Received: 28 December 2021 | Accepted: 14 November 2022

DOI: 10.1111/1460-6984.12827

International Journal of Language & Communication Disorders

REVIEW

Naming as a window to word retrieval changes in healthy and pathological ageing: Methodological considerations

Silvia Martínez-Ferreiro

Gerontology and Geriatrics Research Group, Department of Physiotherapy, Medicine and Biomedical Sciences, University of A Coruña, A Coruña, Spain

Correspondence
Silvia Martínez-Ferreiro, Gerontology and Geriatrics Research Group, Department of Physiotherapy, Medicine and Biomedical Sciences, Faculty of Health Sciences, University of A Coruña, Campus de Oza, E-15006, A Coruña, Spain.
Email: s.martinez.ferreiro@udc.es

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Number: Ramón y Cajal (RYC2020-028927-1); Funding for open access charge: Universidade da Coruña/CISUG

Abstract

Background: Word retrieval skills change across the lifespan. Permanent alterations in the form of decreased accuracy or increased response time can be a consequence of both normal ageing processes or the presence of acquired and neurodegenerative disorders (e.g., aphasia and dementia). Despite the extensive literature exploring the neuroanatomical underpinnings of word retrieval, psycholinguistic, biolinguistic and theoretical explanations, and the vast amount of evidence from primary and secondary language disorders, the best approach to consistently capture these changes is yet to be discovered.

Aims: The goal of this paper is to determine which method(s) stand(s) as the most suitable candidate(s) to provide an accurate picture of word retrieval in the oral production of different groups of adult speakers, including cases of healthy ageing, preclinical Alzheimer's disease (AD), mild cognitive impairment (MCI), aphasia and dementia.

Methods & Procedures: Using an integrative review of recent peer-reviewed journal articles, we provide an overview of the different behavioural methods traditionally used to measure oral naming skills in research-oriented and clinical protocols and discuss their main advantages and limitations.

Main Contribution: Most existing studies are based on the results of people with diagnosed language disorders. Despite the growing interest, the reliability of the majority of the tasks to detect subtle changes associated with healthy ageing, MCI and preclinical AD are yet to be demonstrated, and the delicate balance between informativeness and efficiency (especially in terms of administration time and variable control) in experimental protocols is yet to be achieved. In this article we propose the pursuit of an integrative overarching methodology to characterize all naming deficits (from anecdotal to permanent) and all adult populations (from healthy to pathological ageing).

Conclusions & Implications: A combination of spontaneous speech data and results from structured tasks stands as the best approach to capture changes in

word retrieval skills of adult speakers with and without observable deficits. This review can guide future reflections on the necessary prerequisites of purpose-oriented, sensitive and reliable protocols for the detection of incipient word retrieval problems, thus contributing to the early diagnosis and the design of personalized multicomponent treatments.

KEYWORDS
word retrieval, object and action naming, healthy ageing, mild cognitive impairment, aphasia, dementia

WHAT THIS PAPER ADDS

What is already known on this subject

Word retrieval skills change during adulthood as a consequence of the neurological degradation associated with ageing. These changes are more dramatic in the event of acquired and neurodegenerative disorders. Numerous studies based on people with observable language disorders have addressed the multiplicity of factors involved in word retrieval and provided evidence of potential loci of impairment from a neuroanatomical, cognitive and/or (psycho-)linguistic perspective.

What this paper adds to existing knowledge

This study focuses on methodological strategies to assess naming skills and provides a reflection on generally accepted good practices and unresolved challenges to inform task selection, emphasizing the necessity for a combination of methods to best capture the actual problems and needs of people confronting word retrieval difficulties in their daily lives. Task selection, variable control and administration time stand as key concepts to adjust to the requirements of research and clinical contexts.

What are the potential or actual clinical implications of this work?

The results of this review can orient future research towards the creation of sensitive, reliable and (ecologically) valid materials for the (early) detection of word retrieval deficits and for the customization of treatment protocols to alleviate or palliate their effects.

INTRODUCTION

People's ability to express themselves and understand others in one (or more) language(s) varies over time. This is due to a conglomerate of both internal factors such as the cognitive changes associated with normal ageing and with acquired or degenerative neurological conditions such as aphasia or Alzheimer's disease (AD), but also due to the influence of external factors such as exposure to more than one language or situations leading to language attrition. Language changes can be traceable across all the languages mastered by an individual (L1, L2, L*n*), all modalities (production and comprehension), and at all linguistic levels (e.g., syntax, semantics, pragmatics, discourse). The current paper focuses on internal factors, more specifically, in a widely debated topic: word retrieval in healthy and pathological ageing.

It only takes a quick search in PubMed[1] including the keyword 'word retrieval' to get an idea of the relevance and the complexity of the topic and the overwhelming amount of multidisciplinary research conducted so far (4162 results between the date of the first record in 1963 and 2022; 3459 after 2000, including 59 systematic reviews covering educational, clinical and research-oriented contributions). The effect is more stunning if we use 'naming' to guide our search (149,421 results between 1804 and 2022;

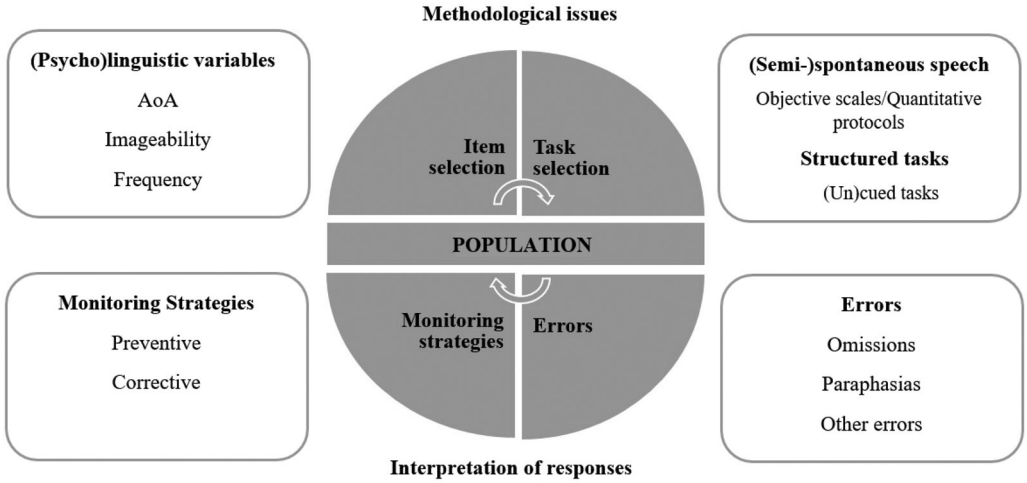

FIGURE 1 Methodological considerations: Sources of variability

124,548 after 2000; 755 systematic reviews). Such an extensive body of research is the result of the collaborative effort of neurologists, neuropsychologists, cognitive psycholinguists, clinical linguists, and speech and language therapists. In this special issue we concentrate on methodological issues, providing an integrative overview of the methods most extensively and effectively used across disciplines to assess and characterize changes in oral naming skills in the healthy and injured brain (Figure 1).

After examining common practices and some of the major findings in different groups of adult speakers, we discuss the advantages and disadvantages offered by different production methods. The novelty resides in the inclusion of a reflection on generally accepted good practices and challenges to guide the design of purpose-oriented protocols aiming at the provision of an accurate picture of word retrieval. With such a broad topic, the author cannot do justice to all relevant sources. Consequently, to minimize the profusion of in-text references, with very few exceptions, this integrative review focuses on a subset of articles published from 2000 onwards and addresses the reader to the seminal references cited there. Given that protocols need to serve different purposes (clinical versus research-oriented protocols) and that the main focus of the existing experimental studies is on young adults and individuals with observable language disorders, we depart from the following hypotheses: To date, (1) method selection has been constrained by the purpose of the study and by the characteristics of the target population, and (2) in isolation, none of the available behavioural methods is subtle enough to accurately detect subtle changes in the naming skills of ageing individuals or set clear boundaries across groups of speakers (e.g., normal ageing, preclinical AD and mild cognitive impairment—MCI).

OBJECT AND ACTION NAMING AS A WINDOW TO WORD RETRIEVAL SKILLS

Naming is essential to communication. But hitherto, to speak of naming studies is to speak of two inherently different word classes, nouns and verbs, two complex categories consisting of different sets of conceptual, semantic, syntactic, morphological and phonological features (for detailed characterizations, see Conroy et al., 2006; and Mätzig et al., 2009, *passim*). Differences between nouns and verbs and their impact on the performance of different groups of speakers have been the object of a lively debate during the past 50 years. In this section, we summarize the most widely discussed views regarding three main themes, namely (1) the spatial and temporal resolution of object and action naming, (2) the conflicting experimental results existing in the literature, and (3) the methodological challenges to be faced to properly capture preserved versus impaired oral production skills across word classes.

