel Rahman, R. (2017). Semantic similarity promotes interference in the continuous naming paradigm: Behavioural and electrophysiological evidence. *Language, Cognition and Neuroscience, 32*(1), 55–68. https://doi.org/10.1080/23273798.2016.1212081

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(4), 753–763. https://doi.org/10.1037/xlm0000592

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods, 49*(5), 1864–1881. https://doi.org/10.3758/s13428-016-0830-1

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, hosting and recruiting: A brief introduction to running behavioral experiments online. *Brain Sciences, 10*(4), 251. https://doi.org/10.3390/brainsci10040251

Sauter, M., Stefani, M., & Mack, W. (2022). Equal quality for online and lab data: A direct comparison from two dual-task paradigms. *Open Psychology, 4*(1), 47–59. https://doi.org/10.1515/psych-2022-0003

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*. https://doi.org/10.1016/j.jml.2019.104038

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language, 75*, 27–44. https://doi.org/10.1016/j.jml.2014.04.006

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language, 54*(2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences, 106*(1), 322–327. https://doi.org/10.1073/pnas.0805874106

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102. https://doi.org/10.1016/0749-596X(90)90011-N

Schwartz, M. F. (2014). Theoretical analysis of word production deficits in adult aphasia. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), Article 20120390. https://doi.org/10.1098/rstb.2012.0390

Schwarz, J., Li, K. K., Sim, J. H., Zhang, Y., Buchanan-Worster, E., Post, B., Gibson, J. L., & McDougall, K. (2022). Semantic cues modulate children's and adults' processing of audio-visual face mask speech. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.879156

Scott, R. M., & Wilshire, C. E. (2010). Lexical competition for production in a case of non fluent aphasia: Converging evidence from four different tasks. *Cognitive Neuropsychology, 27*(6), 505–538. https://doi.org/10.1080/02643294.2011.598853

Spitzer, L., Binkofski, F., Willmes, K., & Bruehl, S. (2021). The novel cognitive flexibility in aphasia therapy (CFAT): A combined treatment of aphasia and executive functions to improve communicative success. *International Journal of Speech-Language Pathology, 23*(2), 168–179. https://doi.org/10.1080/17549507.2020.1757152

Stadie, N., Cholewa, J., & de Bleser, R. (2013). *LEMO 2.0 Lexikon modellorientiert. Diagnostik für Aphasie, Dyslexie und Dysgraphie*. natverlag.

Stark, K., van Scherpenberg, C., Obrig, H., & Abdel Rahman, R. (2022). Web-based language production experiments: Semantic interference assessment is robust for spoken and typed response modalities. *Behavior Research Methods, 55*(1), 236–262. https://doi.org/10.3758/s13428-021-01768-2

Starreveld, P. A., & La Heij, W. (2017). Picture-word interference is a stroop effect: A theoretical analysis and new empirical findings. *Psychonomic Bulletin & Review, 24*(3), 721–733. https://doi.org/10.3758/s13423-016-1167-6

Stoehr, A., Souganidis, C., Thomas, T. B., Jacobsen, J., & Martin, C. D. (2023). Voice onset time and vowel formant measures in online testing and laboratory-based testing with(out) surgical face masks. *The Journal of the Acoustical Society of America, 154*(1), 152–166. https://doi.org/10.1121/10.0020064

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*(4), 584–589. https://doi.org/10.3758/BF03193889

van Scherpenberg, C., Just, A., & Hauber, R. (2020). *Check voice onset times from Chronset with Praat script*. https://osf.io/fmwqb/.

van Scherpenberg, C., Rahman, R. A., Regenbrecht, F., & Obrig, H. (2021). Semantic interference through multiple distractors in picture naming in people with aphasia. *Journal of Cognitive Neuroscience, 33*(8), 1612–1633. https://doi.org/10.1162/jocn_a_01731

Vogt, A., Hauber, R., Kuhlen, A. K., & Abdel Rahman, R. (2022). Internet-based language production research with overt articulation: Proof of concept, challenges, and practical advice. *Behavior Research Methods*, 54(4), 1954–1975. https://doi.org/10.3758/S13428-021-01686-3

Walker, G. M., Hickok, G., & Fridriksson, J. (2018). A cognitive psychometric model for assessment of picture naming abilities in aphasia. *Psychological Assessment, 30*(6), 809–826. https://doi.org/10.1037/pas0000529

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom mapping using support vector regression. *Human Brain Mapping, 35*(12), 5861–5876. https://doi.org/10.1002/hbm.22590