

<https://doi.org/10.1038/s41514-025-00267-4>

Deep learning of conversation-based ‘filmstrips’ for robust Alzheimer’s disease detection



Arthur Trognon^{1,2}✉, Coralie Duman^{1,2}, Gwladys Vittart^{1,3}, Natacha Stortini^{1,2}, Loann Mahdar-Recorbet¹ & Hamza Altakrouy¹

Early detection of Alzheimer’s disease remains complex and costly despite advancements in neurobiological markers. We propose an innovative approach based on the topological and kinetic analysis of verbal exchanges to distinguish patients from healthy individuals. Without requiring full transcription, we leverage a convolutional network capable of identifying discursive patterns indicative of cognitive impairments. Our experiments, conducted with 80 participants, demonstrate performance levels exceeding 95% in cross-validation, comparable to computational approaches relying on biological markers. This robust and minimally invasive methodology could be easily integrated into clinical protocols, enhancing current diagnostics. It also holds the promise of cost-effectively extending monitoring to other neurodegenerative or psychiatric diseases.

As of today, Alzheimer’s disease (AD) remains a substantial public health and social challenge, given its high incidence of 1–3% and a prevalence of 10–30% in the population aged over 65 years^{1–4}.

Disease detection generally relies on the identification of abnormal protein deposits (tau, β -amyloid), biomarkers in cerebrospinal fluid, or brain imaging (amyloid aggregation, cortical atrophy), explaining why most contemporary algorithms (supervised or unsupervised) are based on these signals for early screening or tracking disease progression^{5–12}.

However, despite their technical qualities and precision, these methods appear to face major limitations that can be categorized into two groups: (a) challenges related to the quality and reproducibility of the data, as well as the identification of consensual Regions-Of-Interest, thereby limiting performance¹³; and (b) limited accessibility, as many clinics may lack the necessary infrastructure to produce comparable data, restricting the broader application of these models. Faced with these technical and logistical constraints, exploring simple, non-invasive, or ecological markers could represent a relevant avenue to complement existing methods.

Typically associated with memory and cognitive declines, some studies have already successfully focused on training detection models using neuropsychological parameters, thus distinguishing acquired pathology from milder forms (i.e., mild cognitive impairment)^{14–16}. Nevertheless, more ecological and easily accessible assessment approaches could readily be imagined. For instance, the disease also appears to profoundly impair patients’ ability to engage in everyday conversations: individuals with AD

may produce excessive speech, repeat themselves, or even lose the thread of conversations^{17,18}, as evidenced by recent studies identifying connected speech as a valuable psychomarker for tracking the progression of Alzheimer’s disease^{19,20}. Such impairments likely reflect disrupted synaptic homeostasis within key functional networks, such as the default mode network (DMN)²¹, leading to cascading dysfunction from posterior cortical regions to medial temporal and prefrontal areas implicated in language, behavior and memory^{22–24}.

Following these observations, we propose a novel methodological framework focused on the topological and kinetic architecture of conversations. Its originality lies in the fact that it does not rely on transcribed textual data, representing a qualitative leap in terms of practicality and ease of implementation in routine clinical settings—since transcription and annotation of corpora often constitute a burden for traditional Natural Language Processing methods.

This approach was applied to data from 40 individuals diagnosed with AD and 40 matched healthy controls, collected by a team from the University of Montpellier (France)²⁵. In this study, Participants were required to complete two autobiographical recall tasks in a dialogical setting: one task focused on recounting their most cherished memory, while the other explored their most unpleasant one. The exercise was conducted in dyads with a linguist, who provided thematic prompts to mimic a natural interaction akin to a clinical consultation, thus constituting a low-cost and easy-to-implement setup.

¹CLINICOG, Nancy, France. ²Faculty of Arts and Humanities Campus, Lorraine University, Nancy, France. ³Pont de Bois Campus, Lille University, Villeneuve d’Ascq, France. ✉e-mail: arthur.trognon@clinicog.fr

