Neuropsychologia 231 (2026) 109552

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

A cross-sectional study of category fluency serial recall order in Italian individuals with amnestic mild cognitive impairment from the TREDEM registry: The role of word length

Matteo De Marco [a,*], Chiara Da Ronch [b], Annalena Venneri [c,d], Maurizio Gallucci [b]

[a] Department of Psychology, Brunel University of London, Uxbridge, UK
[b] Cognitive Impairment Centre, Local Health Authority n. 2 Marca Trevigiana, Treviso, Italy
[c] Department of Medicine and Surgery, University of Parma, Parma, Italy
[d] Department of Neuroscience, University of Sheffield, Sheffield, UK

A R T I C L E I N F O

Keywords:
Semantic fluency
Item-level
Verbal fluency
Perirhinal
Transentorhinal

A B S T R A C T

Category Fluency (CF) "item-level" scoring methodologies identify neurocognitive processes that are distinct from those of standard quantitative scoring. One of these accounts for words' serial recall order (SRO) and calculates the correlation between SRO and item-level complexity to operationalise the tendency to name increasingly complex words. We hypothesised this tendency would be altered in individuals with amnestic mild cognitive impairment (aMCI), to reflect lexical-semantic organisational decline resulting from potential functional alterations of perirhinal cortex. We expected aMCI individuals to show higher correlational indices than people with no objective cognitive decline (NOCD) due to these trends being more influenced by word count than overall complexity.

Sixty aMCI and 54 NOCD individuals, recruited by the TREDEM initiative and representative of the North-East Italian population were administered a 1-min "animals" CF. Four complexity features were scored: frequency, age-of-acquisition, typicality and graphemic length. Individual *z*-converted correlation coefficients were calculated between SRO and complexity. The two groups were then compared via uncorrected *t*-tests, age- and education-corrected *ANCOVA*s, and further *ANCOVA*s also correcting for quantitative CF scores.

aMCI individuals named fewer words, and words of lower and more variable complexity. They also showed a significantly stronger "SRO-graphemic length" correlation. This emerged from the uncorrected models and from the first set of *ANCOVA*s.

Significant differences in lexical-semantic organisation exist in aMCI. Length is a diagnostically relevant feature of Italian words generated as part of CF. In line with cross-linguistic generalisability of findings, potential mechanisms differentiating phonological and orthographic length in languages with transparent/opaque orthographies are discussed.

1. Introduction

The Category Fluency Test (CFT) is among the most widely used instruments to assess semantic memory abilities in clinical and research settings (Venneri et al., 2018). A recent meta-analysis that has assessed CFT performance in Alzheimer's disease, in fact, highlights the centrality of semantic abilities and associative retrieval in supporting CFT scores (Olmos-Villaseñor et al., 2023). CFT, however, is also strongly influenced by other non-semantic abilities. These include memory resources such as episodic memory (Greenberg et al., 2009) and memory search (Zemla et al., 2023), lexical generation (Baciu et al., 2021), but also executive abilities like set-shifting (Rende et al., 2002) and speed of processing (Elgamal et al., 2011). The role played by executive functioning has emerged from a number of studies carried out in healthy adults (Aita et al., 2019; Amunts et al., 2021; Shao et al., 2014; Whiteside et al., 2016), and the composite score for executive functioning based on data from the Alzheimer's Disease Neuroimaging Initiative includes CFT performance as part of its structure (Gibbons

* Corresponding author. Department of Psychology, College of Health, Medicine and Life Sciences, Brunel University of London, Kingston Lane, Uxbridge, Middlesex, UB8 3PH, UK.

E-mail address: matteo.demarco@brunel.ac.uk (M. De Marco).

https://doi.org/10.1016/j.neuropsychologia.2026.109552
Received 14 November 2025; Received in revised form 2 July 2026; Accepted 15 July 2026
Available online 16 July 2026

et al., 2012). Semantic control skills, i.e., the ability to manipulate semantic knowledge for the purpose of a task (Lambon Ralph et al., 2017), moreover, also play an integral role in semantic memory testing (Noonan et al., 2013).[1] A schematic representation of some of the abilities that influence CFT performance is shown in Fig. 1. Activation studies provide confirmatory evidence of this multidimensional set of resources, as they show an involvement of inferior-frontal, middle-frontal and anterior-cingulate regions during task engagement (Humphreys and Lambon Ralph, 2025).

While CFT is well-known to rely on diverse cognitive abilities, clinical research in the field of early-stage Alzheimer's disease has focussed on this test because of its role in semantic-memory testing (Venneri et al., 2018). Semantic memory retrieval is supported by the transentorhinal portion of the anterior parahippocampal gyrus (Bastin and Delhaye, 2023; Mishkin et al., 1997), also known as “perirhinal cortex” (i.e., the cytoarchitectural territory corresponding to Brodmann Area 35). This is also the region of the cortex that is affected by early Alzheimer's-related neuropathological changes; in fact, hyperphosphorylated tau changes (i.e., accumulating as neurofibrillary tangles and neuropil threads) are observed in the transentorhinal cortex during Braak Stage I (Braak et al., 2006). This corresponds to a preclinical stage in which no objective episodic-memory symptoms are yet observed, as these become clinically detectable typically only when a person has reached Braak Stages II to IV (Markesbery, 2010). During this preclinical stage, semantic memory alterations are objectively quantifiable, or manifest as a subjective complaint, but not to a degree to emerge as a deficit when measured with standard semantic memory test scores (Huyghe et al., 2025; Marier et al., 2026). It is to address these clinical demands that neuropsychological research can make a significant contribution and propose novel instruments and methodologies that can be more specific to subtle semantic memory decline. Based on these principles, a fine-grained investigation of semantic memory can lead to the discovery and definition of markers of perirhinal/-transentorhinal alterations that can be informative in the context of detection of Alzheimer's disease at the preclinical stage.

Item-level methodologies are an alternative approach to standard CFT scoring. While standard scores are a simple count of correct entries generated by the participant during the 1-min performance, item-level scores are based on the characterisation of each individual entry. Each word is assigned a score that is representative of how semantically complex it is. The concept of item-level complexity has been operationalised in a multitude of ways, such as frequency of use (Pakhomov et al., 2016), age of acquisition (Forbes-McKay et al., 2005), and typicality (Vita et al., 2014). A recent meta-analysis indicates that individuals with a clinical diagnosis of mild cognitive impairment or early Alzheimer's disease tend to generate CFT words that are, on average, more frequent and acquired earlier in life (De Marco et al., 2025).

While item-level scores have now become a consolidated approach to CFT scoring that receives significant attention from clinical researchers (Ferrante et al., 2024; Henderson et al., 2023; Vonk et al., 2023), no consensus has yet been reached on how the set of individual scores should be processed to operationalise test performance. While the vast majority of studies has focused on the average of all individual item-level scores (e.g., by averaging all CFT words), a few studies have proposed alternative scores based on the average of the first three words (Sailor et al., 2004), the first five words (Forbes-McKay et al., 2005), or the five most complex words (Vonk et al., 2023). Recently an alternative method has been proposed to account for word order generation: instead of calculating an average, each word is scored in relation to its serial position within the list of recalled words (De Marco et al., 2021a). The correlation coefficient between item-level complexity and serial recall order (SRO) indicates the degree to which the tested individual can generate words that are increasingly more complex as they search the category in more depth. Serial recall is a fundamental property of memory retrieval that can be associated with biomarker changes proper of early Alzheimer's disease (Bruno et al., 2026). On these grounds, a correlational index that quantifies the association between a consolidated property of retrieval and numerical indices of semantic complexity can represent an intuitive operationalisation of retrieval from semantic memory. As these correlational indices are devised to be less influenced by attentional and executive resources than standard scores (De Marco et al., 2021b, 2025), they can help characterise lexical-semantic alterations. A directional trend is observed in the performance of healthy adults in relation to each index of item-level complexity (De Marco and Venneri, 2022). When compared with young adults aged 18-20, adults older than 70 showed a significant difference in the correlational index between SRO and “valence” (i.e., a property indicating words' pleasantness), with younger adults generating significantly more pleasant words at the start (i.e., words 1-5), and a steeper decrease throughout the rest of the performance (De Marco et al., 2021a). These correlational indices, however, have not yet been investigated in clinical populations.

The assumption on which this study is based suggests that these correlational indices will be valuable to characterise difficulties in semantic memory retrieval in individuals with a clinical diagnosis suggestive of a potential underlying neurodegenerative aetiology. We tested for the first time the role of correlational indices between SRO and item-level complexity in a group of individuals with single-domain amnestic mild cognitive impairment (aMCI) (Petersen et al., 2014), recruited at the Treviso Cognitive Impairment Centre (TCIC) in Treviso, North-Eastern Italy. Although aMCI is not always indicative of its underlying biological changes, it is a syndrome that can have a significant impact on healthcare resources, which emphasises the needs for early detection (Ritchie et al., 2025). In this respect, a better characterisation of lexical-semantic abilities can lead to significant findings that are in line with this priority.

