Clustering and Switching Dynamics in Alzheimer's Disease During Property Listing

Diego Morales (dmoralesb@alumnos.uai.cl)[1], Felipe D. Toro Hernandez[2], Sergio E. Chaigneau[3] & Enrique Canessa[4]

[1]Faculty of Engineering and Science, Universidad Adolfo Ibáñez, Santiago, Chile
[2]Center of Mathematics, Computing and Cognition, Federal University of ABC, São Paulo, Brazil
[3]Center for Social and Cognitive Neuroscience, Universidad Adolfo Ibáñez, Santiago, Chile
[4]Escuela de Ingeniería Informática, Universidad de Valparaíso, Chile

Abstract

Clustering and switching analyses have characterized semantic memory deficits in Alzheimer's disease (AD) using fluency tasks, where participants list category members within time constraints. We applied these analyses to the property listing task (PLT). The PLT differs fundamentally from fluency tasks: it is non-timed, involves an open-ended rather than a finite set of responses, and includes varied property types (perceptual, functional, taxonomic). Using temporal slope difference clustering on data from 24 AD and 35 control participants, we found that AD participants generated fewer properties and exhibited reduced switch rates between temporal clusters, even when controlling for task duration. Cluster size and within-cluster timing were preserved, but AD participants showed longer switch times and reduced within-cluster semantic variability. These findings reveal how clustering and switching manifest in property generation, suggesting that AD affects strategic exploration while preserving local retrieval within temporal clusters.

Keywords: property listing task; Alzheimer's disease; clustering; switching; semantic memory

Introduction

Semantic memory refers to our multi-modal knowledge of objects, people, facts, words, concepts, and the properties associated with those concepts (Lambon Ralph et al., 2017). Understanding how people access and navigate semantic memory is central to cognitive science. In Alzheimer's disease (AD), semantic memory impairments are prominent and can be detected even in prodromal stages (Fagundo et al., 2008; Saranpää et al., 2022). The semantic fluency task (SFT), which requires generating category members (e.g., animals) within a time limit, has proven particularly sensitive in differentiating AD individuals from older adults who are cognitively healthy (Fagundo et al., 2008; Weakley & Schmitter-Edgecombe, 2014).

In the SFT, participants are asked to generate as many exemplars as possible from a given semantic category (e.g., animals) within a fixed time limit, typically 60 seconds. The SFT performance can be decomposed into two theoretically distinct component processes (Troyer et al., 1997). Clustering, the production of words within a semantic subcategory (e.g., listing "dog, cat, hamster" as pets), is thought to reflect the integrity of semantic memory stores and temporal lobe functioning. Switching, the transition between subcategories (e.g., moving from pets to farm animals), is thought to reflect executive control processes such as cognitive flexibility and strategic search, associated with frontal lobe integrity. Evidence for this dissociation comes from lesion studies, where frontal lesions were associated with impaired switching (Okruszek et al., 2013; Troyer et al., 1998) and left temporal lobe lesions reduced clustering (Troyer et al., 1998). Functional neuroimaging has further revealed hippocampal and cerebellar activation during switching, with ramping neural activity patterns consistent with strategic monitoring processes (Lundin et al., 2023).

This clustering and switching decomposition has proven valuable for characterizing cognitive profiles in clinical populations. Studies of semantic fluency in AD have revealed complex patterns of impairment in both processes. Fagundo et al. (2008) found that AD patients generated fewer correct words and showed reductions in both cluster size and number of switches compared to healthy controls. They also found that in a longitudinal sample of individuals with memory complaints, cluster size was the best predictor of subsequent conversion to AD, outperforming total word count and switching measures. Weakley and Schmitter-Edgecombe (2014) similarly found that AD patients produced fewer words, fewer switches, and smaller clusters on semantic fluency.

The pattern of impairment varies across disease stages. Bertola et al. (2014) found that the number of new subcategories generated differentiated healthy controls from all clinical groups, including early-stage amnestic mild cognitive impairment (MCI), while switching was reduced only in more advanced stages. Computational analyses further show that AD patients return to previously visited subcategories less often than controls, suggesting compromised strategic re-exploration of semantic space (Saranpää et al., 2022). Additionally, a recent meta-analysis (De Marco et al., 2025) confirmed that individuals with MCI and AD generate words of lower semantic complexity, specifically words of higher frequency and later age-of-acquisition, compared to healthy controls.