Regarding the spatial and temporal course of noun and action naming, cognitive and language production models show that word retrieval is a complex process engaging a multi-component network (e.g., Levelt et al., 1999). It is generally accepted that word retrieval starts with the activation of the word (conceptual stage); however, the participation of the sensory-motor system at this early stage is still an open debate. According to classical theories, conceptual knowledge is stored in the anterior end of the temporal lobe, independently of the sensory-motor systems (e.g., Lambon Ralph et al., 2017). This view is contested by strong and weak grounded cognition approaches. According to these proposals, conceptual knowledge entails the bilateral activation of the sensorimotor cortex, the visual, and the auditory areas associated

with the features of the target word (Kiefer & Pulvermüller, 2012; Popp et al., 2019). Popp et al. (2019) revisit studies supporting amodal and modality-specific approaches to the representation of conceptual knowledge and show feature-specific conceptual processing differences between objects and actions. Whereas objects are strongly linked to visual-related semantic content (Setti et al., 2009), action-related semantic content is predominant in verbs (Moseley & Pulvermüller, 2014).

Independently of the theoretical stance, most authors agree that after conceptual preparation, the semantic (lemma retrieval), the syntactic and the morphological representation of the item (grammatical encoding) start. According to Indefrey (2011) and Indefrey and Levelt (2004), lemma retrieval takes place at around 200 ms in the left MTG and is immediately followed by lemma selection. Next, during the word-form processing or articulation stage, the phonological representation of the word is activated before moving on to articulatory planning and articulation (phonological and phonetic encoding). Phonological code retrieval is thought to take place in the left posterior STG at around 275 ms, giving way to syllabification at 355 ms. Phonological encoding occurs in the left portions of the IFG at 455 ms. Other areas such as the left precentral gyrus, the left thalamus, and the cerebellum are also involved in word retrieval, as they contribute to articulation. Depending on the methodology used to elicit a given word, a preliminary stage may enter the equation. If word retrieval occurs as a response to an experimental condition including images or words presented auditorily or in their written form, the recognition and identification of the stimulus will trigger the activation of the visual representation, the phonological or the graphemic input lexicon (respectively) before access to the semantic representation is granted. It is important to note that, as the disruption of visual or auditory sensory-specific mechanisms corresponds to pre-linguistic processing, its influence in naming performance has to be kept apart from deficits occasioned by semantic and phonological disruptions (Carreiras et al., 2014; Hillis, 2008; Indefrey, 2011; Indefrey & Levelt, 2004; Race & Hillis, 2015; Raymer, 2015; for a recent review, see Lukic et al., 2021).

In such a distributed multi-component network, numerous types of brain injury or degeneration may differentially disrupt noun and verb retrieval, potentially leading to the multiplicity of linguistic profiles profusely characterized in the literature. Whereas both object and action naming crucially rely on shared left-lateralized areas of the brain, verbs may depend on the recruitment and coordination of additional subcortical structures and right hemisphere areas (Alyahya et al., 2018; Rofes & Mahon, 2021; Rofes & Miceli, 2014).[2] Ever since the 1970s, behavioural studies have shown dissociations between nouns and verbs in different clinical groups (e.g., Kavé & Goral, 2017; Lukic et al., 2021; Mason & Nickels, 2021, *passim*, for a comprehensive summary of evidence from healthy ageing, aphasia and neurodegenerative diseases). According to Miceli et al. (1984), lesion location may be one of the factors underlying the dissociation between nouns and verbs observed in aphasia, as patients with more posterior lesions (e.g., anomic patients) usually experience more difficulties with nouns, whereas those patients with lesions affecting relatively more anterior areas generally suffer from deficits with a higher impact on verbs (for supporting evidence, see Laiacona & Caramazza, 2004, *passim*). Lukic et al. (2021) report a similar pattern in patients with primary progressive aphasia (PPA). Patients with semantic PPA experienced more problems with nouns, whereas for patients with non-fluent/agrammatic PPA verbs were more problematic. However, taken together, empirical results seem to be only loosely connected to lesion location and syndromic classifications (see Conroy et al., 2006; Crepaldi et al., 2006; Kambanaros, 2010; Mätzig et al., 2009, *passim*, for aphasia). This is mainly due to the existence of contradictory results across aphasia types, across individuals in the same group, and even variation within individuals (Luzzatti et al., 2001; Miceli et al., 1984; Howard & Gatehouse, 2006).

Independently of potential differences regarding the neural bases of verb and noun retrieval, verbs are generally accepted to be more demanding than nouns (Mätzig et al., 2009). Many studies based on patient observation agree that object retrieval is generally better preserved than action retrieval. However, the exact underlying causes justifying such a pattern of performance are not clear (Bird et al., 2000; Kambanaros, 2010). Deficits have been attributed to selective damage at different stages of word retrieval or at various linguistic levels (see Crepaldi et al., 2006; and Alyahya et al., 2018, for a summary). Noun–verb dissociations have been associated with lemma or lexeme deficits associated with the phonological or orthographic output lexicons (see Caramazza & Hillis, 1991, and much subsequent work) and to semantic degradation (e.g., Bird et al., 2000), which are claimed to affect nouns and verbs selectively (Rapp & Caramazza, 2002). Various proposals focusing on different language domains have also been formulated in the past decades to attempt an explanation of the (apparently) conflicting naming results (see Miceli et al., 1984, for noun–verb double dissociation; and Luzzatti et al., 2001). These include lexical accounts, based on the assumption of separate mental lexicons for nouns and verbs susceptible of being damaged independently, and semantic, syntactic and morphological accounts, which attribute the noun–verb dissociation to the additional complexity of verbs (as opposed to nouns) at each of these levels (see Miceli et al., 1984; Bird et al., 2000; Kim & Thompson,

2000; and Tsapkini et al., 2002, respectively). None of these theories can fully account for the diversity of data, making it difficult to verify or falsify any hypotheses.

Two of the factors that are currently recognized as potential sources of variability in the experimental results are task and item selection (Vigliocco et al., 2011). Regarding the task, context has been found to play a role in aphasia. Evidence comes from intervention studies that show better outcomes when words are uttered in phrasal or sentence context (noun syntax therapy; Herbert et al., 2012) than in isolation (lexical therapy; e.g., Wisenburn & Mahoney, 2009; for a review, see Herbert et al., 2014, *passim*). Not only do different tasks impose different cognitive demands or have differential facilitating effects, but differences in the (psycho)linguistic properties of the experimental items have a direct impact on performance (Alyahya et al., 2018; Conroy et al., 2006; Luzzatti et al., 2002). Several (psycho)linguistic variables are known to affect word retrieval in both healthy and clinical populations. Linguistic variables include, among others, grammatical, semantic and lexical features such as word class, semantic similarity, target word length, grammatical and phonological regularity, and complexity of the syllabic structure. Among psycholinguistic factors, age of acquisition (AoA), educational level, animacy, imageability, instrumentality, and lemma and lexeme frequency have also been claimed to play a role in naming accuracy and response time.

It is generally accepted that AoA has a strong effect on word retrieval in young and elderly adults, with words acquired earlier being easier/faster to process than those acquired later (Bates et al., 2001; Cuetos et al., 2002; Hodgson & Ellis, 1998; Johnston & Barry, 2006). Spontaneous speech analyses have shown that AoA and educational level also affect word retrieval in patients with AD (Gayraud et al., 2013; Forbes-McKay & Venneri, 2005) and that AoA can reliably be used to detect early manifestations of the disease (Forbes-McKay et al., 2005). Similar claims have been made for imageability and frequency (e.g., see Bird et al., 2000; and Howard & Gatehouse, 2006, for evidence from people with aphasia) or word length (see Hodgson & Ellis, 1998, for healthy ageing; and Crepaldi et al., 2006, for aphasia). Notably, frequency seems to be a strong predictor of degradation in preclinical AD and MCI (Clarke et al., 2021).

Although lack of control over psycholinguistic variables may be held (at least partially) responsible for the observed dissociation between nouns and verbs, this line of research is not free from conflicting results, and the weight attributable to each of the (psycho)linguistic variables differs across languages and studies (Alyahya et al., 2018; Bastiaanse et al., 2016; Crepaldi et al., 2006; Cuetos et al., 2002; Hodgson & Ellis, 1998). One of the cases that have generated a wide debate is frequency. While claimed to be acquired later than nouns and have lower imageability, certain common verbs reach higher frequency rates than most nouns. However, verb retrieval difficulties are common to many groups of adult speakers. An interaction between AoA, imageability and frequency effects could be operative in the case of verbs. Although authors such as Cuetos et al. (2002) found an effect of AoA, visual complexity, object familiarity and word frequency on naming accuracy in Spanish-speaking people with aphasia, Bastiaanse et al. (2016) claim that when other factors such as imageability are controlled for, the weight of lemma and lexeme frequency is reduced to a minor effect in object (but not in action) naming. Also, no word frequency effects were found in Hodgson and Ellis's (1998) study of healthy ageing.

These inconsistencies call for a systematic evaluation of (psycho)linguistic variables, as control stands as one of the keys to unravelling the complex (sometimes contradictory) patterns of performance observed across healthy and clinical groups. However, finding a balance between ecological validity and experimental control is not easy to achieve. When exerting excessive control, there is a risk of obtaining biased observations based on a reduced subset of 'laboratory' items distant from the relevant real-life vocabulary of the general population (e.g., words such as *trellis* are included in the BNT; Kaplan et al., 1983). The use of low-frequency words in standardized tests increases the chance of giving rise to (lack of) familiarity and vocabulary knowledge effects, sometimes associated with educational level (Hamberger & Seidel, 2003).[3] In addition, not all available tests use equally strict criteria for the inclusion of items. Sometimes frequency data may be absent or come from written materials. Differences in mean word frequency across tools render the interpretation of results difficult (Yochim et al., 2013).