Typically, aMCI individuals show impoverished CFT performance when compared with cognitively unimpaired adults, and this is likely due to a multitude of cognitive mechanisms (Fig. 1). In this context, qualitative and quantitative aspects are thought to play a central role in defining SRO trends. In fact, while we expected both cognitively unimpaired people and people with aMCI to name some of the simplest (if not *the* simplest) category exemplars at the start of their performance, people with aMCI would generate: 1) fewer exemplars (i.e., poorer quantitative performance); and 2) on average, less complex exemplars (i.e., poorer qualitative performance). These two mechanisms should have an opposite effect on correlational trends, with poorer quantitative performance steepening them, and poorer qualitative performance flattening them (Fig. 2). We expected aMCI participants to show poorer quantitative and qualitative CFT performance, and we then assumed that the former would be more influential than the latter on correlational trends, as it has been much more consistently reported in the

[1] It is useful to distinguish between “semantic control” and “executive control” (or executive functioning) in terms of terminology and definitions, as at face value these might be seen as synonyms, or as contiguous or partially overlapping processes. Semantic control is an integral part of semantic cognition (Jackson, 2021) and consists of executive resources that are specific to the semantic domain. During CFT testing, semantic control is the ability one relies on “*to guide retrieval to relevant but less dominant exemplars*”, but also “*to select among competitors*” (Thompson et al., 2026, page 156)” General executive functioning, on the other hand, refers to abilities that are not domain-specific and that can be called upon if demanded by a specific task. The 60-s limit imposed by the CFT is a methodological add-on that is foreign to our regular daily-life use of semantic cognition for lexical-semantic retrieval. For this reason, we argue that processing speed and rapid set-shifting (as shown in Fig. 1) can be intended as part of general executive abilities, rather than domain-specific semantic control. On this note, it is also important to point out that the distinction between “semantic control” and “executive control” is also reflected by the neural patterns that uniquely sustain each function, which contribute to defining the two forms of control as associated yet distinct (Chiou et al., 2023).

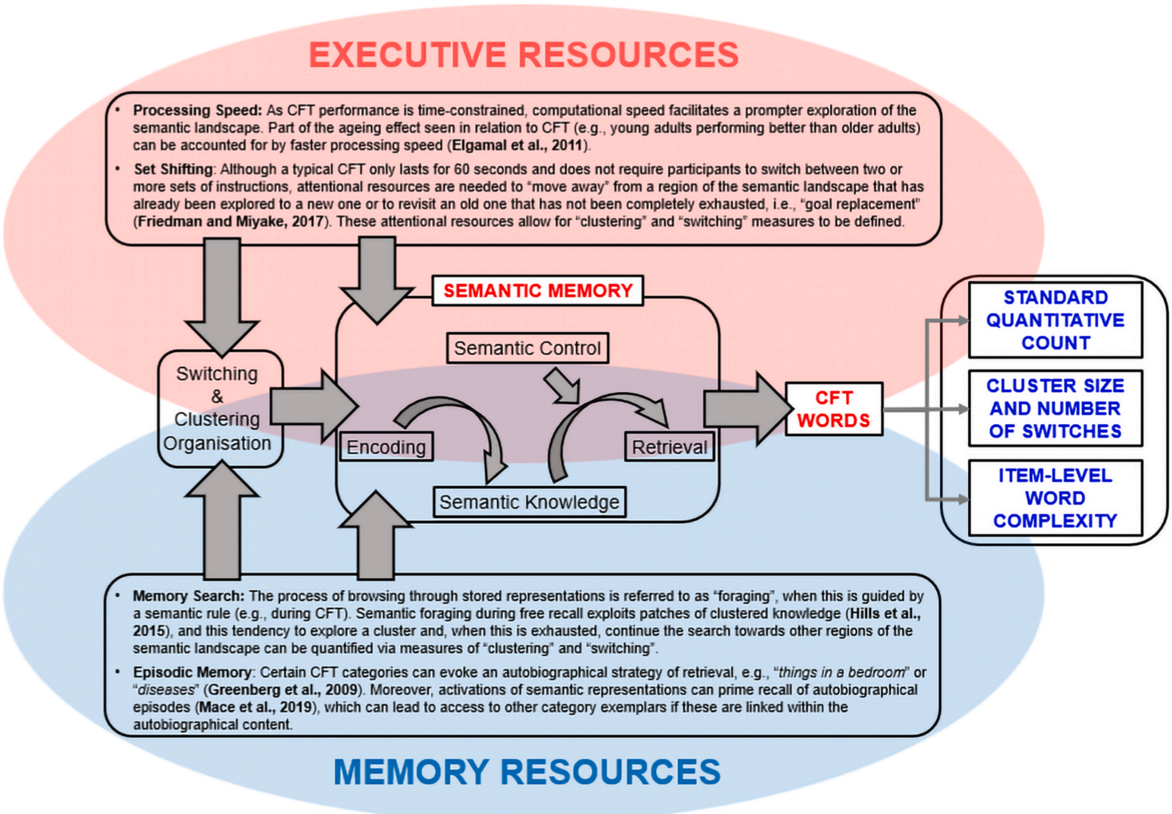


Fig. 1. Schematic representation of CFT performance in relation to the sub-components of semantic memory and the influence played by extra memory and executive resources (illustrated by the blue and red area, respectively).
The image includes four examples of memory (Greenberg et al., 2009; Hills et al., 2015; Mace et al., 2019) and executive (Elgamal et al., 2011; Friedman and Miyake, 2017) skills that have been reported to play a role in influencing CFT performance.

clinical literature. This would lead to steeper correlations in this group reflecting a less thorough exploration of the relevant semantic category. Conversely, individuals with no objective evidence of cognitive decline should be able to search and retrieve semantic information in a more efficient and structured way, and this would translate into more words, and in a more gradual increase in item-level complexity.

Moreover, this study represented an opportunity to test for the first time the SRO-based scoring methodology in a language different from English, as cross-linguistic differences exist in lexical and semantic CFT word properties, such as those of clustering (Supplementary Table 1) and contextual diversity (Supplementary Table 2). Finally, to comply with the principles of Sex and Gender Equity in Research (SAGER) guidelines (Heidari et al., 2016), we report that no sex- or gender-based hypotheses were formulated.

2. Materials and methods

2.1. Participants

A total of 114 CFT performances were included in this cross-sectional study. These were obtained from 60 individuals with a diagnosis of aMCI and 54 individuals who did complete a neuropsychological assessment to address subjective cognitive complaints (or a cognitive concern voiced by an informant) but showed no objective evidence of cognitive decline (NOCD). All participants had attended the TCIC between October 2015 and July 2023 and had been enrolled as part of the TREviso DEMentia (TREDEM) Study and Registry, a clinical-research cohort initiative aimed at studying cognitive impairment and dementia in North-East Italy (Gallucci et al., 2012). All participants included in this study were evaluated by a senior neurogeriatrician or neurologist expert in the field of cognitive impairment and were referred to a neuropsychologist for an extensive neuropsychological examination (Gallucci et al., 2018). All consensus clinical diagnoses of single-domain aMCI followed Petersen's criteria (Petersen et al., 2014). Individuals with NOCD, vice versa, had no objective evidence of cognitive decline, and their scores were indicative of normal cognitive functioning. Both groups were recruited at the same centre and within the same timeframe. All participants were of white ethnicity, and their sociodemographic status was in line with the distribution reported for North-East Italy by the Italian National Institute of Statistics (www.istat.it). They were also asked to report their sex and gender, and these two measurements corresponded for all our participants. As a major long-term objective is clinical implementation of item-level CFT scores, effects sizes were deemed of interest if of sufficient magnitude. Sample size was thus defined to be sufficiently large as to detect an effect size of $d > 0.5$. A *Cohen's d* of 0.5 is typically considered a medium effect size in medical research.

The data analysed and described in this manuscript are part of the TREDEM Study and Registry, that were approved by the Treviso Province Ethics Committee (28th October 2010, and protocol no. 42326 of 10th April 2015, respectively). All procedures were carried out in compliance with the 1964 Declaration of Helsinki, with the local institutional regulations and working practices, and with the privacy rights of participants, as outlined by the European Union's General Data Protection Regulation. Written consent was obtained from each participant.