Theoretical frameworks from optimal foraging theory (Hills et al., 2015) conceptualize semantic memory search as internal foraging, where clustering reflects local exploitation of a semantic "patch" and switching reflects exploration for new regions when the current patch is depleted. From this perspective, the pattern of impairments observed in AD (reduced switching and smaller clusters) suggests an exploitation bias (Wyatt et al., 2024), where patients show reduced exploration of the semantic space and over-reliance on readily accessible

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

neighborhoods. If this exploitation bias reflects a domain-general impairment in semantic search strategy rather than a task-specific deficit, it should manifest similarly across different semantic memory tasks. However, the clustering and switching framework has only been validated in the SFT. A complementary task that probes semantic memory in a structurally different way could test whether these deficits generalize beyond category member retrieval.

The Property Listing Task and the Present Study

One such complementary task is the property listing task (PLT; also known as the feature listing task), in which participants list properties that are typically true of a given concept (e.g., for "dog", one might list "has four legs", "is a pet", "barks"). Unlike the SFT, where participants retrieve coordinate members of a category (e.g., different animals), the PLT requires listing features or properties associated with a single concept (McRae et al., 2005; Muraki et al., 2020). These structural differences are theoretically important. First, while category members in the SFT form a relatively finite set, conceptual properties may be open-ended (Canessa & Chaigneau, 2020; Chaigneau et al., 2018). Second, the PLT typically does not impose a strict time limit, allowing for a potentially different exploration of semantic space than the constrained search required in the SFT. Third, the semantic space traversed differs: in the SFT, participants move between subcategories within a superordinate category (e.g., pets to farm animals to wild animals), whereas in the PLT, they navigate across different types of properties (perceptual, functional, categorical, thematic). Properties can also vary in specificity, from general features (e.g., "has a color") to highly specific ones (e.g., "barks loudly when strangers approach") (Chaigneau et al., 2018).

These structural differences suggest that the PLT may reveal aspects of semantic memory organization not fully captured by the SFT. While the SFT primarily tests navigation among coordinate category members, the PLT tests the ability to access and organize various types of semantic information associated with a single concept. The PLT has been shown to be sensitive to neurodegenerative diseases. Toro-Hernández et al. (2024) found that Parkinson's disease patients produced more concrete and imageable property words than controls and showed reduced semantic variability. These patterns successfully discriminated patients from healthy controls. However, no studies have yet applied clustering and switching metrics (which successfully differentiate AD from controls in the SFT) to PLT data in clinical populations. As free-generation paradigms belonging to the same family of semantic memory tasks (Lagos et al., 2026), both involve the sequential production of semantically related content from memory, suggesting that optimal foraging theory (Hills et al., 2015), which conceptualizes such production in terms of local exploitation (clustering) and strategic exploration (switching), may provide a productive analytical framework for the PLT as well.

The key methodological challenge in applying this framework to the PLT is defining what constitutes a cluster. While the SFT has traditionally relied on pre-defined semantic subcategories (e.g., pets, farm animals), the open-ended nature of property generation makes such pre-specification difficult. We therefore adopt a temporal approach to cluster identification, defining clusters based on the temporal dynamics of property generation rather than on predetermined semantic categories. This approach is supported by evidence that inter-response times reflect semantic distance. Semantically related items are produced more rapidly in succession than items from different semantic neighborhoods (Montez et al., 2015). By analyzing these temporal patterns using the slope difference technique (Bushnell et al., 2022), we can identify clusters and switches without relying on subjective categorization schemes.

In the present study, we analyze PLT data from healthy older adults and individuals with AD, applying clustering and switching analyses to sequences of generated properties. Based on evidence that AD reflects an exploitation bias in semantic search (Wyatt et al., 2024), we predict that AD patients will: (1) produce fewer properties overall, reflecting reduced semantic search capacity; (2) generate smaller semantic clusters, indicating reduced local exploitation of semantic neighborhoods; and (3) exhibit fewer switches between property types, suggesting impaired strategic exploration between clusters. Observing these patterns would suggest that the exploitation bias in AD generalizes beyond exemplar generation (SFT) to feature listing (PLT), revealing how AD affects strategic navigation of semantic memory across different task structures and types of semantic relations.