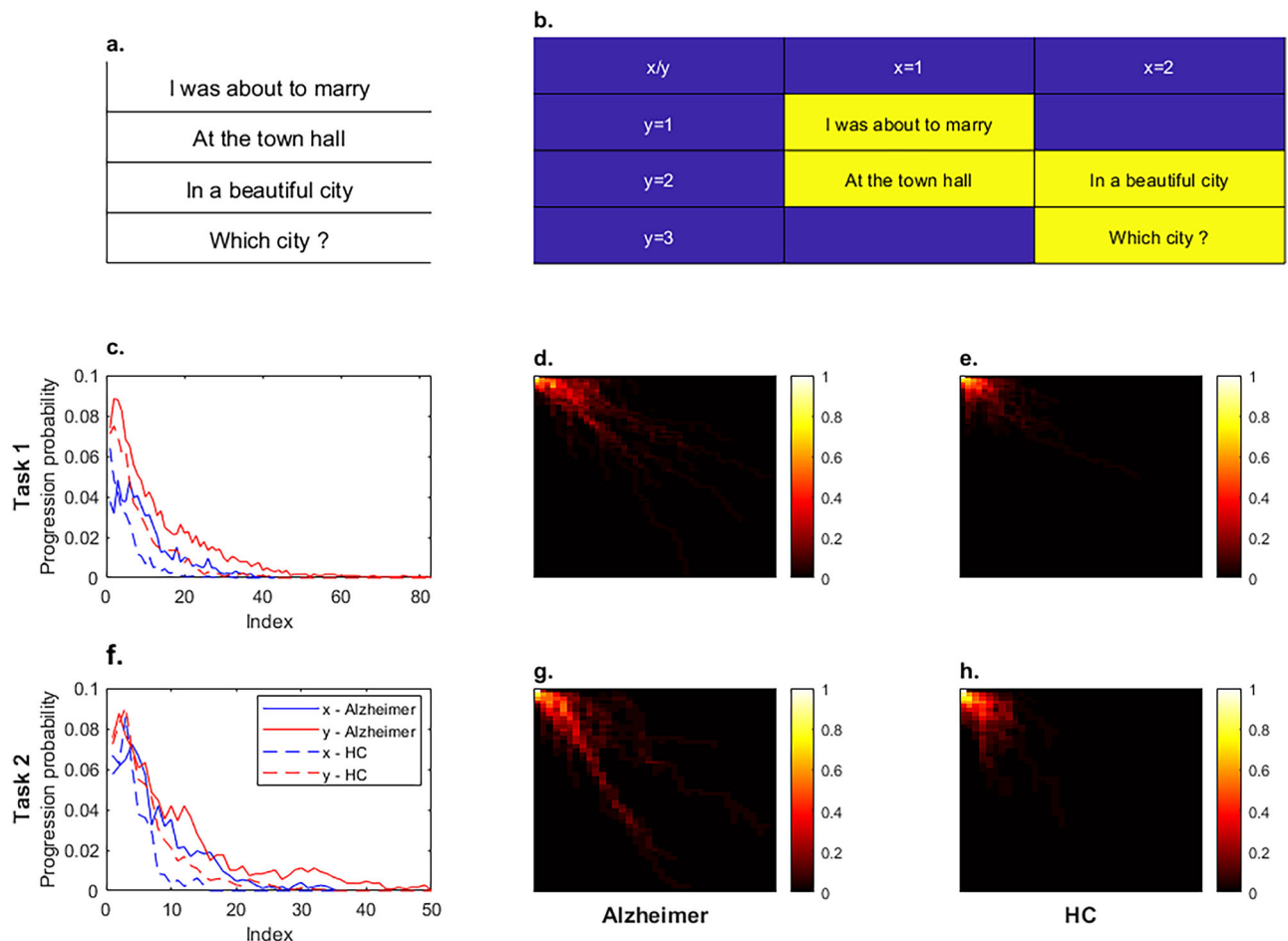


Fig. 1 | Topological mapping of conversation and progression probabilities. **a, b** Example of statements (i.e., speech acts) mapped onto the topological space (x, y): incrementing along x represents the opening or reiteration of a theme, while incrementing along y indicates the elaboration of a sub-theme. **c, f** Progression probabilities along the x -axis (blue) and y -axis (red) for both recall tasks (Task 1:

most pleasant memory; Task 2: least one). Patients with Alzheimer's disease (solid) display slower decay rates and more dispersed peaks in transition probabilities compared to controls (dashed). **d, e, g, h** Visualization of aggregated (x_{\max}, y_{\max}) topological matrices: in Alzheimer's patients, a more diffuse distribution in topological positions is observed compared to controls.

Encoding was achieved using the Trognon Topological and Kinetic Model of dialogue (2TK)^{26–28}. This framework operates through a modular and parametric function, the interaction function, denoted as $\Psi(\xi, \varepsilon, \sigma_n)$, where $\xi = (x, y)$ represents the topological position of a speech act, ε represents the incremental temporal marker, and σ_n (not used in this study) refers to a set of parameters describing metrics such as the type of speech act or alignment measures between interlocutors.

In this specific model, speech acts are mapped onto a hierarchical representation where conversational topics progress along the x -axis, representing the emergence of a theme or its elaboration into sub-discursive themes (e.g. « I was about to marry » [$x = 1$] : « At the town hall » [$x = 1$] + « In a beautiful city » [$x = 2$]; Fig. 1a, b). Meanwhile, the elaboration of specific information is represented by an increment along the y -axis (e.g. « In a beautiful city » [$y = 2$] -> « Which city? » [$y = 3$]; Fig. 1a, b). Thus, this formalism captures conversational dynamics by focusing on thematic development rather than detailed annotated transcription, offering a less resource-intensive alternative to classical NLP methods.