2.2. Materials

All participants completed the CFT during their neuropsychological assessment. This consisted of four distinct categories (1 min each):

Healthy Control (example)

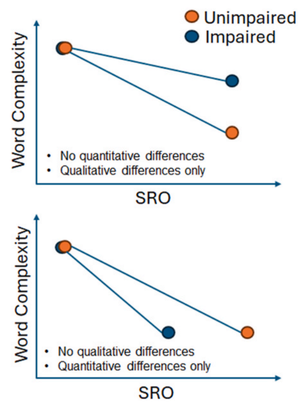
Word	SRO	Age of Acquisition	Frequency
CAT	1	3.073	4.830
DOG	2	3.174	5.168
SHEEP	3	4.181	4.667
GOAT	4	4.682	4.117
ORANGUTAN	5	7.830	2.276
BABOON	6	8.079	3.318
GORILLA	7	6.449	3.726
ANELOPE	8	8.397	3.198
DEER	9	5.212	4.259
ZEBRA	10	5.262	3.692
LION	11	4.430	4.453
TIGER	12	4.418	4.359
BEAR	13	3.925	4.877
KOALA	14	7.598	3.497
PIG	15	4.090	4.511
RAT	16	4.919	4.226
GERBIL	17	8.159	2.823
HAMSTER	18	5.808	3.669
PORCUPINE	19	6.890	3.063
FOX	20	4.045	4.457
PIRANHA	21	8.740	2.949
COYOTE	22	8.788	2.869
CATFISH	23	7.393	2.972
GOLDFISH	24	4.703	3.507

SRO-Age of Acquisition (*Spearman's rho*) = 0.341
SRO-Frequency (*Spearman's rho*) -0.434

MCI Participant (example)

Word	SRO	Age of Acquisition	Frequency
DOG	1	3.174	5.168
CAT	2	3.073	4.830
FISH	3	4.291	5.192
RABBIT	4	3.977	4.389
HORSE	5	3.925	4.990
BIRD	6	3.726	4.853
FLY	7	3.569	4.904
SPIDER	8	3.936	4.242
COW	9	3.515	4.440
ELEPHANT	10	4.702	4.339
GIRAFFE	11	5.120	3.777
ZEBRA	12	5.262	3.692
WILDEBEEST	13	11.310	3.272
AARDVARK	14	9.890	2.741
WALRUS	15	5.224	3.170
SEAL	16	6.162	4.284
BEAR	17	3.925	4.877

SRO-Age of Acquisition (*Spearman's rho*) = 0.689
SRO-Frequency (*Spearman's rho*) -0.662



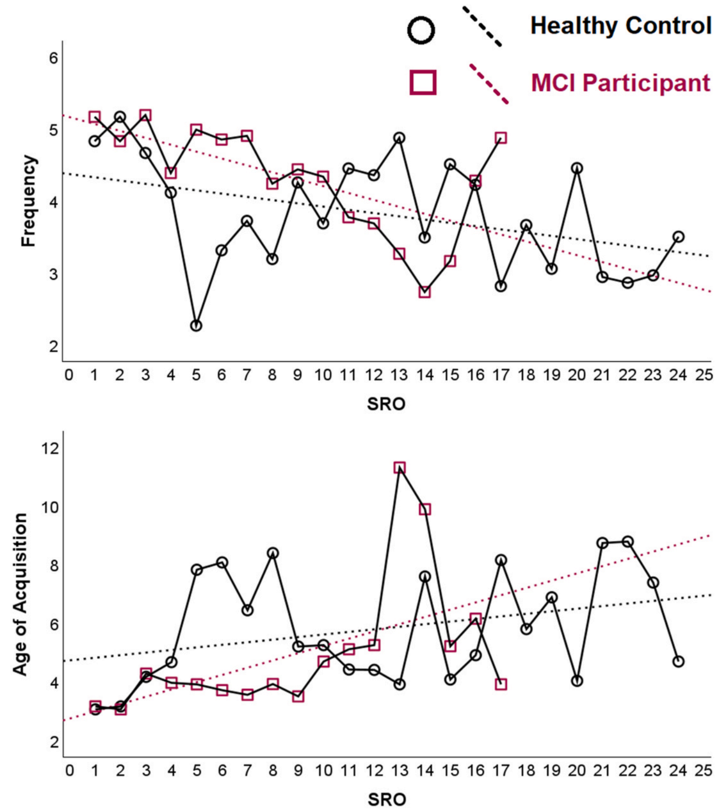


Fig. 2. Schematics of the mechanism at the foundation of the study hypothesis.
CFT performances of a NOCD participant and of a participant with MCI tested in English, with their item-level scores, and visual representation of the rank-based correlation between SRO and two indices of semantic complexity. Age of acquisition and frequency were scored by relying on the English normative data published by van Heuven et al. (2014) and by Kuperman et al. (2012), respectively. The link between poorer quantitative/qualitative performance and steepness of correlational trends is illustrated by the line graph in the middle, i.e., the y axis is in a "more complex to less complex" direction.

animals, fruits, cities and colours. For the purposes of this investigation, the sole "animals" category was retained, as this is by far the most studied category in clinical research (De Marco et al., 2025). All scores on the extended cognitive battery and demographic information about the cohort are reported in Table 1.

CFT words were analysed following the methods outlined in a previous publication (De Marco et al., 2021a). In brief, each individual entry was scored according to its SRO and the following four lexical-semantic features: frequency, age of acquisition, typicality and graphemic length. Words' frequency was scored based on the normative values reported by the SUBTLEX-IT project. These values derive from word counts (i.e., >500,000 words) carried out in a large database of movie subtitles (Crepaldi et al., 2015). Words' age-of-acquisition values were instead drawn from the ItAoA initiative (Montefinese et al., 2019). This database includes ~2000 Italian words, and each value indicates the average estimated age at which the word is learnt. Typicality was not based on normative data, e.g., as those published by Quaranta and colleagues (Quaranta et al., 2016) but reflected the proportion of times a word was named within the cohort (Sailor et al., 2004). Word length, finally, was calculated by counting the number of graphemes used in the orthographic form of the word. These item-level features have been widely used in neurological research (De Marco et al., 2025).

2.3. Scoring procedure

After removing CFT errors (i.e., intrusions and perseverations) four coefficients of correlation were calculated for each individual participant. SRO-frequency, SRO-age of acquisition, SRO-typicality and SRO-length. These were based on Spearman's *rho* since SRO is an ordinal variable. A *rho-to-z* conversion was then carried out (Zar, 2005) to obtain *z*-statistics suitable for group analyses. While the original methodology for scoring English words is based on a strict (and arbitrary) set of rules defining intrusions and perseverations (De Marco et al., 2021a), in this study we discarded entries only if completely foreign to the category, or if a repetition of a previously recalled word. As carried out in our previous research, missing normative values were treated as missing data when calculating correlational indices, i.e., indices were based on the words with an available score.

In addition to the scoring procedure detailed above, each individual performance was also scored following more traditional approaches: First, a global index was calculated for frequency, age of acquisition, typicality, and length, to quantify the average semantic complexity of words throughout the entire performance. Second, all word were assigned to their respective sub-categories for the calculation of clustering (i.e., mean cluster size) and switching (number of switches) scores, following the procedure described by Troyer (2000). The original list of words was adjusted to be suitable for the exemplars generated by this cohort (i.e., the entire list of entries and sub-categories is reported in Supplementary Table S3).

2.4. Data analysis

Quantitative CFT performance and item-level averages (i.e., calculated over the first five entries and over the entire performance) were tested via independent-sample *t*-tests. These analyses served to assess the theoretical assumptions outlined at the end of the Introduction.

Table 1
Characterisation of the study cohort.

Variable	aMCI ($n = 60$)	NOCD ($n = 54$)	Group Difference (p)
Age (years)	73.53 (9.01)	68.76 (9.17)	0.006
Education (years)	9.75 (3.82)	10.31 (3.61)	0.420
Sex and Gender (F/M)†	30/30	34/20	0.164
Mini Mental State Examination	26.58 (2.32)	28.73 (1.20)°	<0.001
Digit Span Test (forward presentation)	5.15 (0.88)	5.55 (0.74)°	0.012
Prose Memory Test (immediate-delayed recall average)	5.97 (3.82)	12.22 (2.26)°	<0.001
Rey Auditory Verbal Learning Test (immediate recall)	24.02 (7.12)	36.79 (9.90)*	<0.001
Rey Auditory Verbal Learning Test (delayed recall)	2.07 (2.07)	7.23 (2.72)*	<0.001
Digit Cancellation Test	46.35 (5.49)	52.47 (5.55)°	<0.001
Letter Fluency Test	26.13 (10.46)	34.98 (10.08)°	<0.001
Category Fluency Test (animals)	14.78 (5.38)	26.91 (8.60)	<0.001
Token Test	30.62 (4.13)	32.21 (3.03)°	0.027
Visuo constructive Praxis Test	8.17 (1.53)	9.37 (1.44)°	<0.001
Clock Drawing Test	8.22 (1.91)	10.47 (8.74)°	0.055
Cognitive Estimates Test	4.48 (0.79)	4.80 (0.46)°	0.011
Activities of Daily Living	5.87 (0.34)	5.86 (0.35)°	0.887
Instrumental Activities of Daily Living	6.15 (1.56)	6.84 (1.48)°	0.021
Global Deterioration Scale	3.69 (2.79)	4.32 (3.03)*	0.272
Hamilton Depression Rating Scale	7.28 (5.15)	8.94 (3.91)°	0.066
Neuropsychiatric Inventory (frequency × severity)	14.98 (11.19)	16.27 (10.02)°	0.535

Means and standard deviations are indicated; °: statistics were calculated on $n = 49$ controls; *: statistics were calculated on $n = 48$ controls. †: gender and sex corresponded for all participants included in this study. Diagnoses of aMCI were made based on the age, education, and gender-corrected normative values of each individual test, as this is the clinical standard followed by Italian memory clinics. Between-group differences in uncorrected scores, therefore, are not necessarily indicative of a deficit.

Additionally, measures of clustering and switching were also analysed with independent-sample *t*-tests.