Method

Participants

Sixty older adults participated in this study, comprising 24 individuals diagnosed with Alzheimer's disease (AD group; M_{age} = 76.88 years, SD = 8.80; 14 female, 10 male) and 35 healthy-matched controls (CN group; M_{age} = 71.83 years, SD = 6.12; 23 female, 12 male). All participants were native Spanish speakers from Chile. AD participants were diagnosed by a group of dementia experts following the NINCDS-ADRDA criteria (McKhann et al., 1984). The mean time since diagnosis in the AD group was 1.47 years (SD = 2.60). Control participants had no history of psychiatric or neurological disease. The groups did not significantly differ in years of education (CN: M = 13.71, SD = 4.13; AD: M = 11.88, SD = 4.79; $F(1, 58) = 2.474$, $p = .121$) or sex ($\chi^2 = 0.091$, $p = .763$), but differed in age ($F(1, 58) = 6.768$, $p = .012$); therefore, age was included as a covariate in subsequent analyses. Cognitive status was assessed with the Montreal Cognitive Assessment (MoCA), showing significant differences between groups (CN: M = 24.87, SD = 3.23; AD: M = 13.52, SD = 6.00; $F(1, 58) = 87.95$, $p < .001$). All participants provided their informed consent according to institutional ethical guidelines.

Procedure

Participants completed the property listing task for ten concrete concepts: tree, airplane, bed, house, duck, clown, hair, puma, sun, and shark. The concepts were given orally by the examiner in random order, and the participants were asked to name as many properties as they could regarding each concept, including perceptual features, structural components, typical uses, and locations. The entire procedure was recorded for further audio analysis.

Data Processing

Audio recordings were transcribed verbatim, preserving the temporal sequence of property generation.

Two independent raters coded each transcribed response using a six-category coding scheme: (1) *Valid Property*: a novel property directly derived from the target concept (e.g., for TREE: "has leaves"); (2) *Sub-property*: a property of a previously mentioned property rather than the target concept; (3) *Iteration*: valid properties repeated within the same superordinate dimension (e.g., "can be tall", "can be short"); (4) *Repetition (Property)*: conceptual repetition of a previously coded property; (5) *Repetition (Concept)*: restating the target concept itself; and (6) *Metacognitive/Off-topic*: task-related comments or off-task utterances. The Inter-rater reliability was substantial (Cohen's $\kappa = 0.747$, 85.25% agreement, n = 7775 coded responses). Discrepancies were resolved through discussion between two authors of the study to produce the final consensus codes.

Response times were extracted from the audio recordings using Adobe Audition. Three independent raters annotated (i) the total duration of property generation for each concept, measured as the cumulative elapsed time from the start of the first response interval (including initial pre-response latency) to the end of the final response, and (ii) the latency between properties, defined as the time elapsed between the onset of successive properties.

Responses coded as Repetition (Property), Repetition (Concept), or Metacognitive/Off-topic (codes 4, 5, and 6) were excluded from subsequent clustering and similarity analyses, as they do not represent novel conceptual properties. Excluded items were removed from the sequence. Inter-response times between the remaining items were computed from absolute cumulative onset times (e.g., if B is excluded from A–B–C, the inter-response time equals $t_C - t_A$, preserving the full elapsed interval).

To quantify semantic similarity between consecutive properties, we computed pairwise cosine similarity using the BAAI/bge-m3 embedding model (Chen et al., 2024), a multilingual transformer model with 1024-dimensional embeddings that supports over 100 languages, including Spanish. For each participant-concept combination, we generated embeddings for all unique property responses and calculated the cosine similarity between each property and its immediate predecessor in the temporal sequence.

Clustering and Switching Metrics

We identified cluster boundaries using the slope difference technique (Bushnell et al., 2022), a data-driven temporal approach that does not require pre-defined semantic categories. This method offers several advantages over alternatives. Unlike fixed inter-response time thresholds (which may miss individual or concept-level variation in production speed), it adapts to each participant's baseline production rate by fitting the exponential model individually to each participant-concept pair; and, unlike manual semantic coding, it does not require a priori judgments about which properties belong to the same semantic category. We first expressed each participant's performance as a cumulative function of properties generated over time (Figure 1A). For each participant-concept combination, we fit an exponential curve (saturation model; Bousfield and Sedgewick 1944) to this cumulative function: $y = c(1 - e^{-mt})$, where t represents the cumulative time in seconds, c and m are fitted parameters, and e is the base of the natural logarithm. This exponential curve represents the predicted rate of property generation, with the parameter m capturing the rate of exponential saturation.

The algorithm then compared the slope of the actual cumulative curve to the slope of the fitted exponential curve at time points halfway between each consecutive pair of properties (Figure 1B). The slope difference was computed as the difference between the actual slope (change in the number of properties generated divided by change in time) and the predicted slope from the exponential model. Properties produced faster than predicted (positive slope difference) were considered to be semantically related to the previous property and thus part of the same cluster. Properties produced more slowly than predicted (negative slope difference) were considered switches to a new cluster. Bushnell et al. (2022) found that timing-based metrics derived from slope difference cluster boundaries improved prediction of progressive cognitive decline, making this method particularly appropriate for differentiating controls from participants with cognitive impairment. The slope difference method is implemented in the latest version of the *forager* toolbox (Kumar et al., 2023).