After encoding all corpora topologically, we aggregated all the resulting structures into a (x_{\max}, y_{\max}) matrix. Visual observation suggested that AD patients seemed to exhibit unique discursive patterns compared to control subjects, characterized by localized transitions (“stepwise”) along the x and y axes (Fig. 1d, e, g, h).

To validate this impression, we computed the transition probabilities from each position (x, y) to neighboring positions ($x + 1, y$) or ($x, y + 1$).

For example, for the x -axis:

$$P(x \rightarrow x + 1) = \frac{1}{N_y} \sum_{y=0}^{N_y-1} P(x, y) \cdot P(x + 1, y) \quad (1)$$

where N_y is the number of rows (i.e. values of y) in the matrix, $P(x, y)$ is the probability at position (x, y) and $P(x + 1, y)$ is the probability at position ($x + 1, y$).

As shown in Fig. 1c, f, the controls (dashed lines) display a rapid decline in progression probabilities as the index rises. Conversely, AD patients (solid lines) exhibit two notable differences: (a) the decrease is less pronounced, and (b) peaks in progression probabilities are observed both at the beginning of the task (low indices, Fig. 1c) and later in the conversation (medium indices, Fig. 1f). These observations thus suggest a strong tendency for digression in AD patients, as frequently reported by previous studies^{17,18}.

Based on these results, we aimed to develop a simple and reproducible approach to encode transition probabilities for training a decision-making algorithm. Based on the topological structure of the exchange (Fig. 2a), we generated a visual representation of the (x_{\max}, y_{\max}) matrices, where a colored box marks the topological position of the exchange at a specific time point ε (Fig. 2b). Then, for each patient, we sequentially concatenated all the unit images, thereby creating a “film strip” that reflects both the topological architecture and the kinetics of the exchange (Fig. 2c).

Fig. 2 | Spatiotemporal encoding procedure and classification performance. **a** Schematic depiction of four speech acts in the (x, y) matrix. **b** At each temporal increment ϵ , the cell (highlighted in yellow) corresponding to the topological position at that moment is selected, generating a sequence of “snapshots”. **c** The snapshots are then concatenated to form a “filmstrip” (horizontal axis) that simultaneously encodes topological structure (x, y) and kinetic progression ϵ . **d** Results (percentage accuracy) comparing the “Experimental” condition (real dataset: AD vs. HC) and the “Control” condition (artificially mixed groups). **e** Boxplots illustrating the model’s sensitivity (blue) and specificity (orange) for the same comparison. The higher scores for the experimental condition confirm the algorithm’s ability to effectively distinguish AD patients.

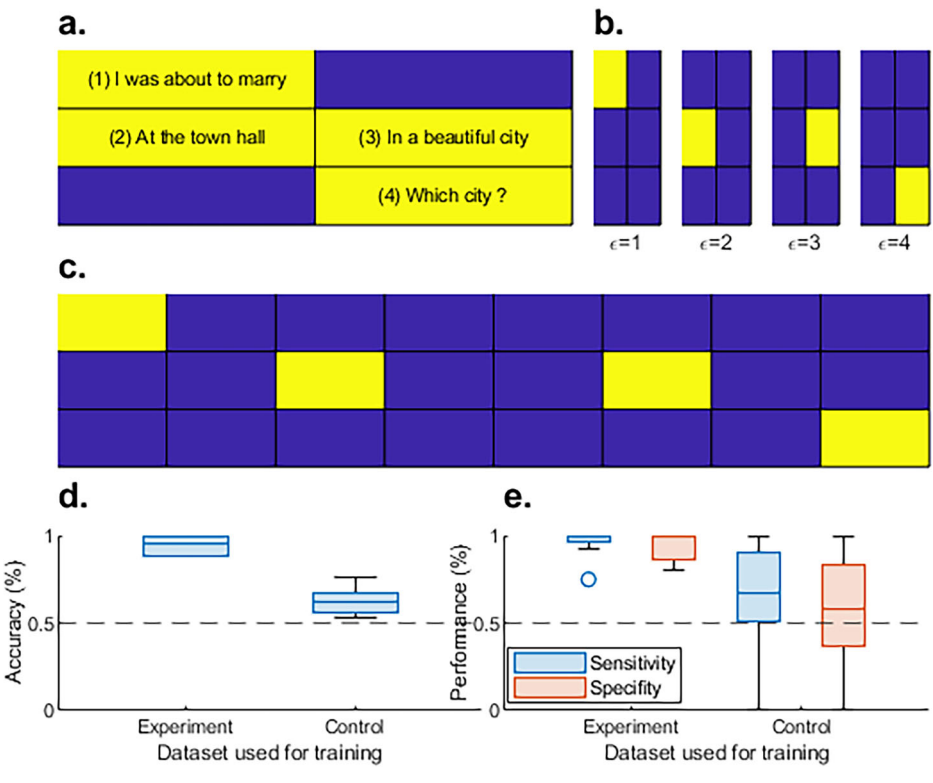


Table 1 | Example corpus for the segmentation of speech turns into speech acts

Locutor	Enunciation	ϵ
E1	(...) would you mind sharing your most cherished memory?	1
P2	Um, I'm not sure... [pause] I feel a bit caught off guard.	2

