# Supplementary Material

## Methods

*Mean Temporal Cluster Size and Number of Temporal Switches.*

Comparable to our previous work (Tröger et al., 2019), we computed clusters and switches temporally. A cluster consisted of a minimum of two words belonging to the same cluster as defined by a temporal threshold. In order to determine *temporal clusters*, each word $w\_{i}$ was assigned a start time $s\_{i}$ and an end time $e\_{i}$ according to its position in the speech recording. Clusters were then determined iteratively. The first word $w\_{0}$ started a new cluster. The next word $w\_{j}$ was part of the cluster, if the distance between its start time $s\_{j}$ and the previous words’ end time $e\_{j-1}$ was below a threshold $t\_{j}$ (i.e., $s\_{j}-e\_{j-1}<t\_{j}$). A base threshold $t$ was determined for a speaker as the mean distance between any consecutive words produced by the speaker. To account for a decrease in word production towards the end of the task, this base threshold was linearly scaled by a maximum factor of two, based on the start of the current word $w\_{j}$ (i.e., $t\_{j}=t\*(\frac{s\_{j}}{60}+1)$). This approach automatically accounts for inter-personal differences in terms of production speed and reaction time. The *mean cluster size* was calculated as the sum of cluster sizes divided by the number of clusters. Finally, the *number of switching clusters* was defined as the total number of switches between clusters, including single word clusters.

## Statistical Analysis

For the analysis of main effects and interactions, we used analysis of variance. For the analysis of repeated assessment effects, we used two planned contrasts (t1, t2, t3 [1, -1, 0] and t1, t2, t3 [0, 1, -1]). Statistical significance levels were set to p < 0.05 and we corrected for multiple testing with the Bonferroni-Holm procedure.

## Results

We found a significant main effect of diagnosis for both mean cluster size [*F* (1,56) = 8.26, *p* < .01] and number of switches [*F* (1,56) = 8.92, *p* < .01]. Patients with MCI showed significantly smaller clusters and less frequent switches than healthy volunteers (Figure S1; Table S1).

For the number of switches, we additionally found a statistical trend for an interaction between diagnosis\*assessment [*F* (2, 112) = 2.62, *p* = .07]. With repeated assessment, healthy volunteers showed more frequent switches, while patients with MCI did not (Figure S1; Table S1).

## Discussion

The results of our additional analysis support our main findings as they also indicate a structural and a procedural deficit. When doing the task for the first time, patients with aMCI produce smaller clusters. Clustering is supposed to reflect semantic retrieval processes and therefore, smaller clusters point to a structural deficit. When doing the task repeatedly, only healthy volunteers were able to change their switching behaviour. Switching is supposed to reflect strategic search processes and thus, a change in switching behaviour points to an effective adaptation of executive control processes.

## References

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation vs. exploration—computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. Neuropsychologia, 131, 53-61.

## Figures

**Figure S1.** Mean cluster size (left) and number of switches (right) with repeated assessment of semantic verbal fluency in healthy volunteers (HC) and patients with amnestic Mild Cognitive Impairment (aMCI). Error bars represent the standard error of the mean.