Inter-response time (IRT) was calculated as the time elapsed between the onset of consecutive property productions, measured from the timestamped audio transcriptions. A cluster (or "chain") was defined as a maximal consecutive sequence of properties connected by positive slope differences. The number of switches was calculated as the total number of negative slope differences across the entire property list for each participant-concept combination.

From the slope difference analysis, we derived multiple dependent variables for each participant-concept combination, organized into four categories (Table 1): *production metrics* (list length, production rate), *clustering and switching metrics* (switch rate, average cluster size), *timing metrics* (overall, within-cluster, and between-cluster temporal dynamics), and *semantic similarity metrics* (cosine similarity between consecutive properties using BGE embeddings, computed both

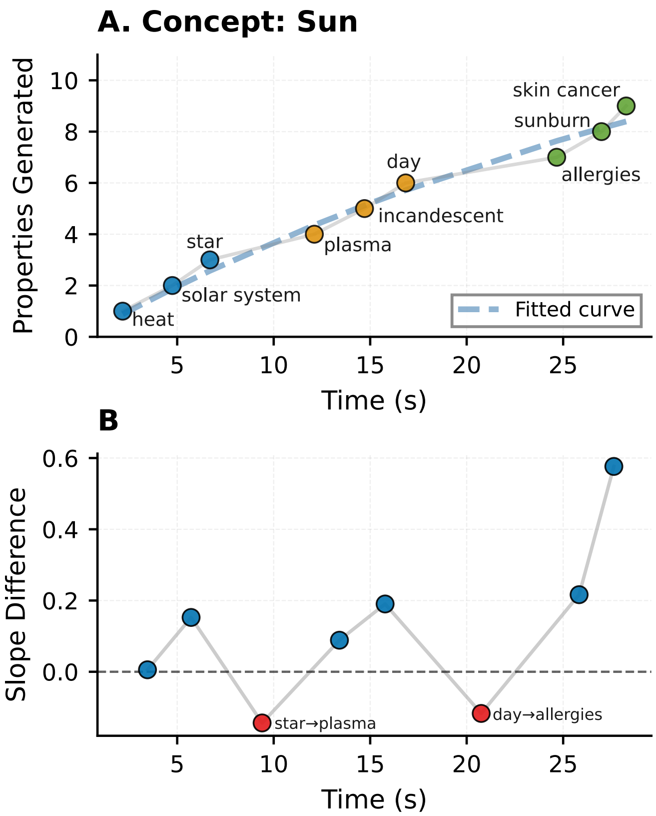

Figure 1: Slope difference method for identifying cluster boundaries. **(A)** Cumulative property generation over time for concept "sun" from a representative study participant, with dot colors indicating cluster membership. The dashed line shows the fitted exponential curve. **(B)** Slope differences at property transitions. Blue circles (positive values) indicate within-cluster retrieval; red circles (negative values) indicate cluster switches, labeled with the corresponding property transition.

within and across cluster boundaries).

Statistical Analysis

We conducted two sets of analyses to characterize diagnostic group differences. First, to test whether the distribution of response types differed between groups, we analyzed all response-level data using chi-square tests of independence and two-proportion z-tests. The chi-square test assessed whether response type distribution was independent of diagnostic group. For post-hoc comparisons of individual response codes (valid properties, sub-properties, iterations, property repetitions, concept repetitions, and metacognitive/off-topic responses), we conducted two-proportion z-tests comparing the proportion of responses in each category between CN and AD groups, with false discovery rate (FDR) correction using the Benjamini-Hochberg procedure to control for multiple comparisons across the six response types.

Second, to assess diagnostic group differences in production, clustering, switching, timing, semantic similarity, and speed metrics while accounting for the nested data structure (multiple concepts per participant), we fit generalized linear mixed-effects models (GLMMs) using the glmmTMB package in R (Brooks et al., 2017). Models included fixed effects of diagnostic group (CN as reference, so positive coefficients indicate AD > CN), the time spent listing properties for each concept, and age, with random intercepts for both participant and concept. Because GLMMs with non-Gaussian families use link functions (log for count and timing variables; logit for proportions), reported β coefficients are on the link scale rather than on the response scale. Cohen's d is reported alongside β as an interpretable effect size. Time was included as a covariate since the PLT was non-timed, allowing us to test whether group differences persist after accounting for task duration. We note that time differences may themselves reflect disease-related processes (e.g., reduced persistence), so group differences in time remain substantively interpretable. GLMMs distributions were assessed using DHARMa R package (Hartig et al., 2016) for residual diagnostics.