Table 2 | Example segmentation of a speech turn into speech acts

Locutor	Enunciation	ϵ
P2a	Um, I'm not sure...	2
P2b	I feel a bit caught off guard.	3

Table 3 | Example corpus for the topological encoding of speech acts

Locutor	Enunciation	ϵ	x	y
E1	(...) could you tell me about your best memory?	1	1	1
P2a	I won't say our wedding	2	1	2
P2b	I'd rather say, uh, the liberation	3	2	2
E3	Liberation from?	4	2	3
P4	Well, the war.	5	2	4

To capitalize on the spatial and temporal dynamics encoded in these filmstrips we trained a convolutional neural network (CNN) within the Teachable Machine 2.0 environment, a web-based platform with no coding required²⁹. Designed initially for visual tasks, CNNs apply sliding convolutional filters to automatically detect and integrate local features, forming a representation optimized for classification tasks³⁰. In our case, the

Table 4 | Example of topological encoding within the 2TK model

	$x = 1$	$x = 2$
$y = 1$	E1(1)	
$y = 2$	P2a(2)	P2b(3)
$y = 3$		E3(4)
$y = 4$		P4(5)

convolutional architecture identifies the spatial and temporal regularities of the horizontally concatenated (x, y, ϵ) matrix, thus revealing thematic progression patterns specific to AD patients. Through this process, we achieve a decision-making algorithm capable of accurately distinguishing conversational profiles between healthy individuals and people with Alzheimer’s disease.

Aligned with practices in computational psychometrics^{31,32}, we implemented a two-step procedure to assess the model’s robustness. On the one hand, an “Experiment” condition where each filmstrip was associated with its true class (AD vs. HC); on the other hand, a “Control” condition where we created two artificial groups (A and B), each containing a mix of 50% AD and 50% HC. This setup follows the principle of a “positive control” (capable of revealing an effect) and a “negative control” (excluding the possibility of genuinely distinguishing the two groups) to ensure that the observed classification does not result from random events or background noise (Tables 1–6).

We performed an analysis of variance on detection metrics derived from 8-fold cross-validation to evaluate significant differences between the two experiments. Regarding accuracy, after a Shapiro test confirmed the non-normal distribution of the data [$p=0.037$], we conducted a Kruskal–Wallis test (Experiment x Accuracy), which confirmed the difference in measured accuracy between the two conditions (Experiment : $\mu_{acc} = 0.95$, SD = 0.06; Control : $\mu_{acc} = 0.62$, SD = 0.09; [$\chi^2_{(1)} = 15$, $p < 0.001$]; Fig. 2d). Similarly, the Scheirer–Ray–Hare test applied to specific metrics

Table 5 | Example corpus for resolving topological ambiguity and proveability matrix of the exchange

ϵ	Speech Act	I have a car	It is red	It is a Ford	To be more precise	It is magenta	It is a Ford Fiesta
1	I have a car	x	1	1	1	1	1
2	It is red	1	x	0	1	1	0
3	It is a Ford	1	0	x	1	0	1
4	To be more precise	0	1	1	x	1	1
5	It is magenta	1	1	0	1	x	0
6	It is a Ford Fiesta	1	0	1	1	0	x

Table 6 | Topological grid of the exchange

	$x = 1$	$x = 2$
$y = 1$	I have a car	
$y = 2$	It is red	It is a Ford
$y = 3$	To be more precise	It is a Ford Fiesta
$y = 4$	It is magenta	–

(Experiment \times Metric \times Value) demonstrated a significant difference in performance between the two experiments (Experiment : $\mu_{sen} = 0.96$, SD = 0.009, $\mu_{spe} = 0.94$, SD = 0.009; Control : $\mu_{sen} = 0.64$, SD = 0.32, $\mu_{spe} = 0.57$, SD = 0.33; [$H_{(1)} = 11.64$, $p < 0.001$]; Fig. 2e). These results suggest that our analytical system can produce computational decisions on dialogically encoded topological data with high performance in screening clinical conditions such as Alzheimer's disease.