Even if the need to factorize AoA and imageability in any experimental design seems undeniable, the frequent use of confrontation naming tasks including visual stimuli creates an additional bias towards the use of highly imaginable words in most tests, an effect which is stronger in the case of nouns than in the case of verbs (Bird et al., 2000; Crepaldi et al., 2006). This is especially relevant if we consider that the inclusion of abstract words may not only help provide a more fine-grained description of word retrieval deficits, but it leads to greater learning and generalization to untrained words when included in treatment protocols (Kiran et al., 2009).

Moreover, so far, we have referred to nouns and verbs (word class) and object and action naming (semantics). However, it is worth noting that, to provide a complete picture of word retrieval skills, nouns other than common concrete countable nouns referring to objects and verbs that do not depict actions should be further

investigated. But, how can we effectively elicit low imageability nouns? Some tests include repetition tasks to compensate for this limitation. However, the mechanisms underlying oral naming and repetition are significantly different, thus, again, imposing difficulties for the interpretation of the results. The same holds, for instance, for psychological verbs (e.g., *please*, *bore*) or words belonging to other word classes. One of the main challenges regarding this specific group of verbs is the difficulty of depicting psychological states without transforming them into actions (e.g., *love*—using a picture of people *hugging* or *kissing*). Videos have been used to overcome these difficulties (Conroy et al., 2006). However, among the criticisms, the additional complexity of videos as compared to static pictures and the supplementary administration times, have to be taken into account. These challenges and the currently available resources are the matter of discussion in the next sections.

METHODOLOGICAL APPROACHES TO WORD RETRIEVAL

Daily communication is full of examples of word-finding difficulties, from anecdotal occurrences of the tip-of-the-tongue phenomenon to more severe deficits jeopardizing the effectiveness of a given communicational exchange. To understand this phenomenon in full, two utterly different methodological routes have been explored: (semi-)spontaneous speech sample analyses and structured tasks.

(Semi-)spontaneous speech analysis

(Semi-)spontaneous speech has been traditionally used by linguists, psychologists, and speech and language therapists for the observation and characterization of a large number of linguistic phenomena, as it provides close insights into the real daily communicative routines of individuals (for instance, see Menn & Obler's, 1990, seminal work on agrammatic aphasia). (Semi-)spontaneous speech analysis grants an advantage over closed naming tasks for assessing 'real' everyday word-finding difficulties, as familiarity effects are generally overruled. Further, word-finding is embedded in sentence generation and does not require the additional complexity added by the recognition/identification of a given stimulus (although it entails sentence processing and phrasal construction, and, consequently, further syntactic complexity that can make it difficult to identify word-retrieval deficits) (see Schmitter-Edgecombe et al., 2000, for evidence from elderly adults; and Cuetos et al., 2002, for aphasia).

In addition to ecological validity, analysing (semi-)spontaneous speech samples offers certain supplementary benefits over more constrained methods in clinical settings. Among others, it can be used to assess language deficits in the absence of standardized tools (Abuom & Bastiaanse, 2012), to detect subtle deviant patterns unnoticeable for standard neuropsychological evaluations (e.g., see Jaecks et al., 2012, for residual aphasias; and Antonsson et al., 2021, for MCI), and to obtain baseline measures to assess recovery after intervention (Brookshire & Nicholas, 1994). Hence, given its susceptibility to reveal subtle changes, spontaneous speech analysis stands a priori as a good tool to detect changes in the word retrieval skills of both healthy and pathological groups of ageing individuals, especially in cases of preclinical AD and MCI.

However, in addition to being highly time-consuming and requiring a deep linguistic knowledge (with transcription and analysis highly susceptible to interrater effects), the use of (semi-)spontaneous speech samples suffers from an important number of shortcomings. Naturally, spontaneous speech varies significantly across individuals, which makes the evaluation of the decrease or degeneration of word retrieval skills dependent on the access to normative or control datasets, usually unavailable for most languages. Also critically, the lack of consensus in data collection procedures and methodological decisions hinders replicability and limits the comparison across studies (Prins & Bastiaanse, 2004).

Various discourse tasks with different cognitive demands have been traditionally used by psychologists, linguists, and speech and language therapists (e.g., Bastiaanse et al., 1996; Bastiaanse & Jonkers, 1998; Goodglass et al., 1993; Nicholas & Brookshire, 1993; Prins & Bastiaanse, 2004; Saffran et al., 1989; Vermeulen et al., 1989; Wagenaar & Prins, 1975). These tasks have been conceived to serve different aims and focus on different discourse types. Whereas traditionally psychologists tended to use descriptive and narrative tasks based on pictures, that crucially involve semantic knowledge and retrieval, or well-known tales and procedures, that, in addition to the former, involve episodic memory, linguists and speech and language therapists tended to rely on semi-standardized interviews and monologist and dialogist conversational analyses (see Prins & Bastiaanse, 2004, for aphasia; and Clarke et al., 2021, for a recent discussion of these tasks in AD). These methods, nowadays used in combination by all sorts of professionals, vary as for the amount of control over the speaker's intention and the ecology of the data (> ecological = < control). Unconstrained interaction data and spontaneous speech analyses are generally dispreferred, as the intention of the speaker may not always be clear, especially in cases of moderate to severe language impairment. Descriptive and narrative discourse

(pictures and tales with or without picture support) allow for better control of the speaker's intentions, especially relevant in the case of moderate and severe language disorders. However, the vocabulary may not be related to everyday language and, with very few exceptions, pictures need to be constantly updated and adapted to match the diverse cultural characteristics of their target audiences to avoid (lack of) familiarity effects, an easier task in the absence of printed materials (Berube et al., 2019; Edwards & Bastiaanse, 2007). Expository discourse (semi-standardized interviews and conversational datasets), is more naturalistic but responses may be too short to allow for quantification, and comprehension of the participants' intentions is not always guaranteed. Lack of control over psycholinguistic, phonological, and lexico-semantic variables is common to all tasks, although to a varying degree (Prins & Bastiaanse, 2004).

Regarding methodological decisions, the adequate sample size (controlled by speech tempo or by the number of words), the required number and the type of variables to be included in the analysis (rating scales, quantifiable linguistic variables), and how to quantify the results (e.g., presence/absence of features versus presence/absence of errors) are just some illustrative examples of the variability attested across studies and disciplines (for a review of aphasia studies, see Prins & Bastiaanse, 2004; for a discussion of sample size in dementia studies, see Ossewaarde et al., 2020). The purpose of the study (research versus clinical goals), the target population (healthy individuals versus individuals with primary and secondary language disorders), and even the typological characteristics of the language spoken by a given participant (e.g., synthetic versus analytic) are taken as decisive factors in the selection of one method over the other (Edwards & Knott, 1994). Finally, scoring is also an important source of variation and an important aspect to consider. Although score sheets should be concise and easy to complete (time and content-wise) to control for potential interrater effects and promote usability in clinical contexts, this is not always the case and their suitability for adaptation to other languages to favour replication is a pending issue (e.g., see Fyndanis et al., 2017, for the necessary changes introduced in the cross-linguistic adaptations of the Comprehensive Aphasia Test (CAT); and Swinburn et al., 2005). (Semi-)spontaneous speech protocols may benefit from the edition of exhaustive explanatory manuals including concise instructions for administration and scoring, as well as scripts (e.g., the AphasiaBank[4] protocol).

Despite the shortcomings, two different methods with a varying degree of demonstrated reliability, objective scales and quantitative linguistic protocols, have been used over the years to analyse language production (Prins & Bastiaanse, 2004; e.g., see Saffran et al.'s, 1989, Quantitative Production Analysis (QPA); and Boxum et al.'s, 2013, *Analyse voor Spontane Taal Bij Afasie* (ASTA)). Traditional measures at the word level frequently include proportions of lexical words, type/token ratios, and noun/verb ratios but also, as in the case of the QPA, noun/pronoun or noun/determiner ratios that can ease the quantification of errors and contribute to the diagnose of specific language disorders. Although originally conceived for the analysis of non-fluent aphasias, these methods for the analysis of (semi-)spontaneous speech have been successfully exploited in the detection of naming impairments in patients with aphasia of varying severity, dementia and fluency problems (e.g., Gordon, 2006). However, they generally fail to provide indications for therapy (Prins & Bastiaanse, 2004). To date, the most widely used comprehensive neuropsychological evaluations include picture description tasks. Some of the best-known examples are the cookie theft picture (BDAE; Goodglass et al., 2001) or the picnic scene (WAB-R; Kertesz, 2007). However, although a certain control of the linguistic and psycholinguistic properties of the words involved is always possible (notice for instance the recurrent use of *boy*, *girl*, *mother/woman*, *cookie*, *water*, *stool*, *washing*, *falling* in the cookie theft picture), the results of these tasks tend to be used to measure expressive skills focusing on aspects such as articulatory agility, utterance length, grammaticality and complexity, prosodic aspects, and content (BDAE scoresheet; Goodglass et al., 2001).