Results

Response Type Distribution

Response type distribution differed significantly between groups, $\chi^2(5) = 203.25$, $p < .001$, Cramér's $V = 0.19$. Post-hoc two-proportion z-tests with FDR correction revealed four significant differences. CN participants produced more valid properties (69.7%) than AD participants (54.8%), $z = 11.09$, FDR-adjusted $p < .001$. Conversely, AD participants produced more metacognitive/off-topic responses (18.1% vs. 7.1%), $z = -12.58$, FDR-adjusted $p < .001$, more concept repetitions (6.9% vs. 4.5%), $z = -3.71$, FDR-adjusted $p < .001$, and more property repetitions (7.6% vs. 6.0%), $z = -2.25$, FDR-adjusted $p = .037$. Sub-properties and iterations did not differ between groups (both FDR-adjusted $p > .40$). This pattern suggests that AD participants generate a smaller proportion of conceptually distinct properties and show increased repetition and off-task behavior, consistent with semantic memory and executive control deficits and with the elevated perseveration rates reported in AD fluency (Zemla & Austerweil, 2019).

Total Task Duration

Before examining production and cognitive metrics, we assessed whether diagnostic groups differed in total time spent on the task. AD participants spent significantly less time listing properties (M = 32.90 s, SD = 16.32) compared to CN participants (M = 38.38 s, SD = 13.18), $t(375.25) = 4.20$, $p < .001$. Given that the PLT was non-timed and this difference may reflect reduced engagement or earlier exhaustion of accessible properties, all subsequent analyses included total time as a covariate to isolate cognitive processes beyond overall task engagement.

Production and Clustering Metrics

To assess group differences, we fitted GLMMs with random intercepts for participants and concepts, with fixed effects of

Table 1: Dependent variables derived from the slope difference clustering analysis.

Metric	Definition
Production Metrics	
List length	Total number of valid properties generated (excluding repetitions and metacognitive utterances)
Production rate	Number of valid properties divided by total task duration (properties/second)
Clustering and Switching Metrics	
Switch rate	Number of switches divided by list length (switches/property)
Average cluster size	Mean number of consecutive properties within each semantic cluster
Timing Metrics	
Mean IRT	Mean inter-response time (onset-to-onset) across all consecutive property pairs (seconds)
Within-cluster IRT	Mean inter-response time for properties within the same cluster (seconds)
Switch delay	Mean pause duration before the first property of a new cluster (seconds)
Time in cluster	Mean duration from onset of first property to offset of last property in each cluster (seconds)
Mean switch time	Mean inter-response time (onset-to-onset) at cluster boundaries; unlike switch delay, includes articulation of the last property in the preceding cluster (seconds)
Semantic Similarity Metrics	
Within-cluster similarity	Mean cosine similarity between consecutive properties within the same cluster (BGE embeddings)
Between-cluster similarity	Mean cosine similarity across cluster boundaries (BGE embeddings)
Within-cluster variability	Standard deviation of cosine similarities within clusters
Between-cluster variability	Standard deviation of cosine similarities across cluster boundaries

diagnostic group (CN as reference), total time, and age.

Production metrics. Even after controlling for total task time, AD participants generated significantly fewer properties (CN: M = 8.28, SD = 3.80; AD: M = 6.00, SD = 2.74; β = -0.25, z = -3.64, p < .001, d = 0.67) and showed a lower production rate (CN: M = 0.23, SD = 0.09; AD: M = 0.21, SD = 0.09 properties/s; β = -0.19, z = -2.80, p = .005, d = 0.22) compared to CN participants. Mean inter-response time was significantly longer for AD participants (CN: M = 5.13 s, SD = 3.01; AD: M = 5.55 s, SD = 3.07; β = 0.41, z = 2.82, p = .005, d = 0.14), reflecting slower property generation. These findings indicate that reduced production in AD participants reflects genuine impairment in semantic retrieval capacity rather than simply spending less time on the task.

Clustering and switching metrics. AD participants showed a significantly lower switch rate (CN: M = 0.29, SD = 0.13; AD: M = 0.26, SD = 0.16 switches/property; β = -0.68, z = -2.27, p = .023, d = 0.21), indicating fewer transitions between temporal clusters per property generated. Average cluster size did not significantly differ between groups (CN: M = 2.45, SD = 0.82; AD: M = 2.45, SD = 1.01; β = 0.11, z = 0.72, p = .473). The combination of reduced switching but preserved cluster size suggests that AD participants produce fewer, but not smaller, temporal clusters. The divergence between switch rate and cluster size is expected. Switch rate is defined as $(C-1)/N$, so shorter lists yield lower rates even with identical cluster sizes, which accounts for the significance difference between these metrics.