To determine whether the topological-kinetic alterations captured by our CNN map onto usual clinical markers of cognitive decline, we computed each (x, y, ϵ) trajectory into eight complementary human-readable metrics and correlated them with the Mini Mental-State Examination (MMSE) score, a test widely used in clinical routine. Among these metrics, the *lateral-to-vertical ratio* $R_{LV} = \frac{N_{(x \rightarrow x+1, y)}}{N_{(x, y \rightarrow y+1)}}$ (2) quantifies the prevalence of thematic jumps relative to thematic elaborations. The *transition entropy* $H = -\sum_{i,j} p_{i,j} \log p_{i,j}$ (3) captures the unpredictability of successive moves across the grid, while the *back-tracking index* $B = \frac{N_{(x_{t+1} < x_t)}}{E}$ (4) isolates explicit returns to abandoned themes. A linear slope fitted to the $x_{unique}(\epsilon)$ (cumulative count of over time) provides a novelty gradient, and the *mean thematic dwell time* $\bar{\tau}_x = \frac{\sum_{\epsilon=1}^E \mathbb{1}_{[x_{\epsilon} = x, x_{\epsilon+1} = x]}}{N_x}$ (5) records how long a topic is sustained. In parallel, the *topological dispersion* $\bar{d} = \frac{1}{E-1} \sum_{\epsilon} \sqrt{(\Delta_{x_{\epsilon}})^2 + (\Delta_{y_{\epsilon}})^2}$ (6) estimates the average spatial amplitude of successive steps, whereas the lag-1 autocorrelation $\rho_x(1)$ and the vertical burstiness coefficient $B_y = \frac{\sigma_{\Delta_y} - \mu_{\Delta_y}}{\sigma_{\Delta_y} + \mu_{\Delta_y}}$ (7) index, respectively, short-range thematic persistence and intermittent bursts of detail along the y -axis.

All eight markers were significantly associated with the MMSE score ($|r| \in [0.18, 0.36]$, $p_x < 0.001$). The strongest signals came from the back-tracking index ($r = 0.36$) and topological dispersion ($r = 0.27$), consistent with the clinical picture of patients drifting across themes, revisiting abandoned ones, and describing an increasingly scattered conversational path as cognition wanes. Even subtler metrics (such as dwell time or burstiness) retained weaker but reliable links (with $r \geq 0.24$), suggesting that deterioration permeates many facets of discourse organization rather than a single dominant feature.

In conclusion, our work demonstrates that a topological and kinetic encoding of dialogical data, combined with a convolutional architecture, can accurately distinguish individuals with Alzheimer's

disease from healthy subjects. Moreover, given that our system only exploits the spatial and temporal progression of the conversation, it eliminates the need for complete textual transcription and thus reduces the technical complexity often associated with traditional NLP approaches. Statistical analyses conducted across both experimental and control conditions confirmed the robustness of our model, with high accuracy, sensitivity, and specificity scores. This performance is particularly notable as it relies on a simple, cost-effective, and easily reproducible protocol—a short, guided interview followed by minimal topological encoding—indicating its potential applicability in various contexts, including teleconsultations or hospital settings.

Moreover, the strong distinction observed in the topological-kinetic profiles between the two groups highlights the interest in further exploring discursive dynamics as a potential psychomarker. In the long term, our approach could be integrated into low-cost early screening or longitudinal monitoring systems, complementing current imaging or biomarker-based methods. Additionally, this approach could be tested on other categories of patients with cognitive or psychiatric disorders, with the aim of developing a panel of characteristic conversational signatures.

Methods

University of Montpellier dataset

A total of 80 native French speakers were included in this archival dataset, consisting of 40 individuals clinically diagnosed with Alzheimer's disease (AD) and 40 healthy controls. All participants were recruited in the Montpellier region from healthcare facilities (for the AD group) or community organizations (for controls). Clinical diagnoses of AD conformed to the standard NINCDS-ADRDA criteria³³, targeting the most frequent amnesic hippocampal form of the disease. Control participants were selected to match the AD group on age, sex distribution and sociocultural level, ensuring comparable demographic profiles (age range: 64–89 years for AD versus 65–85 years for controls, p : n.s.; sex ratio: AD group, 28 females and 12 males versus controls, 20 females and 20 males, p : n.s.; sociocultural level: mean \pm SD, 2.5 ± 1.01 for AD group versus 2.9 ± 1.14 for controls, range 1–4 for both groups, p : n.s.). Each participant was screened for significant neurological or psychiatric history apart from AD. The Mini Mental State Examination (MMSE) was used to confirm cognitive status (mean MMSE score: 21.53 ± 2.81 for AD versus 30 ± 0 for controls, $p < 0.001$). Most patients (34 of 40) were at a mild disease stage (MMSE between 20 and 25). Informed consent was systematically collected from participants or their legal representatives when required, and the study was deemed observational by the relevant ethics committee (Comité de Protection des Personnes de Montpellier Sud Méditerranée II), negating further regulatory filing. All interviews were carried out in a quiet setting (hospital unit, care facility or community center) and digitally recorded (44 kHz, 16-bit, mono). Participants were asked to describe salient life events, with a focus on (a) their most positive and (b) most unpleasant autobiographical memories. Interviewers used thematic prompts to encourage detailed narration. The entire set of audio files was subsequently transcribed using CHAT convention in the CLAN software³⁴, yielding a corpus of 54,454 total word tokens. No significant difference in corpus size was detected between AD and control groups ($p = 0.155$).