Three sets of models were designed to test the study hypothesis. First, uncorrected inferential models were set up: independent-sample *t*-tests were calculated to compare the two clinical groups' sets of correlational indices. Second, one-way analyses of covariance (*ANCOVAs*) were designed to infer group differences while controlling for age and years of education. Third, a further set of *ANCOVAs* was designed correcting for age, years of education and quantitative CFT scores. These analyses were also carried out via hierarchical regressions, to help compare the models by inspecting the r^2 coefficient. In line with the study hypothesis, one-sided thresholds of significance (i.e., $p < 0.05$) were applied.

3. Results

3.1. General neuropsychological profiles

As shown in Table 1, the group of aMCI participants was significantly older than the NOCD individuals, but no between-group differences in sex ratios or years of education were found. Mini Mental State Examination scores were significantly lower in the group of aMCI yet, on average, well above the cut-off expected in this diagnostic group. No differences, moreover, were found in activities of daily living between the two groups, in line with Petersen's criteria (Petersen et al., 2014). Although significant group differences were also seen in non-memory tests, the aMCI group had scores below normative cut-offs only on memory tests.

3.2. CFT quantitative counts, item-level averages, and clustering-switching scores

A statistically significant difference was found between the two groups in the quantitative count of valid CFT entries, with aMCI participants generating significantly fewer valid entries ($t_{87.102} = 8.911$, $p < 0.001$, $d = 1.711$).

The cohort named a total 2340 words, corresponding to 266 different animals. Frequency scores were available for 255 of the 266 animals (i.e., 2329 entries). Given the considerably smaller normative database, age-of-acquisition values were only available for 101 animals (i.e., 1722 entries). As typicality and length were not based on normative data, these were scored for 100% of the entries.

On average, the first five words generated by the two groups did not differ in their age of acquisition ($t_{112} = -0.250$, $p = 0.803$, $d = -0.047$), frequency ($t_{112} = -0.537$, $p = 0.592$, $d = -0.101$), typicality ($t_{112} = 0.074$, $p = 0.941$, $d = 0.014$), or length ($t_{112} = -1.642$, $p = 0.103$, $d = -0.308$).

On average, words generated by aMCI participants calculated over the entire performance were acquired earlier in life ($t_{112} = -3.978$, $p < 0.001$, $d = -0.746$), more frequent ($t_{92.890} = 5.496$, $p < 0.001$, $d = 1.001$), more typical ($t_{101.386} = 5.109$, $p < 0.001$, $d = 0.937$), and shorter ($t_{109.383} = -3.943$, $p < 0.001$, $d = -0.730$) than the NOCD group.

aMCI participants also made fewer switches ($t_{112} = 8.073$, $p < 0.001$, $d = 1.514$) and generated smaller clusters ($t_{112} = 2.713$, $p = 0.008$, $d = 0.509$) than the NOCD group. All these findings are reported in Table 2.

Table 2
Item-level and clustering/switching properties of CFT performance.

CFT Measure	aMCI	NOCD	*p*-value	*Cohen's d*
Age of Acquisition - Global Average	4.24 (0.37)	4.48 (0.26)	<0.001	−0.746
Frequency - Global Average	2023.36 (720.52)	1434.35 (391.23)	<0.001	1.001
Typicality - Global Average	0.33 (0.08)	0.27 (0.05)	<0.001	0.937
Length - Global Average	6.14 (0.52)	6.49 (0.40)	<0.001	−0.73
Age of Acquisition - First 5 Words	3.83 (0.54	3.85 (0.57)	0.803	−0.047
Frequency - First 5 Words	3171.03 (1480.03)	3318.86 (1452.98)	0.592	−0.101
Typicality - First 5 Words	0.45 (0.12)	0.45 (0.13)	0.941	0.014
Length - First 5 Words	5.74 (0.76)	5.98 (0.79)	0.103	−0.308
Number of Switches	7.77 (3.64)	13.80 (4.33)	<0.001	−1.514
Mean Cluster Size	1.62 (0.54)	1.89 (0.53)	0.008	−0.509
Age of Acquisition - Intra-cluster change	−0.09 (0.65)	−0.25 (0.83)	0.262	0.214
Frequency - Intra-cluster change	850.22 (2453.08)	989.42 (1520.22)	0.714	−0.067
Typicality - Intra-cluster change	0.04 (0.14)	0.06 (0.09)	0.448	−0.140
Length - Intra-cluster change	0.10 (1.37)	−0.18 (1.17)	0.238	0.222

Means and standard deviations are indicated; aMCI: amnestic mild cognitive impairment; NOCD: participants with no objective cognitive deficits.

3.3. SRO-based correlational trends

None of the correlational indices was associated with age but all of them correlated with years of education, with absolute r-scores between 0.24 and 0.34. SRO-frequency and SRO-age of acquisition were not significantly correlated with quantitative CFT scores, while SRO-typicality and SRO-length were (i.e., $r = 0.223$ and $r = -0.293$, respectively; Fig. 3). The variance inflation factor, however, was <1.1 in both cases, indicating a modest relationship between quantitative and SRO-based qualitative aspects of performance.

Uncorrected analyses revealed no differences between groups in the correlational indices between SRO and age of acquisition ($t_{107.28} = 1.24$, $p = 0.108$, $d = 0.172$), typicality ($t_{112} = -1.62$, $p = 0.054$, $d = 0.306$), or frequency ($t_{112} = -0.914$, $p = 0.181$, $d = 0.231$). A significant difference was instead found in association with the SRO-length index (Fig. 4), with aMCI individuals showing significant larger correlation coefficients ($t_{88.21} = 2.92$, $p = 0.002$, $d = 0.539$). While this index had a positive sign (indicating a positive correlation, i.e., words were increasingly longer during test performance) in both groups, this trend was more exacerbated in the aMCI group ($z = 0.289$, versus a $z = 0.126$ calculated in the group of NOCD).

The first set of analysis-of-covariance models (i.e., corrected for age and years of education) confirmed the results of the t tests: no effect of diagnosis emerged from the analysis of the correlational indices between SRO and frequency ($F_{1,110} = 1.25$, $p = 0.266$, *partial* $\eta^2 = 0.011$), typicality ($F_{1,110} = 2.37$, $p = 0.127$, *partial* $\eta^2 = 0.021$), or age of acquisition ($F_{1,110} = 0.88$, $p = 0.350$, *partial* $\eta^2 = -0.008$). A significant effect of diagnosis was instead found in association with the SRO-length index ($F_{1,110} = 8.96$, $p = 0.003$, *partial* $\eta^2 = 0.075$).

The second set of analysis-of-covariance models (i.e., corrected for age, years of education and quantitative CFT scores) showed no significant effect of diagnosis: SRO-frequency: $F_{1,109} = 0.339$, $p = 0.562$, *partial* $\eta^2 = 0.003$; SRO-typicality: $F_{1,109} = 0.180$, $p = 0.672$, *partial* $\eta^2 = 0.001$; SRO-age of acquisition: $F_{1,109} = 0.037$, $p = 0.847$, *partial* $\eta^2 < 0.001$; SRO-length: $F_{1,109} = 1.644$, $p = 0.203$, *partial* $\eta^2 = 0.025$.

In order to obtain a more intuitive statistical index that could allow us to compare the effect of diagnostic group on each z-converted correlational index, all three sets of analyses were rerun as hierarchical regression models. This analytical approach was used to quantify and compare the r^2-*change* coefficients across models. These are shown in Table 3.

The two analyses that reported between-group differences were significant at a p-value lower than the 0.0042 corresponding to a Bonferroni-corrected threshold accounting for 12 independent models. All sets of inferential models were also rerun by retaining only participants who generated ≥15 words (i.e., 28 aMCI and 53 NOCD; see Supplementary Fig. S1 for a distribution of scores among those participants producing <15 words). All findings were confirmed.

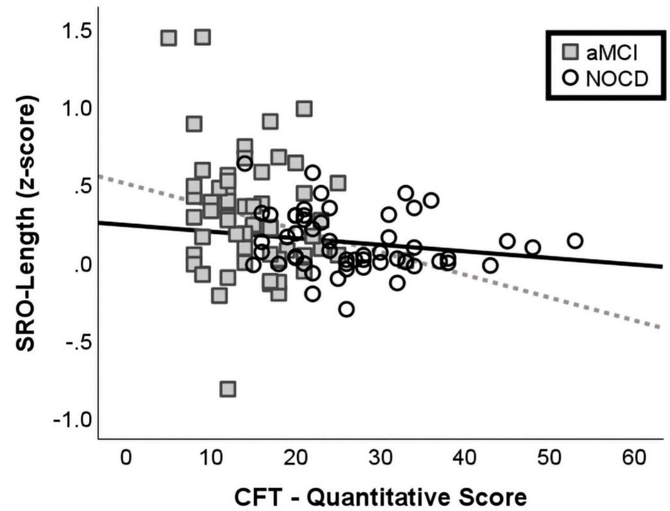


Fig. 4. Scatterplot showing the association between quantitative CFT scores and the correlational index between SRO and word length (i.e., the only one associated with a significant effect of group from the set of uncorrected models). The continuous black line and the grey dotted line were fitted to represent datapoints in the NOCD and aMCI group, respectively.