Temporal Dynamics

Within-cluster timing. Within-cluster inter-response time (CN: M = 3.27 s, SD = 2.76; AD: M = 3.42 s, SD = 2.35; β = 0.21, z = 1.45, p = .147) and time in clusters (CN: M = 6.79 s, SD = 5.01; AD: M = 7.32 s, SD = 6.08; β = 0.27, z = 1.77, p = .076) did not significantly differ between groups. This suggests that once within a temporal cluster, AD participants retrieve properties at a similar pace as controls.

Switch timing. AD participants showed significantly longer mean switch time (CN: M = 8.46 s, SD = 4.34; AD: M = 9.93 s, SD = 6.26; β = 0.39, z = 2.65, p = .008, d = 0.28). However, switch delay (the silence period from the offset of the last property in a cluster to the onset of the first property of the new cluster) did not significantly differ (CN: M = 3.26 s, SD = 2.47; AD: M = 2.96 s, SD = 2.62; β = -0.24, z = -1.00, p = .318). This dissociation indicates that the longer switch times in AD reflect slower articulation of the last property in each cluster rather than extended pre-cluster pauses.

Semantic Similarity

Mean similarity. Average within-cluster similarity (CN: M = 0.54, SD = 0.07; AD: M = 0.54, SD = 0.07; β = -0.0003, z = -0.02, p = .982) and between-cluster similarity (CN: M = 0.51, SD = 0.07; AD: M = 0.51, SD = 0.09; β = -0.001, z = -0.10, p = .922) did not significantly differ between groups. This indicates that the overall semantic coherence of clusters and the semantic distance across cluster boundaries are similar in both groups.

Similarity variability. AD participants showed significantly lower variability in within-cluster similarity (CN: M = 0.08, SD = 0.04; AD: M = 0.06, SD = 0.04; β = -0.012, z = -2.56, p = .010, d = 0.50), suggesting more homogeneous semantic neighborhoods within clusters. Between-cluster similarity variability did not significantly differ (CN: M = 0.06, SD = 0.03; AD: M = 0.06, SD = 0.04; β = 0.006, z = 1.16, p = .246). The reduced within-cluster variability suggests that AD participants may rely on more stereotypical or canonical properties, producing semantically tighter clusters with less diversity in property types.

Discussion

The present study applied clustering and switching analyses to the property listing task in Alzheimer's disease. Individuals with AD generated fewer properties overall, exhibited a reduced rate of switching between temporal clusters, and showed longer pauses when making such transitions. However, cluster size and within-cluster production timing did not differ between groups. These results suggest that AD affects strategic exploration across semantic space while local retrieval processes remain relatively intact.

The reduced switch rate in AD participants should be interpreted alongside the preserved cluster size and reduced list length, as these metrics are mathematically interrelated. Even after controlling for total task time, individuals with AD made fewer transitions between temporal clusters per property produced. This pattern aligns with previous semantic fluency findings showing reduced switching in AD (Bertola et al., 2014; Fagundo et al., 2008; Weakley & Schmitter-Edgecombe, 2014), suggesting that strategic exploration deficits generalize across different types of semantic retrieval tasks. Within the optimal foraging framework (Hills et al., 2015), this pattern suggests impaired exploration, consistent with the exploitation bias documented in AD and aging populations (Wyatt et al., 2024). The combination of reduced switch rate and preserved cluster size indicates that AD participants generate roughly the same number of properties per cluster as controls, but produce fewer clusters overall. Although mean switch times were longer in AD participants, the silence between clusters (switch delay) did not differ between groups. This suggests slower speech production at cluster boundaries in AD rather than prolonged search for new properties, though the slope difference method's definition of cluster boundaries may also have contributed to this pattern.

One methodological consideration is that the slope difference method defines cluster boundaries based on temporal dynamics. Properties within a cluster are, by definition, produced at shorter intervals than those across boundaries. The lack of group differences in within-cluster IRT may therefore partially reflect the algorithm rather than purely cognitive processes. Despite this, AD participants showed longer mean switch times, indicating genuine difficulty navigating between temporal clusters.