Assessor background and preparation

The original interviews were conducted by a doctoral researcher holding a Master's degree in Linguistics, and subsequent 2TK annotation was carried out by two first-year clinical-psychology master's students who had received a single three-hour lecture on the 2TK framework during the preceding semester. A final examination was carried out by an experienced 2TK analyst to ensure annotation fidelity.

Data preprocessing

Segmentation of speech turns into speech acts

Within the 2TK framework, conversational analysis involves breaking down speech turns into minimal units of information (i.e. speech acts). This means that each speech act is considered as a small, semantically coherent block, distinct from a purely syntactic sentence. For instance, consider the following dialogue between the interviewer (E) and the participant (P; Table 1):

In this case, the participant's turn (P2) can be segmented into two separate speech acts (Table 2):

Where:

- P2a corresponds to an initial act reflecting the participant's hesitation and uncertainty (« Um, I'm not sure... »).
- P2b constitutes a separate act emphasizing a sense of surprise or lack of readiness ("I feel a bit caught off guard").

Topological encoding

In the 2TK model, each speech act κ is assigned an address ξ , defining its position within the exchange. Expressed as $\xi = (x, y)$, this address captures two key dimensions:

- x : the axis of thematic progression, which indicates when a new theme/sub-theme is introduced, or a transversal link to an already-discussed theme is made.
- y : the informational depth axis, marking instances where clarifications, details, or examples are requested for an ongoing topic.

In parallel, the ε index provides a temporal marker, organizing speech acts in their chronological order of occurrence.

Let us consider the following dialogue between the interviewer (E) and the participant (P; Table 3):

In this example:

- E1 introduces the request to mention "the best memory," initiating a new theme; it is thus associated with $(x = 1, y = 1)$.
- P2a continues the same theme ($x = 1$) by mentioning a first idea ("I won't say our wedding"), hence the incrementing of y to $(y = 2)$ to signify further elaboration.
- P2b introduces a new perspective ("the liberation"), at the same degree of elaboration: so x shifts from 1 to 2, while keeping y at $(y = 2)$.
- E3 seeks clarification ("Liberation from?"), refining the current theme ($x = 2$) and increasing the degree of elaboration on this topic, advancing y to $(y = 3)$.
- P4 responds ("Well, the war"), adding another layer of depth ($y = 4$) within the same theme ($x = 2$).

Matrix conversion

Once each speech act is positioned in the topological space via its address $\xi = (x, y)$, exchange is converted into a matrix. Specifically, a two-dimensional matrix is constructed, where the columns correspond to the different values of x (i.e. thematic progression) and the rows correspond to the different values of y (i.e. elaboration).

In each cell, the speaker's identifier is stored, followed by the temporal position (ε) of the speech act. For instance, in the precedent exchange where x et y respectively take the values $\{1, 2\}$ and $\{1, 2, 3, 4\}$, the resulting matrix could appear as follows (Table 4):

In this example:

- The cell $(x = 1, y = 1)$ contains «E1(1)», which signifies that the speech act performed at $(\varepsilon = 1)$ by investigator E is mapped to the

address $\xi = (1, 1)$.

- The cell $(x = 2, y = 2)$ corresponds to "P2b(3)", representing the speech act carried out at $(\varepsilon = 3)$ by participant P.
- The absence of explicit speech acts in empty cells is particularly significant. These voids highlight thematic coherence: if $(x = 2, y = 2)$ is populated while $(x = 2, y = 1)$ remains empty, this signals that sub-theme ($x = 2$) ties back to a thread initiated at $(x = 1, y = 1)$ rather than indicating a radical thematic shift requiring a speech act at $(x = 2, y = 1)$.

In this sense, this annotated topological matrix allows for a structural mapping of the conversation: rather than retaining the entire verbatim transcript, it captures only the thematic anchoring (x) and the degree of elaboration (y) for each speech act, as well as the chronological order (ε). This abstraction makes it possible (i) to dispense with a literal transcription of the discourse and (ii) to facilitate automated analysis, whether for calculating transition probabilities or constructing image representations for a convolutional neural network.