Table 3
Uncorrected and corrected models analysed via hierarchical regression models.

Hierarchical Regression Model	Age of Acquisition	Frequency	Typicality	Graphemic Length
Uncorrected Analyses				
BLOCK 1: Diagnostic Group	0.013 (**1.3%**)	0.007 (**0.7%**)	0.023 (**2.3%**)	0.067 (**6.7%**)
Corrected Analyses - Set 1				
BLOCK 1: Age and Years of Education	0.063 (**6.3%**)	0.139 (**13.9%**)	0.068 (**6.8%**)	0.064 (**6.4%**)
BLOCK 2: Diagnostic Group	0.070 (**+0.7%**)	0.149 (**+1%**)	0.088 (**+2%**)	0.134 (**+7%**)
Corrected Analyses - Set 2				
BLOCK 1: Age, Years of Education, and Quantitative CFT Count	0.077 (**7.7%**)	0.148 (**14.8%**)	0.099 (**9.9%**)	0.143 (**14.3%**)
BLOCK 2: Diagnostic Group	0.077 (**+0%**)	0.150 (**+0.2%**)	0.100 (**+0.1%**)	0.156 (**+1.3%**)

r^2 (BLOCK 1) and r^2-*change* (BLOCK 2) coefficients are shown. Percentages and block-to-block increases in percentages are reported.

3.4. Intra-cluster change in word complexity

While naming increasingly difficult words is a tendency that

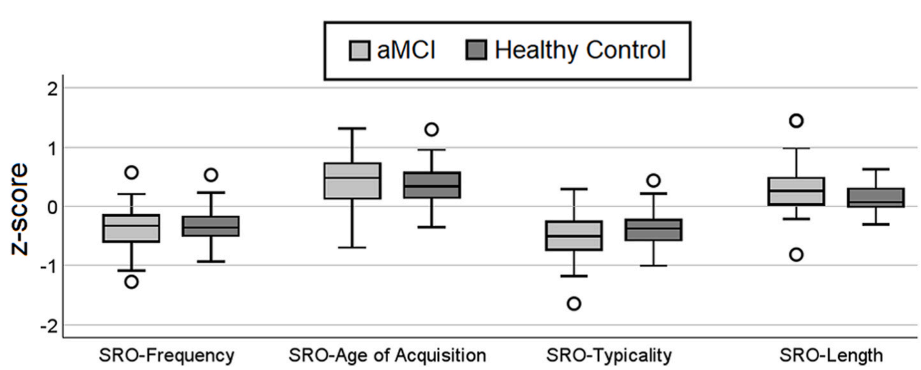


Fig. 3. Boxplot graphs showing the distribution of correlational indices in the two diagnostic groups. Supplementary Fig. S1 illustrates where, within these distributions, those who generated <15 words are located.

characterises CFT performance over the entire 1-min performance, this trend may also be influenced by the cluster-based organisation of retrieval. We tested this additional hypothesis *post hoc*, by focussing on each individual cluster, and evaluating the intra-cluster change in word complexity. To do so, we calculated the difference between the first and the last word of each cluster, and we averaged these change-scores within each participant's performance. One sample *t*-tests were then run to test the one-tailed hypothesis by which change scores would be significantly different from zero and indicative of an intra-cluster trend towards increased complexity. A significant and strong effect was found in the NOCD group in relation to decreasing frequency ($t_{53} = 4.783$, $p < 0.001$, $d = 0.651$) and decreasing typicality ($t_{53} = 5.057$, $p < 0.001$, $d = 0.688$), while a more modest effect was found in relation to increasing age of acquisition ($t_{53} = -2.165$, $p = 0.008$, $d = -0.295$). In the aMCI group, only decreasing frequency ($t_{59} = 2.685$, $p = 0.005$, $d = 0.347$) and decreasing typicality ($t_{59} = 2.483$, $p = 0.017$, $d = 0.318$) demonstrated a moderately significant effect, while the findings for age of acquisition were not significant ($t_{56} = -1.007$, $p = 0.159$, $d = -0.133$). Finally, the change scores for word length did not emerge as significant in any of the two groups (NOCD: $t_{53} = -1.147$, $p = 0.128$, $d = -0.156$; aMCI: $t_{59} = -0.577$, $p = 0.283$, $d = 0.075$). We also compared intra-cluster change scores between the two groups, but none of the models resulted in significant differences.

4. Discussion

This is the first study investigating the differences in CFT performance between aMCI individuals and individuals with subjective deficits but no objective evidence of cognitive decline, with a specific focus on the link between word complexity and the order of word retrieval. SRO is an ordinal variable that defines the hierarchy of free recall within the semantic constraints of the target category. The association between SRO and an index of item complexity indicates the propensity of an individual to access a semantically-cued lexicon and generate words that are increasingly difficult. It was previously demonstrated that these correlational indices show a significant directional trend in healthy adults (De Marco and Venneri, 2022). Our findings (reported in Fig. 4) confirm these trends when the test is administered in a language different from English, and are found alongside a reduced quantitative output, and an overall tendency to name simpler entries.

Although four correlational indices were investigated, only one of these showed significant differences between the two diagnostic groups: that between SRO and word length. This emerged from the set of uncorrected analyses, as well as from the analyses corrected for age and years of education. Non-significant trends emerged from the uncorrected models in association with the other three indices, but these trends were considerably downsized after controlling for age and years of education. In comparison with frequency, age of acquisition and typicality, graphemic length does not convey any semantic content, but reflects a lexical property of words. This indicates that the main property on which individuals with NOCD and those with aMCI differ is the degree to which their performance consists of increasingly longer words. Moreover, as we had assumed quantitative differences to be more influential on these trends than qualitative differences (Fig. 2), the lack of statistical findings emerging from the third set of inferential models (i.e., those also controlling for quantitative performance) does support our assumption.

Although the aMCI participant with the strongest SRO-length correlation (i.e., equal to a z-score of 1.44) had generated <10 words, the other 11 participants who had similarly generated <10 words had less strong correlational z-scores (i.e., averaging at 0.380). This confirms lack of a strong link between quantitative scores and SRO-length correlational indices (which is also illustrated in Fig. 4), and the fact that higher coefficients in the aMCI group are not exclusively a by-product of this group generating fewer words. All findings, in fact, were replicated after retaining in the analyses only participants with a quantitative score of 15 and above. Quantitative scores, however, did play a significant role in the second set of *ANCOVA* models. In fact, when analyses included quantitative scores as a correction factor, the results were no longer significant. The vast majority of studies that have investigated item-level CFT scores by studying the difference in performance between a group of controls and a group of participants with a neurological condition did not control for quantitative CFT scores (De Marco et al., 2025). To date, only five studies have included quantitative performance as a control factor (De Marco et al., 2021a; van den Berg et al., 2024; Vita et al., 2014; Vonk et al., 2019, 2023). This is a methodological aspect that should be considered in future studies testing item-level CFT scores.

4.1. Constraints on generalisation

The effect documented in relation to word length deserves further attention since, differently from English, the Italian language has a transparent orthography. Published evidence indicates that orthographic characteristics of a language influence speech production, e.g., (Peereman et al., 2009; Rastle et al., 2011). When languages are characterised by a transparent grapheme-to-phoneme correspondence, orthographic length of words tends to coincide with phonological length. As a result, it is possible that orthographic length plays a more relevant role in CFT testing in transparent languages. An investigation carried out in English and Spanish, however, has shown that both orthographic and phonological length are significantly correlated with semantic features of words like frequency or age of acquisition, and this appears to be the case for both languages (Davies et al., 2016). Another study, however, reported that phonological (but not orthographic) similarity of words generated during a CFT was predictive of quantitative performance in native English speakers with MCI (Clark et al., 2014), suggesting a dissociation between the two indices at this diagnostic stage. In addition, phonological length of English CFT words has been associated with executive functioning, as shorter words possess a wider phonological neighbourhood that needs to be inhibited during task performance (Rofes et al., 2020).[2] We argue that this, however, would mostly affect shorter words generated during the second part of the test, as early words are typically generated via automatic processing (Hurks et al., 2004). It is also important to point out that the impact of word length on time-constrained word generation may differ from language to language. While a very large number of animals in English are represented by one-syllable words, this is not the case in other languages, and this might lead to cross-linguistic differences in CFT performance (Kempler et al., 1998).

When performance is analysed as a function of generation time, in 10/15-s quantiles in cognitively unimpaired adults and individuals with MCI, the majority of words are produced at the start of performance (Fernaeus et al., 2008). This suggests that correlational indices calculated across the entire performance are more influenced by automatic (than controlled) word retrieval, as more of the earlier words contribute to the calculation of the coefficients. To have a graphical overview of the overall findings and help interpret our data, we plotted word length as a function of order of recall for the first 20 words in the two diagnostic groups. No trend can be immediately recognised by inspecting length of individual words or of groups of 5 consecutive words (Fig. 5). What emerges is a considerably larger variability in the aMCI group, possibly due to a range of different types of episodic memory difficulties, with

[2] Operationalising the orthographic neighbourhood via the 'Levenshtein Distance' metric (i.e., the number of words differing from the target by one graphemic insertion, deletion, or substitution, as scored by the resources available at https://www.dcode.fr/levenshtein-distance), short words such as HORSE (16 neighbours), MOUSE (19 neighbours), or GOAT (20 neighbours) have larger orthographic neighbourhoods than longer words such as STARLING (9 neighbours), PHEASANT (3 neighbours) or SQUIRREL (1 neighbour).