Although not specifically envisioned for the detection of fine-grained patterns of disruption at the word level, under certain conditions, especially in cases of moderate-to-severe stroke-induced aphasia, picture description tasks may arguably be reduced to a mere confrontation naming exercise. However, recent results show that confrontation naming abilities strongly relate to informativeness in picture description tasks (Boucher et al., 2020). Contrary to the former, picture description tasks are claimed to provide further complex information about more ecological communication settings (Boucher et al., 2020).

In their comprehensive review of aphasia studies, Kavé and Goral (2017) show that, in connected speech, patients' word-finding difficulties generally lead to a more reduced output with limited lexical variety and containing more dysfluencies and substitution errors than those detected in their healthy counterparts. A similar pattern has been reported after the analysis of discourse samples in patients with AD (Gayraud et al., 2013; Forbes-McKay & Venneri, 2005; Forbes-McKay et al., 2005). More controversial is the adequacy of spontaneous speech samples to characterize word retrieval skills in healthy ageing, preclinical AD and MCI. Kavé and Goral (2017) fail to identify a clear pattern of performance in the speech output of elderly adults

based on spontaneous speech results alone. However, Schmitter-Edgecombe et al. (2000) report significant differences in the number of semantic errors between elderly and young adults, thus claiming that discourse tasks may be better than structured tasks for assessing word-finding difficulties in healthy adults. Although further evidence is still needed to confirm the adequacy of (semi-)spontaneous speech tasks as a meaningful screening tool in cases of preclinical AD and MCI (Antonsson et al., 2021), task selection seems to be crucial to accurately detect early symptoms. Despite the paucity of data, narrative tasks stand as the most adequate task to differentiate these individuals from healthy elderly adults (see Clarke et al., 2021, *passim*, for dementia). Narrative tasks have also been successfully used to discriminate word production skills in people with semantic dementia, progressive non-fluent aphasia and healthy controls (Fraser et al., 2014).

Structured tasks

According to the American Psychological Association, the label 'naming task' is used to refer to all those tasks in which participants are required to name objects or actions, independently of the specific type of task or the aim of the study (e.g., assessment versus treatment). Although naming tasks can also be used without standardized tools (e.g., it is common for SLPs to ask people with aphasia to name items in the room during screening), standard neuropsychological evaluations typically rely on structured tasks to detect the presence of anomia, impaired word-retrieval abilities, and semantic memory difficulties (e.g., for the CAT, see Swinburn et al., 2005), thus constituting an important diagnostic tool for acquired and degenerative neurological disorders. The most widely used tasks for assessment purposes are *confrontation naming tasks*, which involve the attribution of a specific label to a corresponding stimulus, generally a picture. Stimuli can also be auditory or text-based (e.g., naming after a description). Presentation mode has a direct effect on lexical processing as it triggers the activation of different brain areas before accessing the semantic representation (Hillis, 2008; Raymer, 2015). The effect of presentational mode in the healthy and injured brain has not yet been studied in detail. However, preliminary differences in the results of auditory and visual tasks in individuals with word-finding complaints suggest that auditorily presented stimuli may help identify deficits that can go unnoticed in visual confrontation naming tasks (Hamberger & Seidel, 2003).

In addition to uncued tasks, *cued naming tasks* are also frequently used in research, assessment, and therapy. Different types of (mis)cues can be provided to facilitate or obstruct participants' performance, altering accuracy rates, error patterns, and response times. Tasks including *semantic cues* target the meaning of the word as a way to test the semantic system, including semantic memory. Different cues address the activation of semantic features. Gestures and descriptions, for instance, aim at activating a specific meaning (*responsive naming*, e.g., shape, function). Other association tasks consist of providing semantic features such as the category the word belongs to (e.g., naming words belonging to a specific semantic field to measure category fluency—*divergent naming*) to elicit a target word or a word list, usually under certain time constraints (*rapid naming*). Naming the category a given word belongs to (*convergent naming*) or providing lists of synonyms or associated words are other available methods. Finally, sentence completion tasks may equally fall under this category.

Tasks including *phonemic cues* provide information about the first sound(s) or syllable of a target word or the motor configuration required for the production of these segments (visual phonemic cues). Similar to semantic tasks, tasks providing *orthographic cues* include spelling information (e.g., the first letter or syllable or even spelling the word aloud). Whereas phonemic cues are expected to ease the access to the phonological representation of a given word, orthographic cues target the activation of the graphemic representation, which may allow for the self-generation of phonemic cues (e.g., Howard & Gatehouse, 2006). Among the tasks falling in this group, we find the production of word lists with certain characteristics (e.g., words starting with a specific letter or syllable orally or visually presented, as in *phonemic fluency tasks*). Other sound-based cues used in speech and language therapy are tapping out the number of syllables or providing rhyming words. The effectivity of some cues over others varies from subject to subject (Nickels, 2002). According to Meteyard and Bose (2018) phonological cues have a more significant impact on accuracy than semantic cues. Contrary to the general assumption that phonological cues favour phonological processes, increased accuracy is attributed to the effect of phonological information on picture recognition and categorization, which may justify inconsistencies across subjects.

Among the tasks including miscues, one commonly used paradigm is the *semantic and/or phonologic picture–word interference paradigm*, which consists of the presentation of semantically and/or phonologically related distractors during picture naming. These tasks aim at raising competition at either the lemma or the phonological level. The selection or combination of tasks and the inclusion of (mis)cues vary with the aim of the study. Given the impossibility of predicting treatment needs and outcomes, confrontation naming tasks are generally completed by

cue-based tasks during speech and language therapy to serve facilitatory, repairing, or re-educational purposes (Nickels, 2002). The effectiveness of different types of cues have been found to vary across syndromes. Whereas phonological cues have been reported to be beneficial in any type of anomia, semantic cues are only relevant in cases of lexical–phonological or mixed anomia (Howard et al., 1985; Python et al., 2021).

Structured tasks show several advantages over the analysis of spontaneous speech samples, such as reduced administration and data processing time, further availability of normative data, and better control for (psycho)linguistic variables. Nonetheless, these tasks also present a significant number of limitations. First, the number and type of items included in each task are important factors to control for. In protocols subject to time constraints, finding a balance between representativity and efficiency may lead to misjudgements. If the number of items per category is insufficient, one error may have a strong statistical weight in the interpretation of the results, resulting in the underestimation or overestimation of word-finding problems (Kambanaros, 2010). Another factor has to do with feature control, as controlling for a diversity of (psycho)linguistic variables may lead to the selection of denatured items and, consequently, to familiarity and educational level effects (Hamberger & Seidel, 2003). Granting access to the list of tested items, the name agreement results (in picture-based tasks), and the values of the (psycho)linguistic variables of the experimental items is also essential, especially for cross-linguistic adaptations and comparisons as the values, and consequently, the adequacy or relevance of a given stimulus can vary significantly across languages (Hodgson & Ellis, 1998).

Quantitative results from confrontation naming and fluency tasks are nowadays the standard used to diagnose word retrieval deficits in different clinical groups (e.g., BNT, Kaplan et al., 1983; OANB, Druks & Masterson, 2000; and CAT, Swinburn et al., 2005; envisioned for people with aphasia and dementia). Although with some exceptions (Kambanaros, 2010, *passim*), according to Kavé and Goral (2017) and Mason and Nickels (2021), the results of structured tasks in people with aphasia tend to show the same pattern of performance as those obtained from the analysis of (semi-)spontaneous speech samples; that is, the more severe the word retrieval difficulties, the more reduced output and lexical variety and the more dysfluencies and substitution errors. However, Schmitter-Edgecombe et al. (2000) report that in healthy ageing, naming accuracy can be higher in picture naming than in discourse tasks. Also, the use of picture description tasks alone may have an impact on the understanding of how early AD pathology reflects in speech (Clarke et al., 2021), and semantic fluency tasks have resulted to be critical to distinguish between MCI and dementia (Belleville et al., 2017). This evidence suggests the need for a combination of methods to provide an accurate overarching account of word retrieval problems in healthy ageing and preclinical and clinical groups.

ON THE INFORMATIVENESS OF ERRORS AND ERROR TYPES AND THE ROLE OF MONITORING STRATEGIES

In addition to the number of correct and incorrect responses, the nature of the errors can provide information relevant (1) for deepening our understanding of how word retrieval takes place and how it is implemented in the brain, (2) for further delving into how the different processes involved in lexical retrieval can be selectively disrupted, and (3) for achieving accurate differential diagnoses. According to classical sources in the field of aphasiology, deficits may be due to impaired semantic specifications (e.g., Hillis et al., 1990), failure in the phonological output (e.g., Kay & Ellis, 1987), or lost representations (Howard, 1995). The classification of errors and the identification of sensitive or potential breakdown points are essential for the early detection of syndromes, their diagnosis and the consequent intervention planning. Grima and Franklin (2017) showed that the presence of atypical errors is key to distinguishing between mild anomic impairments in healthy ageing and those in people with aphasia, an unnoticeable difference in the quantitative test scores. Similarly, Silagi et al. (2015) showed that variability in the pattern of errors could help distinguish between controls and mild and moderate AD patients. However, offline evidence is not always easy to interpret and some of the observable errors can be attributed to more than one potential disruption site (see Howard & Gatehouse, 2006, for aphasia), thus limiting treatment planning capacity.