Resolving topological ambiguity

The 2TK framework typically resolves ambiguity by appealing to a purely topological criterion, independent of any discursive annotation. Concretely, the first act that substantiates its super-ordinate element is placed at $(x, y + 1)$ with respect to that anchor; the second corroborative act is mapped to $(x + 1, y + 1)$, the third to $(x + 2, y + 1)$, and so forth. In other words, the proof chain fans out laterally while remaining on the same depth line, guaranteeing a deterministic placement even under strong pragmatic ambiguity.

In a minority of exchanges, two successive speech acts may simultaneously refine a previously established theme and open a collateral sub-theme, thereby blurring the customary "vertical-versus-horizontal" decision rule. When uncertainty persists, we build a proveability matrix (i.e. a symmetric table whose cells encode the presence (1) or absence (0) of a justificatory link between every pair of speech acts). The example below (Table 5) illustrates the procedure on a six-turn micro-dialogue taken from ref. 28:

Applying the rule step-by-step:

- $\varepsilon = 1$. Anchor of the discourse $\Rightarrow (x = 1, y = 1)$.
- $\varepsilon = 2$. Single high link with $\varepsilon = 1$ ("colour" qualifies the car) \Rightarrow vertical elaboration $\rightarrow (x = 1, y = 2)$.
- $\varepsilon = 3$. High affinity with $\varepsilon = 1$ but none with $\varepsilon = 2$; introduces a brand attribute \Rightarrow lateral shift at same depth $\rightarrow (x = 2, y = 2)$.
- $\varepsilon = 4$. Ties equally to colour and brand branches; meta-comment requesting further precision \Rightarrow transversal elaboration anchored to the brand clarified next (colour) $\rightarrow (x = 1, y = 3)$.
- $\varepsilon = 5$. Precision on colour only \Rightarrow remain in colour branch, increase depth $\rightarrow (x = 1, y = 4)$.
- $\varepsilon = 6$. Precision on brand branch \Rightarrow remain in brand branch, increase depth $\rightarrow (x = 2, y = 3)$.

The resulting grid is therefore (Table 6):

Filmstrip generation. Filmstrip creation was performed using Python (v3.12.4) to translate each participant's topological matrix into a sequence of images that jointly encode temporal (ε) and spatial (x, y) progression. First, the maximum values of x and y (x_{\max}, y_{\max}) were determined from the Excel file storing the topological matrices. Next, for each participant and each non-empty cell (x, y) in the matrix, the temporal index ε was extracted, and a single-frame matrix of size (x_{\max}, y_{\max}) was generated. All entries were initialized to zero, except for the cell at (x, y) , which was set to one, indicating that at time ε , the conversation occupied position (x, y) . The resulting frames—one per speech act—were ordered chronologically according to their ε values. Finally, each single-frame matrix was saved as a PNG image (axes and annotations removed) and

concatenated horizontally to create a single filmstrip that captures the spatiotemporal evolution of the conversation. For each participant, one filmstrip was generated per autobiographical recall task, leading to a total of 143 filmstrips across all participants (4 participants did not complete task 1 and 15 participants did not complete task 2 in the original dataset).

Model and training. Filmstrips were processed through Teachable Machine v2.0, a MobileNet-V2 (width-multiplier = 1.0, input = $224 \times 224 \times 3$) CNN automatically instantiated in the no-code environment proposed by Google to process image data²⁹. All depth-wise separable convolutional blocks are frozen; only the global-average-pooling output (1280 units) feeds a new dense layer (soft-max, 2 classes) that is fine-tuned. Learning is performed for 50 epochs, batch-size = 16, with adam optimiser, and initial learning-rate = 1×10^{-3} ^{35–39}. All 143 filmstrips (two per participant when available) were pooled and labelled AD or HC. For each of eight independent repetitions, Teachable Machine applied its built-in stratified hold-out (85% training, 15% test) at the image level; the test subset was never accessed during training. Because the partition occurs per image, the two autobiographical tasks of a given participant can appear in different splits. This choice was intentional: positive- and negative-memory recalls engage distinct autobiographical networks and exhibit non-redundant neurophysiological pathways^{40,41}.

Computational psychometrics experiment. To evaluate the robustness of our approach, we implemented a procedure inspired by computational psychometrics³¹, based on the positive/negative control principle commonly used in animal experiments⁴². Specifically, we defined two conditions: in the “Experiment” condition, the topological profiles were associated with their true group (AD vs. HC), whereas in the “Control” condition, we artificially created two mixed groups (each containing 50% AD profiles and 50% HC profiles). This second condition aimed to estimate the “background noise,” as no class difference was actually expected. In both conditions, we then applied the same prediction protocol (convolutional network within the Teachable Machine 2.0 environment and 8-fold cross-validation) to calculate sensitivity, specificity, and accuracy. The performance gap between the “Experiment” and “Control” conditions thus provides information on the model’s effective validity, distinguishing true signals from random noise⁴³.