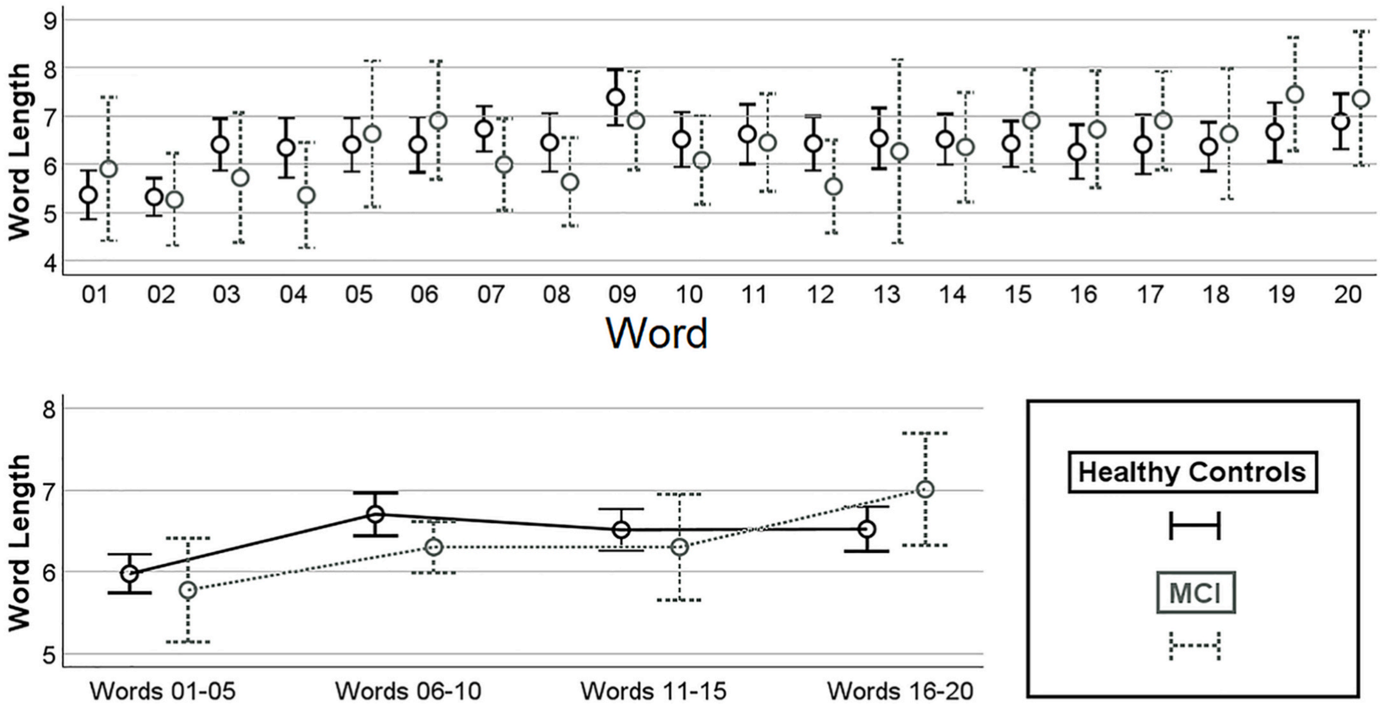


Fig. 5. Visual representation of word length in association with up to the first 20 words generated by each participant.
Individual words and the mean of words 1-5, 6-10, 11-15 and 16-20 are shown. Error bars represent the 95% confidence intervals.

these expressing as deficits on either or both the Rey Auditory Verbal Learning Test or the Prose Memory Test, and as a result of impairment in a range of processes being involved, i.e., encoding, access, retrieval, and control resources. Future studies based on larger cohorts should consider taking this clinical element into account and could extend the focus on the comparison of aMCI and non-amnestic MCI, or of single-domain and multiple domain MCI, or should possibly include measures of Alzheimer's biomarkers, to focus on groups that are mechanistically more homogeneous. To date, in fact, the literature on the study of item-level CFT scoring is almost entirely based on studies of clinically-diagnosed participants (De Marco et al., 2025), with one study only reporting the use of biomarkers (Ferrante et al., 2024). On this note, while multiple aetiologies may cause amnestic symptoms, Alzheimer's disease or another neurogenerative condition is typically a prime aetiological suspect when a clinical diagnosis of aMCI is made. For this reason, given the established link between semantic memory and integrity of the perirhinal cortex, item-level approaches to CFT scoring like the one presented in this study can be an informative yet inexpensive route to early detection of individuals with possible perirhinal alterations.

Beyond its use in clinical research, item-level scoring holds promise as a method for investigating memory-related search. The perirhinal cortex is responsible for the retrieval of information that is “context free” (Mishkin et al., 1997). When a semantic category is searched (or “foraged”), however, this is also sustained by an increase in hippocampal theta oscillations (Viganò et al., 2025), which characterise the operational mode of the hippocampus (Buzsáki, 2002). It is possible that an altered cross-talk between perirhinal and hippocampal areas may spare the ability to retrieve context-free category exemplars, in a general sense, but would not allow deployment of hippocampal resources that can grant contextual-binding facilitations, such as, for example, a more efficient cluster-to-cluster strategy, or the use of other resources, such as visual imagery or autobiographical episodes. From this viewpoint, the effect observed in relation to graphemic length could indicate over-reliance in the aMCI group on a mechanism that, not reflecting any contextual binding, and not embodying any semantic information, is not useful at supporting performance. In this respect, this study expands research in the field of semantic foraging and underscores the need to approach this area by focussing on multiple languages.

4.2. Study limitations

This study is not free from limitations. While word length, typicality and frequency of use were scored for the entire (or vast majority of the) database, age of acquisition norms were only available for 73.6% of the total CFT words generated by the entire cohort (i.e., we were able to score 1722 of 2340 words that represent the global output). Although we cannot rule out that different SRO correlational indices would have been obtained if 100% of the words had been scored in relation to each lexical and semantic feature, the existing normative values did allow us to score the majority of the words produced by our participants and to calculate individual age of acquisition trends (i.e., more details are provided in Supplementary Fig. S2). In this respect, however, while a large amount of normative data is available for English words, fewer options are available, as of now, in other languages (Banks and Connell, 2023). Secondly, while the positional order of each word operationalised by SRO reflects a general property of retrieval from long-term lexical-semantic memory, we do not know to what degree this is linked to the mechanisms of access and elaboration of lexical-semantic memories. Studies investigating ‘relational’ properties of retrieved words such as clustering and switching suggest an elaboration of semantic content occurring during CFT (i.e., as it is an elaboration of content that cues retrieval of semantically associated words). However, it is likely that words rapidly generated during the first part of the test are simply accessed but not elaborated in depth. As a result, it needs still to be established to what degree correlational indices linked to SRO are influenced by the various mechanisms of memory retrieval.

4.3. Conclusions

In summary, we found a significant difference between aMCI and individuals with NOCD attending a memory clinic in the correlational index between CFT SRO and word length. aMCI individuals showed stronger correlational indices than NOCD individuals, indicating a more pronounced propensity to generate words that are increasingly longer during the course of the 1-min performance. This effect, however, was no longer significant after controlling for CFT quantitative performance (i.e., the number of words generated). No significant findings emerged instead from the analysis of the correlational indices between SRO and words' frequency, age of acquisition, or typicality. This suggests that, when CFT is tested in Italian individuals with aMCI, lexical, rather than semantic coding seems to play a more distinctive role in the processes of normal-vs.-abnormal ageing. More studies are warranted to explore this area of study in depth, possibly thanks to the application of novel and lexically-richer normative data, which are largely available in English, but less so in other languages.

Data statement

All raw data collected and scored as part of this project are available on Brunel University of London Figshare repository: https://doi.org/10.17633/rd.brunel.28532192.v1.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

No generative AI software or service was used in the manuscript preparation process.

Funding

This work was supported by a grant from the Alzheimer's Association (23AARG-1030190), awarded to MDM. AV is supported by funding obtained under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 341 of 15/03/2022 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU, Project code PE0000006, Concession Decree No. 1553 of 11/10/2022 adopted by the Italian Ministry of University and Research, CUP D93C22000930002, "A multiscale integrated approach to the study of the nervous system in health and disease" (MNESYS).

CRediT authorship contribution statement

Matteo De Marco: Conceptualization, Formal analysis, Funding acquisition, Methodology, Visualization, Writing – original draft. **Chiara Da Ronch:** Data curation, Writing – review & editing. **Annalena Venneri:** Funding acquisition, Resources, Writing – review & editing. **Maurizio Gallucci:** Investigation, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

For the purposes of open access, the author has applied a Creative Commons Attribution (CC BY) Licence to any Accepted Author Manuscript version arising from this submission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2026.109552.