Moreover, error interpretation crucially depends on the classification of responses, which is not always straightforward. Challenges are more evident when analysing spontaneous speech samples, as the lack of experimental constraints allows for the exploitation of a variety of mechanisms and compensatory strategies that coexist in the same individual (Baciu et al., 2021). Both healthy older adults and people with neurological conditions (including psychogenic disorders) implement preventive (pre-articulatory) self-monitoring strategies (e.g., avoidance, Heeschen, 1985; ellipsis and morphosyntactic simplifications, Hofstede, 1992; Kolk, 1995; and substitutions, Nespoulous, 1996). Resorting to these strategies can go unnoticed. Lack of knowledge of the speakers' intention makes it impossible to evaluate certain responses as

correct, that is, the produced word corresponds to the original intention of the speaker, or incorrect, as in certain cases of ellipsis or substitutions by synonyms and hypernyms (that may result into a reduction of lexical diversity). As avoidance is more limited in structured tasks, where stimuli are generally selected to evoke a univocal response, the use of strategies is mostly restricted to repair (post-articulatory) strategies such as self-corrections (Hartsuiker & Kolk, 2001), which are easier to detect. The application of self-monitoring strategies may result in higher accuracy scores in (semi-)spontaneous speech tasks as compared to structured tasks. The availability of self-monitoring strategies may not only vary across tasks but also across populations, as monitoring is assumed to be comprehension and/or production-based and consequently depends on preserved production and/or comprehension skills as well as in cognitive functioning (Kolk, 2006; Nickels & Howard, 1995; Levelt et al., 1999; Roelofs, 2020). Different degrees of impairment in the ability to implement preventive and corrective strategies have been reported in fluent and non-fluent aphasias, as well as in other conditions such as apraxia of speech (Oomen et al., 2001). Differential impairment has also been key to distinguishing among degenerative disorders such as frontotemporal dementia (FTD) and PPA (Banks & Weintraub, 2008).

Regarding the type of errors, some of the most common word retrieval errors are omissions and *don't know* responses, the production of paraphasias (phonological, semantic, and unrelated), perseverations, filler words, neologisms, and circumlocutions. Each of these errors can provide information to circumscribe the specific stage at which word retrieval broke down and to identify task and item-related factors affecting production. In people with aphasia, omission errors have been found to appear as a consequence of lexical–semantic deficits (rather than phonological deficits), due to the need to select from a competitor set (Chen et al., 2019). According to Friedmann et al. (2013) review of lexical retrieval in acquired and developmental language disorders, problems with the conceptual system may lead to the production of unrelated paraphasias. Although in structured tasks, they may also be an indicator of problems with a specific experimental item (vocabulary knowledge). Semantic paraphasias may be attributed to lemma access or lemma retrieval deficits, that is, to problems with the semantic lexicon, which is also susceptible to imageability, category specificity, and typicality effects. Impaired lexical and phonological representations can result in the use of both semantic and phonological paraphasias, although phonological paraphasias (together with morphological substitutions) may also indicate problems with the phonological output buffer. The phonological lexicon is sensitive to word frequency effects (see also Bastiaanse et al., 2016). Length and syllable frequency effects are also thought to interfere with production at this stage (see Howard & Gatehouse, 2006, for an interesting reverse-length effect in a patient with aphasia). Other responses such as circumlocutions (from sporadic—tip of the tongue phenomenon—to pathological) are compatible with multiple disruption sites.

ABOUT CHOICES

Uttering the right word is a complex neuropsycholinguistic process. In this paper, we have briefly sketched the neural bases and the temporal resolution of word retrieval, how different (psycho)linguistic factors interfere, the processes it involves, and how it is disrupted. The number and type of errors have permitted us to measure the degree of preservation of word retrieval skills allowing to distinguish between young and elderly adults (Schmitter-Edgecombe et al., 2000), but also between normal ageing and the effects of primary and secondary language disorders such as aphasia (Friedmann et al., 2013) and dementia (Clarke et al., 2021). However, as we have pointed out along these lines, appearances can be deceiving, and, despite our vast knowledge, characterizing naming impairments in terms of specific disruption sites is still an unattainable goal, as different errors are compatible with different disrupted processes (Howard & Gatehouse, 2006), and the integrity of the entire language network seems to be necessary for word production (Døli et al., 2021).

Additionally, a significant part of the studies come from cases of aphasia after stroke, more specifically Broca's and anomic aphasia in speakers of a reduced subset of languages. Aphasias due to other etiologies, mixed and residual aphasias, and aphasias among speakers of non-Indo-European languages are still minimally represented (Beveridge & Bak, 2011). Despite a growing interest in the study of dementia and healthy ageing in the past decades, a systematic evaluation of (cross-)linguistic symptoms also remains a need. Not to mention that, in comparison, the dissection of impairments in MCI and preclinical AD is still in its infancy.[5] Although the use of a common methodology is desirable in terms of comparability, only further data from different groups of adult speakers will help evaluate the suitability of these methods to discriminate between healthy and pathological ageing and across different forms of impairment, especially when it comes to describing typical patterns or mild word retrieval difficulties (e.g., Rohrer et al., 2008, for progressive aphasias). Differences in skill level across adult populations may call for the examination of words of diverse complexity (e.g., low-frequency items, abstract words) in order to be able to detect subtle changes.

Even for the most studied disorders, the number of participants per sample remains a common complaint. The bulk of data still comes from single case or case-series reports (e.g., Bird et al., 2000; Miceli et al., 1984; Rapp & Caramazza, 2002; Tsapkini et al., 2002). Studies including large cohorts of participants are to date still scarce due to the difficulty of recruiting homogeneous groups (e.g., Alyahya et al., 2018; Luzzatti et al., 2001). Small cohorts, combined with the methodological differences across studies, render difficult the formulation of strong theoretical and methodological generalizations regarding word retrieval and give rise to competing theories at all levels. Enhancing replicability and data sharing could help palliate this effect.

The recent tendency towards data sharing, the creation of open databases including exhaustive characterizations of linguistic and psycholinguistic properties of a wide range of words in certain languages (McEnery & Hardie, 2012), and the compilation of normative data from large groups of speakers without language complaints across different age, gender and educational attainment groups can help estimate what a given person should know to better adjust our metrics and interpretation of the results of both types of tasks. The proper contextualization of the currently observed behaviour can shed light on the sometimes controversial use of the word 'significant' in test results not only from a research perspective, which tends to adhere to a more prescriptive approach, but also in clinical contexts, and very importantly, from the perspective of the people living with naming difficulties.

The inconsistencies (and sometimes contradictions) across results show that, on their own, none of the available behavioural methods is sensitive enough to capture the complexity of factors intervening in word retrieval across different groups of participants, thus favouring the multiplication of methodological approaches to address specific phenomena. As naming and discourse tasks impose different cognitive demands, they can provide complementary information. So far, choices have been heavily determined by the purpose of the study and by the characteristics of the target population: from the linguistic characterization of changes to the identification of potential disruption sites across populations to serve assessment, treatment, and outcome measurement purposes. However, all this work has not been fruitless. As recommended by Mason and Nickels (2021), taken together, and once the advantages and disadvantages of each method have been evaluated, the experimental results confirm the need to promote the rigorous development of protocols including a combination of both (semi-)spontaneous speech and structured tasks, consistent with the procedures currently applied in the clinical practice, where most batteries rely on a conglomerate of tasks (BDAE, Goodglass, Kaplan & Barresi, 2001; CAT, Swinburn, Porter & Howard, 2005). Among others, the need for (semi-)spontaneous speech tasks can be justified in terms of ecology, sensitivity to preclinical distortions, and avoidance of (un)familiarity and word knowledge effects (e.g., Jaecks et al., 2012). Given that picture naming tasks pose certain problems (Mason & Nickels, 2021), narrative tasks should be further investigated as an alternative method (Clarke et al., 2021). The inclusion of structured tasks is also justified as they allow for greater control over (psycho)linguistic variables and for their theoretical and clinical relevance (specificity, administration time, normative data, consequences for therapy). Both confrontation naming and fluency tasks are highly informative (Belleville et al., 2017). The possibility of supplementing visual confrontation naming paradigms with auditory tasks (Hamberger & Seidel, 2003) deserves further investigation. In what follows, we provide an overview of certain additional factors and control mechanisms that seem to be key to achieving a better understanding of oral naming data and furnish some recommendations regarding further methodological choices and future directions.

The role of linguistic and psycholinguistic variables

Even if testing nouns alone is still a widely generalized practice, nowadays it is undeniable that *word class* is one of the factors that play a crucial role in word retrieval. Eliciting both nouns and verbs is critical to provide a more fine-grained picture of overall word retrieval skills. However, it is important to note that the image is still incomplete as objects and actions represent a minimal subset of the words required for efficient communicative exchanges. Information regarding other categories of nouns and verbs may help decide among competing theories such as the amodal versus modality-specific representation of conceptual knowledge (Popp et al., 2019). Including other word classes such as adjectives and adverbs is also important to augment our knowledge base of lexical diversity and help refine the diagnostic capabilities of the assessment protocols (Kim et al., 2019).