The semantic similarity analyses revealed an unexpected pattern. AD participants showed significantly reduced variability in within-cluster similarity, suggesting more homogeneous semantic neighborhoods. This may reflect reliance on more stereotyped or canonical properties, with less flexible navigation within temporal clusters. This interpretation is consistent with meta-analytic findings that individuals with MCI and AD generate words of higher frequency and later age-of-acquisition (De Marco et al., 2025), suggesting reliance on more accessible, canonical semantic content. Whether this reduced variability reflects constraints on the types of properties accessed or more rigid retrieval strategies remains to be determined through future analyses that code properties by type (e.g., perceptual, functional, taxonomic).

AD participants spent less time on the task than controls despite it being non-timed, possibly reflecting reduced engagement or earlier exhaustion of accessible properties. All analyses controlled for total task time, so group differences cannot be attributed simply to AD participants spending less time. However, the time difference itself may be meaningful. AD participants likely finish earlier because they more rapidly exhaust accessible temporal clusters due to reduced switching ability, which would be consistent with an exploration deficit (Bertola et al., 2014; Fagundo et al., 2008; Weakley & Schmitter-Edgecombe, 2014).

Several limitations should be noted. First, the non-timed nature of the PLT introduces variability in task engagement that may differ between clinical groups. Future studies could examine performance under timed conditions and test other clustering approaches, including similarity-based methods such as the similarity-drop model (Hills et al., 2012; Kumar et al., 2025). Second, we did not assess the semantic types of properties generated, which might differ between groups and contribute to the reduced within-cluster variability in AD. Third, future work should characterize concept-level variability in clustering and switching, and verify the robustness of similarity-based findings with alternative embedding models. Finally, our AD sample showed considerable clinical heterogeneity (MoCA SD = 6.00), and the sample size limited within-group analyses of severity. Future work with larger samples and more detailed characterization should examine whether clustering and switching metrics scale with disease severity.

A further limitation concerns the random-effects structure of the GLMMs. Random-intercept-only models can inflate Type I error when effects vary across grouping units (Barr et al., 2013). Maximal random-effects models replicated production and semantic similarity findings, but switching effects attenuated to non-significance, likely reflecting concept-level heterogeneity in the AD–CN contrast and the unbalanced design (33% of AD vs. 14% of CN participants completed fewer than 10 concepts). Switching findings should therefore be interpreted with caution. Future work should select random-effects complexity using a principled approach that balances Type I error and statistical power (Matuschek et al., 2017).

The present study showed that clustering and switching analyses can be productively applied to property listing tasks in clinical populations. Individuals with Alzheimer's disease showed reduced strategic exploration across temporal clusters while maintaining relatively intact local retrieval processes. Nonetheless, their within-cluster variability was lower, consistent with difficulties in retrieval strategies. This pattern suggests that the exploitation bias observed in AD generalizes beyond category member retrieval to feature listing, revealing how AD affects strategic navigation of semantic memory across different task structures and types of semantic relations.

Acknowledgments

LLMs were used by the authors to review code for data preprocessing and statistical analysis scripts. The authors take full responsibility for the content of the published article.

References

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

Bertola, L., Cunha Lima, M. L., Romano-Silva, M. A., De Moraes, E. N., Diniz, B. S., & Malloy-Diniz, L. F. (2014). Impaired generation of new subcategories and switching in a semantic verbal fluency test in older adults with mild cognitive impairment. *Frontiers in Aging Neuroscience*, *6*. https://doi.org/10.3389/fnagi.2014.00141

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An Analysis of Sequences of Restricted Associative Responses. *The Journal of General Psychology*, *30*(2), 149–165. https://doi.org/10.1080/00221309.1944.10544467

Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Bushnell, J., Svaldi, D., Ayers, M. R., Gao, S., Unverzagt, F., Gaizo, J. D., Wadley, V. G., Kennedy, R., Goñi, J., & Clark, D. G. (2022). A comparison of techniques for deriving clustering and switching scores from verbal fluency word lists. *Frontiers in Psychology*, *13*, 743557. https://doi.org/10.3389/fpsyg.2022.743557

Canessa, E., & Chaigneau, S. E. (2020). Mathematical regularities of data from the property listing task. *Journal of Mathematical Psychology*, *97*, 102376. https://doi.org/10.1016/j.jmp.2020.102376

Chaigneau, S. E., Canessa, E., Barra, C., & Lagos, R. (2018). The role of variability in the property listing task. *Behavior Research Methods*, *50*(3), 972–988. https://doi.org/10.3758/s13428-017-0920-8

Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024, February 5). *M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation*. arXiv: 2402.03216 [cs]. https://doi.org/10.48550/arXiv.2402.03216

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. *Neuropsychology Review*. https://doi.org/10.1007/s11065-024-09657-z