References

Aita, S.L., Beach, J.D., Taylor, S.E., Borgogna, N.C., Harrell, M.N., Hill, B.D., 2019. Executive, language, or both? An examination of the construct validity of verbal fluency measures. Appl. Neuropsychol. Adult 26, 441–451. https://doi.org/10.1080/23279095.2018.1439830.

Amunts, J., Camilleri, J.A., Eickhoff, S.B., Patil, K.R., Heim, S., von Polier, G.G., Weis, S., 2021. Comprehensive verbal fluency features predict executive function performance. Sci. Rep. 11, 6929. https://doi.org/10.1038/s41598-021-85981-1.

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, 1725–1765. https://doi.org/10.1007/s11357-021-00367-5.

Banks, B., Connell, L., 2023. Category production norms for 117 concrete and abstract categories. Behav. Res. Methods 55, 1292–1313. https://doi.org/10.3758/s13428-021-01787-z.

Bastin, C., Delhaye, E., 2023. Targeting the function of the transentorhinal cortex to identify early cognitive markers of Alzheimer's disease. Cognit. Affect Behav. Neurosci. 23, 986–996. https://doi.org/10.3758/s13415-023-01093-5.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., Del Tredici, K., 2006. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol. 112, 389–404. https://doi.org/10.1007/s00401-006-0127-z.

Bruno, D., Reichert-Plaska, C., Jauregi-Zinkunegi, A., Ashton, N.J., Zetterberg, H., Blennow, K., Pomara, N., 2026. Associations of plasma p-tau 231 with serial position recall performance in free-of-dementia individuals. J. Neuropsychol. 20, 291–300. https://doi.org/10.1111/jnp.70012.

Buzsáki, G., 2002. Theta oscillations in the hippocampus. Neuron 33, 325–340. https://doi.org/10.1016/s0896-6273(02)00586-x.

Chiou, R., Jefferies, E., Duncan, J., Humphreys, G.F., Lambon Ralph, M.A., 2023. A middle ground where executive control meets semantics: the neural substrates of semantic control are topographically sandwiched between the multiple-demand and default-mode systems. Cerebr. Cortex 33, 4512–4526. https://doi.org/10.1093/cercor/bhac358.

Clark, D.G., Wadley, V.G., Kapur, P., DeRamus, T.P., Singletary, B., Nicholas, A.P., Blanton, P.D., Lokken, K., Deshpande, H., Marson, D., Deutsch, G., 2014. Lexical factors and cerebral regions influencing verbal fluency performance in MCI. Neuropsychologia 54, 98–111. https://doi.org/10.1016/j.neuropsychologia.2013.12.010.

Crepaldi, D., Amenta, S., Mandera, P., Keuleers, E., Brysbaert, M., 2015. Quality, not quantity. Register is more important than size in corpus–based frequency [data set]. In: 19th Conference of the European Society for Cognitive Psychology, Paphos, Cyprus, 17-20 September 2015.

Davies, S.K., Izura, C., Socas, R., Dominguez, A., 2016. Age of acquisition and imageability norms for base and morphologically complex words in English and in Spanish. Behav. Res. Methods 48, 349–365. https://doi.org/10.3758/s13428-015-0579-y.

De Marco, M., Blackburn, D.J., Venneri, A., 2021a. Serial recall order and semantic features of category fluency words to study semantic memory in normal ageing. Front. Aging Neurosci. 13, 678588. https://doi.org/10.3389/fnagi.2021.678588.

De Marco, M., Vonk, J.M.J., Quaranta, D., 2021b. Editorial: the mechanistic and clinical principles of item-level scoring methods applied to the category fluency test and other tests of semantic memory. Front. Psychol. 21, 1152574. https://doi.org/10.3389/fpsyg.2023.1152574.

De Marco, M., Wright, L.M., Makovac, E., 2025. Item-level analysis of category fluency test performance: a systematic review and meta-analysis of studies of normal and neurologically-abnormal ageing. Neuropsychol. Rev. https://doi.org/10.1007/s11065-024-09657-z (in press).

De Marco, M., Venneri, A., 2022. Serial recall order of category fluency words: exploring its neural underpinnings. Front. Psychol. 12, 777838. https://doi.org/10.3389/fpsyg.2021.777838.

Elgamal, S.A., Roy, E.A., Sharratt, M.T., 2011. Age and verbal fluency: the mediating effect of speed of processing. Can. Geriatr. J. 14, 66–72. https://doi.org/10.5770/cgj.v14i3.17.

Fernaeus, S.E., Östberg, P., Hellström, A., Wahlund, L.O., 2008. Cut the coda: early fluency intervals predict diagnoses. Cortex 44, 161–169. https://doi.org/10.1016/j.cortex.2006.04.002.

Ferrante, F.J., Migeot, J., Birba, A., Amoruso, L., Pérez, G., Hesse, E., Tagliazucchi, E., Estienne, C., Serrano, C., Slachevsky, A., Matallana, D., Reyes, P., Ibáñex, A., Fittipaldi, S., Gonzalez Campo, C., García, A.M., 2024. Multivariate word properties in fluency tasks reveal markers of Alzheimer's dementia. Alzheimers Dement. 20, 925–940. https://doi.org/10.1002/alz.13472.

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, 1625–1632. https://doi.org/10.1016/j.neuropsychologia.2005.01.008.

Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: individual differences as a window on cognitive structure. Cortex 86, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023.

Gallucci, M., Di Battista, M.E., Battistella, G., Falcone, C., Bisiacchi, P.S., Di Giorgi, E., 2018. Neuropsychological tools to predict conversion from amnestic mild cognitive impairment to dementia. The TREDEM registry. Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn. 25, 550–560. https://doi.org/10.1080/13825585.2017.1349869.

Gallucci, M., Mariotti, E., Saraggi, D., Stecca, T., Oddo, M.G., Bergamelli, C., Boldrini, P., Mazzuco, S., Ongaro, F., Mecocci, P., Di Paola, F., Bendini, M., Forloni, G.L., Albani, D., Antuono, P., Caberlotto, L., Zanardo, A., Siculi, M., Gajo, G.B., Durante, E., Buscato, G., 2012. The Treviso dementia (TREDEM) study: a biomedical, neuroradiological, neuropsychological and social investigation of dementia in north-eastern Italy. J. Frailty Aging 1, 24–31. https://doi.org/10.14283/jfa.2012.5.

Gibbons, L.E., Carle, A.C., Mackin, R.S., Harvey, D., Mukherjee, S., Insel, P., Mackay Curtis, S., Mungas, D., Crane, P.K., for the Alzheimer's Disease Neuroimaging Initiative, 2012. A composite score for executive functioning, validated in Alzheimer's disease neuroimaging initiative (ADNI) participants with baseline mild cognitive impairment. Brain Imaging Behav 6, 517–527. https://doi.org/10.1007/s11682-012-9176-1.

Greenberg, D.L., Keane, M.M., Ryan, L., Verfaellie, M., 2009. Impaired category fluency in medial temporal lobe amnesia: the role of episodic memory. J. Neurosci. 29, 10900–10908. https://doi.org/10.1523/jneurosci.1202-09.2009.

Heidari, S., Babor, T.F., De Castro, P., Tort, S., Curno, M., 2016. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. Res. Integr. Peer Rev. 1, 2. https://doi.org/10.1186/s41073-016-0007-6.

Henderson, S.K., Peterson, K.A., Patterson, K., Lambon Ralph, M.A., Rowe, J.B., 2023. Verbal fluency tests assess global cognitive status but have limited diagnostic differentiation: evidence from a large-scale examination of six neurodegenerative diseases. Brain Comm 5, fcad042. https://doi.org/10.1093/braincomms/fcad042.

Hills, T.T., Todd, P.M., Jones, M.N., 2015. Foraging in semantic fields: how we search through memory. Top. Cogn. Sci. 7, 513–534. https://doi.org/10.1111/tops.12151.

Humphreys, G.F., Lambon Ralph, M.A., 2025. Mapping the task- general and task-specific neural correlates of speech production: meta- analysis and fMRI direct comparisons of category fluency and picture naming. Imaging Neurosci 3. https://doi.org/10.1162/IMAG.a.154. IMAG.a.154.

Hurks, P.P.M., Hendriksen, J.G.M., Vles, J.S.H., Kalff, A.C., Feron, F.J.M., Kroes, M., van Zeben, T.M.C.B., Steyaert, J., Jolles, J., 2004. Verbal fluency over time as a measure of automatic and controlled processing in children with ADHD. Brain Cogn. 55, 535–544. https://doi.org/10.1016/j.bandc.2004.03.003.

Huyghe, L., Salman, Y., Colmant, L., Gérard, T., Malotaux, V., Besson, G., Delhaye, E., Bastin, C., Dessain, Q., Dricot, L., Lhommel, R., Ivanoiu, A., Quenon, L., Hanseeuw, B., 2025. Preclinical detection of Alzheimer's disease pathology using conceptual discrimination abilities. J. Prev. Alzheimers Dis. 12, 100332. https://doi.org/10.1016/j.tjpad.2025.100332.