Regarding *psycholinguistic variables*, AoA and imageability are the most generally accepted factors to control for in experimental designs, as these not only consistently affect word retrieval skills in different groups of healthy speakers (young versus healthy ageing subjects), but they have also proven their diagnostic potential in the detection of early manifestations of certain secondary language disorders such as those observed in AD (Bird et al., 2000; Crepaldi et al., 2006; Forbes-McKay et al., 2005; Gayraud et al., 2013; Hodgson & Ellis, 1998). Other variables have

been shown to interfere with word retrieval (Friedmann et al., 2013). But the more variables that are factorized, the more items are needed to properly capture its effect, which compromises administration times and calls for a selective prioritization based on study objectives. Deepening our knowledge about the individual contribution of each variable or set of variables would undoubtedly contribute to eliminate redundancies and homogenize the list of indices that must be controlled in each experimental design, thus promoting comparability across studies.

The role of monitoring strategies

The use of monitoring strategies may also contribute to the characterization of certain deficits, especially in the case of degenerative conditions. In this respect, the use of structured tasks interferes with the respondents' freedom to omit or substitute potentially conflictive terms, thus again providing more control over the outcomes than that offered by spontaneous speech samples.

Cued versus uncued elicitation

The use of cues and miscues in structured tasks reduces the ecological validity of the outcomes, since these can bring out patterns different from those observed in the daily communicative exchanges of a given individual (Mason & Nickels, 2021). However, cues have been successfully used to maximize treatment outcomes. In line with spreading activation theories and compound cue theories (McKoon & Ratcliff, 1992), facilitatory cues may enhance performance, although variability across and within individuals does not allow prediction of results, calling for treatment customization (Nickels, 2002). Miscues, on the other hand, have been proved useful for the identification of potential disruption sites, as different miscues may target specific phases of word retrieval. However, the outcomes are generally compatible with more than one disruption site and make it difficult to unambiguously interpret the results (Howard & Gatehouse, 2006). Only further dissection of the tasks and the exploration of new datasets will help produce clearer answers.

CONCLUSIONS

In the same way that there is no perfect speaker, there is also no perfect task to characterize his/her production. Isolated tasks fail to accurately detect subtle changes in the naming skills of ageing individuals, set clear boundaries across certain groups of speakers, and help circumscribe the locus of impairment enabling an earlier and customized therapeutic intervention while simultaneously enlarging our theoretical knowledge about word retrieval. Hence, a combination of methods, that is, the use of (semi-)spontaneous speech data supplemented by structured tasks, is key for the creation of reliable, sensitive, and (ecologically) valid standard testing procedures. The balance between informativeness and efficiency (especially in terms of administration time) in experimental protocols is yet to be achieved. Task selection, variable control (especially AoA and imageability), and administration time stand as key concepts to adjust to the requirements of research and clinical contexts.

ORCID

Silvia Martínez-Ferreiro https://orcid.org/0000-0003-2393-1214

NOTES

[1] Retrieved on 27 July 2022.

[2] The potential contribution of the right hemisphere in word retrieval cannot be ignored, as it may play an active part as a compensation mechanism for compromised left-hemisphere structures in clinical populations (Riès et al., 2016).

[3] As pointed out by one of the reviewers, low frequency items are not necessarily distant from the 'relevant real-life vocabulary' of a particular individual (e.g., the frequency of specialized vocabulary changes among professionals). In line with Renvall et al. (2013), we distinguish here between 'personally chosen' and 'generally frequent' items and discuss the available (standardized) tools to characterize word retrieval in healthy and pathological ageing in general. Limiting the assessment of word retrieval to more widely used words enhances comparability, although it may limit the sensitivity of testing (e.g., subtle decline in word retrieval in typical ageing; Schmitter-Edgecombe et al., 2000), as well as the potential topics that a particular person with language difficulties can discuss.

[4] See https://aphasia.talkbank.org/.

[5] The linguistic characterization of tumour and vascular malformation patients would fall into this category.

CONFLICT OF INTEREST

The author reports no conflict of interest.

DATA AVAILABILITY STATEMENT

Not applicable.

REFERENCES

Abuom, T.O. & Bastiaanse, R. (2012) Characteristics of Swahili–English bilingual agrammatic spontaneous speech and the consequences for understanding agrammatic aphasia. *Journal of Neurolinguistics*, 25, 276–293. DOI: 10.1016/j.jneuroling.2012.02.003

Alyahya, R., Halai, A.D., Conroy, P., Ralph, L., & M, A. (2018) Noun and verb processing in aphasia: behavioural profiles and neural

correlates. *NeuroImage: Clinical*, 18, 215–230. DOI: 10.1016/j.nicl.2018.01.023

Antonsson, M., Lundholm Fors, K., Eckerström, M. & Kokkinakis, D. (2021) Using a discourse task to explore semantic ability in persons with cognitive impairment. *Frontiers in Aging Neuroscience*, 12, 607449. https://doi.org/10.3389/fnagi.2020.607449

Baciu, M., Banjac, S., Roger, E., Haldin, C., Perrone-Bertolotti, M., Lœvenbruck, H. & Démonet, J.F. (2021) Strategies and cognitive reserve to preserve lexical production in aging. *GeroScience*, 43(4), 1725–1765. DOI: 10.1007/s11357-021-00367-5

Banks, S. & Weintraub, S. (2008) Self-awareness and self-monitoring of cognitive and behavioral deficits in behavioral variant frontotemporal dementia, primary progressive aphasia and probable Alzheimer's disease. *Brain and Cognition*, 67(1), 58–68. DOI: 10.1016/j.bandc.2007.11.004

Bastiaanse, R., Edwards, S. & Kiss, K. (1996) Fluent aphasia in three languages: aspects of spontaneous speech. *Aphasiology*, 10, 561–575. DOI: 10.1080/02687039608248437

Bastiaanse, R. & Jonkers, R. (1998) Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology*, 12, 951–969. DOI: 10.1080/02687039808249463

Bastiaanse, R., Wieling, M. & Wolthuis, N. (2016) The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology*, 30, 1221–1239. DOI: 10.1080/02687038.2015.1100709

Bates, E., Burani, C., D'Amico, S. & Barca, L. (2001) Word naming and picture naming in Italian. *Memory and Cognition*, 29, 986–999. DOI: 10.3758/BF03195761

Beveridge, M.E.L. & Bak, T.H. (2011) The languages of aphasia research: bias and diversity. *Aphasiology*, 25(12), 1451–1468. DOI: 10.1080/02687038.2011.624165

Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. & Croteau, J. & Consortium for the Early Identification of Alzheimer's Disease—Quebec (2017) Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 328–353. DOI: 10.1007/s11065-017-9361-5

Berube, S., Nonnemacher, J., Demsky, C., Glenn, S., Saxena, S., Wright, A., et al. (2019) Stealing cookies in the twenty-first century: measures of spoken narrative in healthy versus speakers with Aphasia. *American Journal of Speech–Language Pathology*, 28(1S), 321–329. DOI:10.1044/2018_AJSLP-17-0131

Bird, H., Howard, D. & Franklin, S. (2000) Why is a verb like an inanimate object? Grammatical category and semantic category deficits. *Brain and Language*, 72, 246–309. DOI: 10.1006/brln.2000.2292

Boucher, J., Marcotte, K., Brisebois, A., Courson, M., Houzé, B., Desautels, A., Léonard, C., Rochon, E. & Brambati, S.M. (2020) Word-finding in confrontation naming and picture descriptions produced by individuals with early post-stroke aphasia. *The Clinical Neuropsychologist*, 36(6), 1422–1437. DOI: 10.1080/13854046.2020.1817563

Boxum, E., Van der Scheer, F. & Zwaga, M. (2013) Analyse voor Spontane Taal bij Afasie(ASTA). Standaard in samenwerking met de Vereniging voor Klinische Linguïstiek. 4th version. Retrieved from: https://klinischelinguistiek.nl/

Brookshire, R.H. & Nicholas, L.E. (1994) Speech sample size and test–retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research*, 37, 399–407. DOI: 10.1044/jshr.3702.399

Caramazza, A. & Hillis, A. (1991) Lexical organization of nouns and verbs in the brain. *Nature*, 349, 788–790. DOI: 10.1038/349788a0

Carreiras, M., Armstrong, B.C., Perea, M. & Frost, R. (2014) The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, 18(2), 90–98. DOI: 10.1016/j.tics.2013.11.005

Chen, Q., Middleton, E. & Mirman, D. (2019) Words fail: lesion-symptom mapping of errors of omission in post-stroke aphasia. *Journal of Neuropsychology*, 13(2), 183–197. DOI: 10.1111/jnp.12148

Conroy, P., Sage, K. & Ralph, L. & M, A. (2006) Towards theory-driven therapies for aphasic verb impairments: a review of current theory and practice. *Aphasiology*, 20(12), 1159–1185. DOI: 10.1080/02687030600792009

Clarke, N., Barrick, T.R. & Garrard, P. (2021) A comparison of connected speech tasks for detecting early Alzheimer's disease and mild cognitive impairment using natural language processing and machine learning. *Frontiers in Computer Science*, 3, 634360. DOI: 10.3389/fcomp.2021.634360

Crepaldi, D., Aggujaro, S., Arduino, L.S., Zonca, G., Ghirardi, G., Inzaghi, M.G., Colombo, M., Chierchia, G. & Luzzatti, C. (2006) Noun–verb dissociation in aphasia: the role of imageability and functional locus of the lesion. *Neuropsychologia*, 44, 73–89. DOI: 10.1016/j.neuropsychologia.2005.04.006

Cuetos, F., Aguado, G., Izura, C. & Ellis, A.W. (2002) Aphasic naming in Spanish: predictors and errors. *Brain and Language*, 82, 344–365. DOI: 10.1016/S0093-934X(02)00038-X

Døli, H., Andersen Helland, W., Helland, T. & Specht, K. (2021) Associations between lesion size, lesion location and aphasia in acute stroke. *Aphasiology*, 35(6), 745–763. DOI: 10.1080/02687038.2020.1727838

Druks, J. & Masterson, J. (2000) *An object & action naming battery*. Hove, U.K: Psychology Press.