Fagundo, A. B., López, S., Romero, M., Guarch, J., Marcos, T., & Salamero, M. (2008). Clustering and switching in semantic fluency: Predictors of the development of Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *23*(10), 1007–1013. https://doi.org/10.1002/gps.2025

Hartig, F., Lohse, L., & Leite, M. D. S. (2016). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (Version 0.4.7). https://doi.org/10.32614/CRAN.package.DHARMa

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–440. https://doi.org/10.1037/a0027373

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in Semantic Fields: How We Search Through Memory. *Topics in Cognitive Science*, *7*(3), 513–534. https://doi.org/10.1111/tops.12151

Kumar, A. A., Apsel, M., Zhang, L., Xing, N., & Jones, M. N. (2023). Forager: A Python package and web interface for modeling mental search. *Behavior Research Methods*, *56*, 6332–6348. https://doi.org/10.3758/s13428-023-02296-x

Kumar, A. A., Lundin, N. B., & Jones, M. N. (2025). What's in my cluster? Evaluating automated clustering methods to understand idiosyncratic search behavior in verbal fluency. *Journal of Memory and Language*, *141*, 104606. https://doi.org/10.1016/j.jml.2024.104606

Lagos, R., Chaigneau, S. E., Canessa, E., Toro-Hernández, F., Bontempo, A., Carthery-Goulart, M. T., & Medina Marín, F. A. (2026). You cannot just count: A statistical rethinking of semantic richness in free-generation tasks and semantic norms. *Psychological Methods*. https://doi.org/10.1037/met0000820

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., O'Donnell, B. F., & Todd, P. M. (2023). Neural evidence of switch processes during semantic and phonetic foraging in human memory. *Proceedings of the National Academy of Sciences*, *120*(42), e2312462120. https://doi.org/10.1073/pnas.2312462120

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939–939. https://doi.org/10.1212/WNL.34.7.939

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BF03192726

Montez, P., Thompson, G., & Kello, C. T. (2015). The Role of Semantic Clustering in Optimal Memory Foraging. *Cogni-*

tive Science, *39*(8), 1925–1939. https://doi.org/10.1111/cogs.12249

Muraki, E. J., Sidhu, D. M., & Pexman, P. M. (2020). Mapping semantic space: Property norms and semantic richness. *Cognitive Processing*, *21*(4), 637–649. https://doi.org/10.1007/s10339-019-00933-y

Okruszek, Ł., Rutkowska, A., & Wilińska, J. (2013). Clustering and Switching Strategies During the Semantic Fluency Task in Men with Frontal Lobe Lesions and in Men with Schizophrenia. *Psychology of Language and Communication*, *17*(1), 93–100. https://doi.org/10.2478/plc-2013-0006

Saranpää, A. M., Kivisaari, S. L., Salmelin, R., & Krumm, S. (2022). Moving in Semantic Space in Prodromal and Very Early Alzheimer's Disease: An Item-Level Characterization of the Semantic Fluency Task. *Frontiers in Psychology*, *13*, 777656. https://doi.org/10.3389/fpsyg.2022.777656

Toro-Hernández, F. D., Migeot, J., Marchant, N., Olivares, D., Ferrante, F., González-Gómez, R., González Campo, C., Fittipaldi, S., Rojas-Costa, G. M., Moguilner, S., Slachevsky, A., Chaná Cuevas, P., Ibáñez, A., Chaigneau, S., & García, A. M. (2024). Neurocognitive correlates of semantic memory navigation in Parkinson's disease. *npj Parkinson's Disease*, *10*(1), 15. https://doi.org/10.1038/s41531-024-00630-4

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*(6), 499–504. https://doi.org/10.1016/S0028-3932(97)00152-8

Weakley, A., & Schmitter-Edgecombe, M. (2014). Analysis of Verbal Fluency Ability in Alzheimer's Disease: The Role of Clustering, Switching and Semantic Proximities. *Archives of Clinical Neuropsychology*, *29*(3), 256–268. https://doi.org/10.1093/arclin/acu010

Wyatt, L. E., Hewan, P. A., Hogeveen, J., Spreng, R. N., & Turner, G. R. (2024). Exploration versus exploitation decisions in the human brain: A systematic review of functional neuroimaging and neuropsychological studies. *Neuropsychologia*, *192*, 108740. https://doi.org/10.1016/j.neuropsychologia.2023.108740

Zemla, J. C., & Austerweil, J. L. (2019). Analyzing Knowledge Retrieval Impairments Associated with Alzheimer's Disease Using Network Analyses. *Complexity*, *2019*, 4203158. https://doi.org/10.1155/2019/4203158