Jackson, R.L., 2021. The neural correlates of semantic control revisited. Neuroimage 224, 117444. https://doi.org/10.1016/j.neuroimage.2020.117444.

Kempler, D., Teng, E.L., Dick, M., Taussig, I.M., Davis, D.S., 1998. The effects of age, education, and ethnicity on verbal fluency. J. Int. Neuropsychol. Soc. 4, 531–538. https://doi.org/10.1017/s1355617798466013.

Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M., 2012. Age-of-acquisition ratings for 30,000 English words. Behav. Res. Methods 44, 978–990. https://doi.org/10.3758/s13428-012-0210-4.

Lambon Ralph, M.A., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. Nat. Rev. Neurosci. 18, 42–55. https://doi.org/10.1038/nrn.2016.150.

Mace, J.H., McQueen, M.L., Hayslett, K.E., Staley, B.J.A., Welch, T.J., 2019. Semantic memories prime autobiographical memories: general implications and implications for everyday autobiographical remembering. Mem. Cognit. 47, 299–312. https://doi.org/10.3758/s13421-018-0866-9.

Marier, A., Fernández Arias, J., Aumont, É., Hall, B.J., Macedo, A.C., Rahmouni, N., Bezgin, G., Vitali, P., Rosa-Neto, P., Montembeault, M., 2026. Language deficits across PET-based Braak stages of tau accumulation in Alzheimer's disease. Alzheimers Dement. 22, e71286. https://doi.org/10.1002/alz.71286.

Markesbery, W.R., 2010. Neuropathologic alterations in mild cognitive impairment: a review. J. Alzheimers Dis. 19, 221–228. https://doi.org/10.3233/jad-2010-1220.

Mishkin, M., Suzuki, W.A., Gadian, D.G., Vargha-Khadem, F., 1997. Hierarchical organization of cognitive memory. Philos. Trans. R. Soc. Lond. B Biol. Sci. 352, 1461–1467. https://doi.org/10.1098/rstb.1997.0132.

Montefinese, M., Vinson, D., Vigliocco, G., Ambrosini, E., 2019. Italian age of acquisition norms for a large set of words (ItAoA). Front. Psychol. 10, 278. https://doi.org/10.3389/fpsyg.2019.00278.

Noonan, K.A., Jefferies, E., Visser, M., Lambon Ralph, M.A., 2013. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. J. Cogn. Neurosci. 25, 1824–1850. https://doi.org/10.1162/jocn_a_00442.

Olmos-Villaseñor, R., Sepulveda-Silva, C., Julio-Ramos, T., Fuentes-Lopez, E., Toloza-Ramirez, D., Santibañez, R.A., Copland, D.A., Mendez-Orellana, C., 2023. Phonological and semantic fluency in Alzheimer's disease: a systematic review and meta-analysis. J. Alzheimers Dis. 95, 1–12. https://doi.org/10.3233/jad-221272.

Pakhomov, S.V.S., Eberly, L., Knopman, D., 2016. Characterizing cognitive performance in a large longitudinal study of aging with computerized semantic indices of verbal fluency. Neuropsychologia 89, 42–56. https://doi.org/10.1016/j.neuropsychologia.2016.05.031.

Peereman, R., Dufour, S., Burt, J.S., 2009. Orthographic influences in spoken word recognition: the consistency effect in semantic and gender categorization tasks. Psychon. Bull. Rev. 16, 363–368. https://doi.org/10.3758/pbr.16.2.363.

Petersen, R.C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., Fratiglioni, L., 2014. Mild cognitive impairment: a concept in evolution. J. Intern. Med. 275, 214–228. https://doi.org/10.1111/joim.12190.

Quaranta, D., Caprara, A., Piccininni, C., Vita, M.G., Gainotti, G., Marra, C., 2016. Standardization, clinical validation, and typicality norms of a new test assessing semantic verbal fluency. Arch. Clin. Neuropsychol. 31, 434–445. https://doi.org/10.1093/arclin/acw034.

Rastle, K., McCormick, S.F., Bayliss, L., Davis, C.J., 2011. Orthography influences the perception and production of speech. J. Exp. Psychol. Learn. Mem. Cogn. 37, 1588–1594. https://doi.org/10.1037/a0024833.

Rende, B., Ramsberger, G., Miyake, A., 2002. Commonalities and differences in the working memory components underlying letter and category fluency tasks: a dual-task investigation. Neuropsychology 16, 309–321. https://doi.org/10.1037//0894-4105.16.3.309.

Ritchie, C., Trepel, D., Edwards, S., Hahn-Pedersen, J.H., Chan, M.S., Bray, B.D., Clark, A., Wichmann, C.A., Evans, M., 2025. The impact of mild cognitive impairment on healthcare utilization and costs: a UK Biobank study. Alzheimers Dement. 17, e70065. https://doi.org/10.1002/dad2.70065.

Rofes, A., de Aguiar, V., Jonkers, R., Oh, S.J., DeDe, G., Sung, J.E., 2020. What drives task performance during animal fluency in people with Alzheimer's disease? Front. Psychol. 11, 1485. https://doi.org/10.3389/fpsyg.2020.01485.

Sailor, K., Antoine, M., Diaz, M., Kuslansky, G., Kluger, A., 2004. The effects of Alzheimer's disease on item output in verbal fluency tasks. Neuropsychology 18, 306–314. https://doi.org/10.1037/0894-4105.18.2.306.

Shao, Z., Janse, E., Visser, K., Meyer, A.S., 2014. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. Front. Psychol. 5, 772. https://doi.org/10.3389/fpsyg.2014.00772.

Thompson, H.E., Sowden, P.T., Cogdell-Brooke, L., Violante, I.R., Jefferies, B., 2026. Category fluency and creative potential in semantic aphasia. J. Neuropsychol. 20, 154–174. https://doi.org/10.1111/jnp.70019.

Troyer, A., 2000. Normative data for clustering and switching on verbal fluency tasks. J. Clin. Exp. Neuropsychol. 22, 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370.

van den Berg, E., Dijkzeul, J.C.M., Poos, J.M., Eikelboom, W.S., van Hemmen, J., Franzen, S., de Jong, F.J., Dopper, E.G.P., Vonk, J.M.J., Papma, J.M., Satoer, D., Jiskoot, L.C., Seelaar, H., 2024. Differential linguistic features of verbal fluency in behavioral variant frontotemporal dementia and primary progressive aphasia. Appl. Neuropsychol. Adult 31, 669–677. https://doi.org/10.1080/23279095.2022.2060748.

van Heuven, W.J.B., Mandera, P., Keuleers, E., Brysbaert, M., 2014. SUBTLEX-UK: a new and improved word frequency database for British English. Q. J. Exp. Psychol. 67, 1176–1190. https://doi.org/10.1080/17470218.2013.850521.

Venneri, A., Jahn-Carta, C., De Marco, M., Quaranta, D., Marra, C., 2018. Diagnostic and prognostic role of semantic processing in preclinical Alzheimer's disease. Biomark. Med. 12, 637–651. https://doi.org/10.2217/bmm-2017-0324.

Viganò, S., Giari, G., Mai, R., Doeller, C.F., Bottini, R., 2025. Foraging in conceptual spaces: hippocampal oscillatory dynamics underlying searching for concepts in memory. BioRxiv PrePrint. https://doi.org/10.1101/2025.10.07.680852.

Vita, M.G., Marra, C., Spinelli, P., Caprara, A., Scaricamazza, E., Castelli, D., Canulli, S., Gainotti, G., Quaranta, D., 2014. Typicality of words produced on a semantic fluency task in amnesic mild cognitive impairment: linguistic analysis and risk of conversion to dementia. J. Alzheimers Dis. 42, 1171–1178. https://doi.org/10.3233/jad-140570.

Vonk, J.M.J., Flores, R.J., Rosado, D., Qian, C., Cabo, R., Habegger, J., Louie, K., Allocco, E., Brickman, A.M., Manly, J.J., 2019. Semantic network function captured by word frequency in nondemented APOE ε4 carriers. Neuropsychology 33, 256–262. https://doi.org/10.1037/neu0000508.

Vonk, J.M.J., Geerlings, M.I., Avila-Rieger, J.F., Qian, C.L., Schupf, N., Mayeux, R., Brickman, A.M., Manly, J.J., 2023. Semantic item-level metrics relate to future memory decline beyond existing cognitive tests in older adults without dementia. Psychol. Aging 38, 443–454. https://doi.org/10.1037/pag0000747.

Whiteside, D.M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M.R., Roper, B., 2016. Verbal fluency: language or executive function measure? Appl. Neuropsychol. Adult 23, 29–34. https://doi.org/10.1080/23279095.2015.1004574.

Zar, J.H., 2005. Spearman rank correlation: overview. In: Armitage, P., Colton, T. (Eds.), Encyclopaedia of Biostatistics. John Wiley and Sons, Hoboken, NJ. https://doi.org/10.1002/0470011815.b2a15150.

Zemla, J.C., Gooding, D.C., Austerweil, J.L., 2023. Evidence for optimal semantic search throughout adulthood. Sci. Rep. 13, 22528. https://doi.org/10.1038/s41598-023-49858-9.