Edwards, S. & Bastiaanse, R. (2007) Assessment of aphasia in a multi-lingual world. In: Ball, M.J. & Damico, J.S. (Eds.) *Clinical aphasiology: future directions*. New York, NY: Psychology Press, pp. 245–258.

Edwards, S. & Knott, R. (1994) Assessing spontaneous language abilities of aphasic speakers. *Language Testing*, 11(1), 49–64. DOI: 10.1177/026553229401100105

Forbes-McKay, K.E. & Venneri, A. (2005) Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences*, 26, 243–254. DOI: 10.1007/s10072-005-0467-9

Forbes-McKay, K.E., Ellis, A.W., Shanks, M.F. & Venneri, A. (2005) The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age-related cognitive decline. *Neuropsychologia*, 43(11), 1625–1632. DOI: 10.1016/j.neuropsychologia.2005.01.008

Fraser, K.C., Meltzer, J.A., Graham, N.L., Leonard, C., Hirst, G., Black, S.E. & Rochon, E. (2014) Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex*, 55, 43–60. DOI: 10.1016/j.cortex.2012.12.006

Friedmann, N., Biran, M. & Dotan, D. (2013) Lexical retrieval and breakdown in aphasia and developmental language impairment. In: Boeckx, C. & Grohmann, K.K. (Eds.) *The Cambridge Handbook of Biolinguistics*. Cambridge, UK: Cambridge University Press, pp. 350–374.

Fyndanis, V., Lind, M., Varlokosta, S., Gram Simonsen, H., Kambanaros, M., Ceder, K., Rofes, A., Soroli, E., Bjekic, J., Gavarró, A., Grohmann, K., Kuvac, J., Martinez Ferreiro, S., Munarriz, A., Vuksanovic, J., Zakarias, L. & Howard, D. (2017) Cross-linguistic adaptations of The Comprehensive Aphasia Test: challenges and solutions. *Clinical Linguistics & Phonetics*, 31(7–9), 697–710. DOI: 10.1080/02699206.2017.1310299

Gayraud, F., Thibert, C. & Barkat-Defradas, M. (2013) *Age of acquisition affects word retrieval in spontaneous speech produced by patients with Alzheimer's disease*. Toulouse, France: Perspectives neuropsycholinguistiques sur l'Aphasie. ⟨halshs-00779151⟩

Grima, R. & Franklin, S. (2017) Usefulness of investigating error profiles in diagnosis of naming impairments. *International Journal of Language & Communication disorders*, 52(2), 214–226. DOI: 10.1111/1460-6984.12266

Goodglass, H., Christiansen, J.A. & Gallagher, R. (1993) Comparison of morphology and syntax in free narrative and structured tests: fluent vs. nonfluent aphasics. *Cortex*, 29, 377–407. DOI: 10.1016/S0010-9452(13)80250-X

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

Gordon, J.K. (2006) A quantitative production analysis of picture description. *Aphasiology*, 20, 188–204. DOI:10.1080/02687030500472777

Hamberger, M.J. & Seidel, W.T. (2003) Auditory and visual naming tests: normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society: JINS*, 9(3), 479–489. DOI: 10.1017/s135561770393013x

Hartsuiker, R.J. & Kolk, H.H. (2001) Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology*, 42(2), 113–157. DOI: 10.1006/cogp.2000.0744

Herbert, R., Webster, D. & Dyson, L. (2012) Effects of syntactic cueing therapy on picture naming and connected speech in acquired aphasia. *Neuropsychological Rehabilitation*, 22(4), 609–633. DOI: 10.1080/09602011.2012.679030

Herbert, R., Gregory, E. & Best, W. (2014) Syntactic versus lexical therapy for anomia in acquired aphasia: differential effects on narrative and conversation. *International Journal of Language & Communication Disorders*, 49(2), 162–173. DOI: 10.1111/1460-6984.12054

Heeschen, C. (1985) Agrammatism versus paragrammatism: a fictitious opposition. In: Kean, M.L. (Ed.) *Agrammatism*. New York: Academic Press, pp. 207–248.

Hillis, A.E., Rapp, B., Romani, C. & Caramazza, A. (1990) Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology*, 7(3), 191–243. DOI: 10.1080/02643299008253442

Hillis, A.E. (2008) Cognitive neuropsychological approaches to rehabilitation of language disorders: introduction. In: Chapey, R. (Ed). *Language interventions strategies in aphasia and related neurogenic communication disorders*, 5th edition. Philadelphia: Lippincott Williams & Wilkins, pp. 595–604.

Hodgson, C. & Ellis, A.W. (1998) Last in, first to go: age of acquisition and naming in the elderly. *Brain and Language*, 64, 146–163. DOI: 10.1006/brln.1998.1960

Hofstede, B.T.M. (1992) *Agrammatic speech in Broca's Aphasia: strategic choice for the elliptical register*. Ph.D. dissertation, Catholic University of Nijmegen [available as NICI Technical Report 92-07].

Howard, D. (1995) Lexical Anomia: or the case of the missing lexical entries. *The Quarterly Journal of Experimental Psychology Section A*, 48(4), 999–1023. DOI: 10.1080/14640749508401426

Howard, D. & Gatehouse, C. (2006) Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology*, 20(9), 921–950, DOI: 10.1080/02687030600782679

Howard, D., Patterson, K., Franklin, S., Orchard-lisle, V. & Morton, J. (1985) The facilitation of picture naming in aphasia. *Cognitive Neuropsychology*, 2, 49–80. DOI: 10.1080/02643298508252861

Indefrey, P. (2011) The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, 2, 255. DOI: 10.3389/fpsyg.2011.00255

Indefrey, P. & Levelt, W.J.M. (2004) The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144. DOI: 10.1016/j.cognition.2002.06.001

Jaecks, P., Hielscher-Fastabend, M. & Stenneken, P. (2012) Diagnosing residual aphasia using spontaneous speech analysis. *Aphasiology*, 26, 953–970. DOI: 10.1080/02687038.2012.663075

Johnston, R. & Barry, C. (2006) Age of acquisition and lexical processing. *Visual Cognition*, 13(7/8), 789–845. DOI:10.1080/13506280544000066

Kambanaros, M. (2010) Action and object naming versus verb and noun retrieval in connected speech: comparisons in late bilingual Greek–English anomic speakers. *Aphasiology*, 24, 210–230. 10.1080/02687030902958332

Kaplan, E., Goodglass, H. & Weintraub, S. (1983) *Boston Naming Test*. Philadelphia: Lea & Febiger.

Kavé, G. & Goral, M. (2017) Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*, 24(5), 508–527. DOI: 10.1080/13825585.2016.1226249

Kay, J. & Ellis, A. (1987) A cognitive neuropsychological case study of anomia. Implications for psychological models of word retrieval. *Brain: A Journal of Neurology*, 110(Pt 3), 613–629. DOI: 10.1093/brain/110.3.613

Kertesz, A. (2007) *Western Aphasia Battery–Revised*. San Antonio, TX: The Psychological Corporation.

Kiefer, M. & Pulvermüller, F. (2012) Conceptual representations in mind and brain: theoretical developments, current evidence, and future directions. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 48, 805–825. DOI: 10.1016/j.cortex.2011.04.006

Kim, M. & Thompson, C.K. (2000) Patterns of comprehension and production of nouns and verbs in agrammatism: implications for lexical organization. *Brain and Language*, 74(1), 1–25. DOI: 10.1006/brln.2000.2315

Kim, H., Kintz, S., Zelnosky, K. & Wright, H.H. (2019) Measuring word retrieval in narrative discourse: core lexicon in aphasia. *International Journal of Language & Communication Disorders*, 54(1), 62–78. DOI: 10.1111/1460-6984.12432

Kiran, S., Sandberg, C. & Abbott, K. (2009) Treatment for lexical retrieval using abstract and concrete words in persons with aphasia: effect of complexity. *Aphasiology*, 23(7), 835–853. DOI: 10.1080/02687030802588866

Kolk, H. (1995) A time-based approach to agrammatic production. *Brain and Language*, 50(3), 282–303. DOI: 10.1006/brln.1995.1049

Kolk, H. (2006) How language adapts to the brain: an analysis of agrammatic aphasia. *The Syntax of Nonsententials: Multidisciplinary Perspectives*, 229–258. DOI: 10.1075/la.93.11kol

Laiacona, M. & Caramazza, A. (2004) The noun/verb dissociation in language production: varieties of causes. *Cognitive Neuropsychology*, 21(2), 103–23. DOI: 10.1080/02643290342000311

Ralph, L., A, M., Jefferies, E., Patterson, K. & Rogers, T.T. (2017) The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18, 42–55. DOI: 10.1038/nrn.2016.150

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

Lukic, S., Borghesani, V., Weis, E., Welch, A., Bogley, R., Neuhaus, J., Deleon, J., Miller, Z.A., Kramer, J.H., Miller, B.L., Dronkers, N.F. & Gorno-Tempini, M.L. (2021) Dissociating nouns and verbs in temporal and Perisylvian networks: evidence from neurodegenerative diseases. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 142, 47–61. DOI: 10.1016/j.cortex.2021.05.006

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A. & Pinna, G. (2001) On the nature of the selective impairment of verb and noun retrieval. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 37, 724–726.

Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A. & Pinna, G.D. (2002) Verb–noun double dissociation in aphasic lexical impairments: the role of word frequency and imageability. *Brain and Language*, 81(1–3), 432–444. DOI: 10.1006/brln.2001.2536

McEnery, T. & Hardie, A. (2012) *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511981395

Mason, C. & Nickels, L. (2021) Are single-word picture naming assessments a valid measure of word retrieval in connected speech? *International Journal of Speech–Language Pathology*, 6, 1–13. DOI: 10.1080/17549507.2021.1966098

Mätzig, S., Druks, J., Masterson, J. & Vigliocco, G. (2009) Noun and verb differences in picture naming: past studies and new evidence. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 45, 738–58. DOI: 10.1016/j.cortex.2008.10.003

McKoon, G. & Ratcliff, R. (1992) Spreading activation versus compound cue accounts of priming: mediated priming revisited. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 18(6), 1155–1172. DOI: 10.1037//0278-7393.18.6.1155

Menn, L. & Obler, L.K. (1990) *Agrammatic aphasia: a cross-language narrative sourcebook*. Amsterdam: John Benjamins.

Meteyard, L. & Bose, A. (2018) What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. *Journal of Speech, Language, and Hearing Research: JSLHR*, 61(3), 658–674. https://doi.org/10.1044/2017_JSLHR-L-17-0214

Miceli, G., Silveri, M.C., Villa, G. & Caramazza, A. (1984) On the basis for the agrammatic's difficulty in producing main verbs. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 20, 207–240. DOI: 10.1016/s0010-9452(84)80038-6

Moseley, R.L. & Pulvermüller, F. (2014) Nouns, verbs, objects, actions, and abstractions: local fMRI activity indexes semantics, not lexical categories. *Brain and Language*, 132(100), 28–42. DOI: 10.1016/j.bandl.2014.03.001

Nespoulous, J.-L. (1996) Les stratégies palliatives dans l'aphasie [Palliative strategies in aphasia]. *Rééducation Orthophonique*, 34(188), 423–433.

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

Nicholas, L.E. & Brookshire, R.H. (1993) A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, 36, 338–350. DOI: 10.1044/jshr.3602.338

Nickels, L. & Howard, D. (1995) Phonological errors in aphasic naming: comprehension, monitoring, and lexicality. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 31, 209–237. DOI: 10.1016/S0010-9452(13)80360-7

Oomen, C.C., Postma, A. & Kolk, H.H. (2001) Prearticulatory and postarticulatory self-monitoring in Broca's aphasia. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 37(5), 627–41. DOI: 10.1016/s0010-9452(08)70610-5

Ossewaarde, R., Jonkers, R., Jalvingh, F. & Bastiaanse, R. (2020) Quantifying the Uncertainty of Parameters Measured in Spontaneous Speech of Speakers With Dementia. *Journal of Speech, Language, and Hearing Research*, 63(7), 2255–2270. DOI: 10.1044/2020_JSLHR-19-00222

Popp, M., Trumpp, N.M., Sim, E.J. & Kiefer, M. (2019) Brain Activation During Conceptual Processing of Action and Sound Verbs. *Advances in Cognitive Psychology*, 15(4), 236–255. DOI: 10.5709/acp-0272-4

Prins, R. & Bastiaanse, R. (2004) Analyzing the spontaneous speech of aphasic speakers. *Aphasiology*, 18(12), 1075–1091. DOI: 10.1080/02687030444000534

Python, G., Pellet Cheneval, P., Bonnans, C. & Laganaro, M. (2021) Facilitating word retrieval in aphasia: which type of cues for which aphasic speakers? *Frontiers in Human Neuroscience*, 15, 747391. DOI: 10.3389/fnhum.2021.747391

Race, D.C. & Hillis, A.E. (2015) The neural mechanisms underlying naming. In: Hillis, A.E. (Ed.) *The handbook of adult language disorders*. New York: Psychology Press, pp. 151–160.

Rapp, B. & Caramazza, A. (2002) Selective difficulties with spoken nouns and written verbs: a single case study. *Journal of Neurolinguistics*, 15, 373–402.

Raymer, A.M. (2015) Clinical diagnosis and treatment of naming disorders. In: Hillis, A.E. (Ed.) *The handbook of adult language disorders*. New York: Psychology Press, pp. 161–183.

Renvall, K., Nickels, L. & Davidson, B. (2013) Functionally relevant items in the treatment of aphasia (part I): challenges for current practice. *Aphasiology*, 27(6), 636–650. DOI: 10.1080/02687038.2013.786804

Riès, S.K., Dronkers, N.F. & Knight, R.T. (2016) Choosing words: left hemisphere, right hemisphere, or both? Perspective on the lateralization of word retrieval. *Annals of the New York Academy of Sciences*, 1369(1), 111–131. DOI: 10.1111/nyas.12993

Roelofs, A. (2020) Self-monitoring in speaking: in defense of a comprehension-based account. *Journal of Cognition*, 3(1), 18. DOI: 10.5334/joc.61

Rofes, A. & Mahon, B.Z. (2021) Naming: nouns and Verbs. In: Mandonnet, E. & Herbet, G. (Eds.) *Intraoperative mapping of cognitive networks*. Cham: Springer. DOI: 10.1007/978-3-030-75071-8_11

Rofes, A. & Miceli, G. (2014) Language mapping with verbs and sentences in awake surgery: a review. *Neuropsychology Review*, 24(2), 185–199. DOI:10.1007/s11065-014-92510

Rohrer, J.D., Knight, W.D., Warren, J.E., Fox, N.C., Rossor, M.N. & Warren, J.D. (2008) Word-finding difficulty: a clinical analysis of

the progressive aphasias. *Brain*, 131(1), 8–38. DOI: 10.1093/brain/awm251

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

Schmitter-Edgecombe, M., Vesneski, M. & Jones, D.W. (2000) Aging and word-finding: a comparison of spontaneous and constrained naming tests. *Archives of Clinical Neuropsychology*, 15(6), 479–93.

Setti, A., Caramelli, N. & Borghi, A.M. (2009) Conceptual information about size of objects in nouns. *European Journal of Cognitive Psychology*, 21, (7), 1022–1044. DOI: 10.1080/09541440802469499

Silagi, M.L., Bertolucci, P.H. & Ortiz, K.Z. (2015) Naming ability in patients with mild to moderate Alzheimer's disease: what changes occur with the evolution of the disease? *Clinics (Sao Paulo, Brazil)*, 70(6), 423–428. DOI: 10.6061/clinics/2015(06)07

Swinburn, K., Porter, G. & Howard, D. (2005) *The comprehensive aphasia test*. Hove, UK: Psychology Press.

Tsapkini, K., Jarema, G. & Kehayia, E. (2002) A morphological processing deficit in verbs but not in nouns: a case study in a highly inflected language. *Journal of Neurolinguistics*, 15, 265–288. DOI: 10.1016/S0911-6044(01)00039-2

Vermeulen, J., Bastiaanse, R. & Van Wageningen, B. (1989) Spontaneous speech in aphasia: a correlational study. *Brain and Language*, 36, 252–274. DOI: 10.1016/0093-934x(89)90064-3

Vigliocco, G., Vinson, D., Druks, J. & Cappa, S. (2011) Nouns and verbs in the brain: a review of behavioral, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews*, 35, 407–426. DOI: 10.1016/j.neubiorev.2010.04.007

Wagenaar, E., Snow, C. & Prins, R. (1975) Spontaneous speech of aphasic patients: a psycholinguistic analysis. *Brain and Language*, 2, 281–303. DOI: 10.1016/s0093-934x(75)80071-x

Wisenburn, B. & Mahoney, K. (2009) A meta-analysis of word-finding treatments for aphasia. *Aphasiology*, 23, 1338–1352. DOI:10.1080/02687030902732745

Yochim, B.P., Rashid, K., Raymond, N.C. & Beaudreau, S.A. (2013) How frequently are words used on naming tests used in spoken conversation? *The Clinical Neuropsychologist*, 27(6), 973–987. DOI: 10.1080/13854046.2013.797501

How to cite this article: Martínez-Ferreiro, S. (2022) Naming as a window to word retrieval changes in healthy and pathological ageing: Methodological considerations. *International Journal of Language & Communication Disorders*, 1–16. https://doi.org/10.1111/1460-6984.12827