Progression probabilities. To estimate the probability of occurrence of each topological position (x, y) in the conversations, we first grouped the matrices derived from the 2TK encoding for each of the two groups (AD patients and controls) and for each segment of the interview (best memory or most unpleasant). Specifically, for each group, we listed all the matrices produced and then determined the maximum size (in terms of rows and columns) among them. Smaller matrices were then resized into this larger grid by filling empty areas with zeros. Next, we stacked all the adjusted matrices (i.e., now sharing the same dimensions) before calculating the mean occupation at each position (x, y) of the grid. This operation produced a probability map quantifying, for each topological coordinate, the average frequency of a speech act within the group considered. Finally, the aggregated probability maps were saved as a DataFrame to facilitate subsequent statistical analyses and visualization.

Metrics calculation. For each conversation, topological and kinetic metrics were computed using custom Python functions applied to the encoded trajectories stored in DataFrames. Specifically, the lateral-to-vertical ratio was determined by counting lateral movements (increments along the x -axis only) versus vertical movements (increments along the y -axis only), calculating their ratio. Transition entropy was obtained by counting unique pairs of sequential topological transitions and computing the Shannon entropy to quantify unpredictability in conversational structure. The back-tracking index measured how often conversation returned explicitly to previously visited themes by assessing

decreases in x -axis positions. The thematic novelty slope was computed by fitting a linear regression to the cumulative number of unique themes introduced over time (ϵ). Mean thematic dwell time and mean horizontal run length were calculated by identifying continuous sequences with constant x or y values and averaging their durations. Topological dispersion quantified the average spatial distance between consecutive speech acts, measured as the mean Euclidean distance across (x, y) coordinates. Lag-1 autocorrelation assessed short-term thematic persistence along the x -axis, while vertical burstiness captured variability in vertical elaborations on the y -axis, computed as the normalized difference between standard deviation and mean of vertical increments.

Data availability

The raw data used in this study are openly available in the online archives of the University of Montpellier (Lee²⁵). The preprocessed data and the filmstrips can be obtained alongside the present manuscript. The code used in this study is openly available alongside the manuscript on the Code Ocean platform

Received: 29 April 2025; Accepted: 11 August 2025;
Published online: 29 August 2025

References

- Bachman, D. L. et al. Incidence of dementia and probable Alzheimer’s disease in a general population: the Framingham Study. *Neurology* **43**, 515–515 (1993).
- Hebert, L. E. et al. Age-specific incidence of Alzheimer’s disease in a community population. *JAMA* **273**, 1354–1359 (1995).
- Prince, M. et al. *World Alzheimer Report 2015. The Global Impact of Dementia: an analysis of prevalence, incidence, cost and trends* (Alzheimer’s Disease International, 2015).
- Masters, C. L. et al. Alzheimer’s disease. *Nat. Rev. Dis. Prim.* **1**, 15056 (2015).
- Afzal, S. et al. Alzheimer Disease Detection Techniques and Methods: A Review. *Int. J. Interact. Multimed. Artif. Intell.* **6**, 26–38 (2021).
- Marwa, E.-G., Moustafa, H. E.-D., Khalifa, F., Khater, H. & Abdelhalim, E. An MRI-based deep learning approach for accurate detection of Alzheimer’s disease. *Alex. Eng. J.* **63**, 211–221 (2023).
- Goel, T. et al. Multimodal neuroimaging based Alzheimer’s disease diagnosis using evolutionary RVFL classifier. *IEEE J. Biomed. Health Inform.*, **29**, 3833–3841 (2023).
- Bandyopadhyay, A. et al. Alzheimer’s disease detection using ensemble learning and artificial neural networks. In *International Conference on Recent Trends in Image Processing and Pattern Recognition* 12–21 (Springer, 2022).
- Rahim, N. et al. Prediction of Alzheimer’s progression based on multimodal deep-learning-based fusion and visual explainability of time-series data. *Inf. Fusion* **92**, 363–388 (2023).
- Kumar-Singh, S. et al. Mean age-of-onset of familial alzheimer disease caused by presenilin mutations correlates with both increased A β 42 and decreased A β 40. *Hum. Mutat.* **27**, 686–695 (2006).
- Rowe, C. C. et al. Imaging β -amyloid burden in aging and dementia. *Neurology* **68**, 1718–1725 (2007).
- Klunk, W. Imaging brain amyloid in Alzheimer’s disease with Pittsburgh Compound-B. *Ann. Neurol.* **55**, 303–305 (2004).
- Menagadevi, M., Devaraj, S., Madian, N. & Thiyagarajan, D. Machine and deep learning approaches for alzheimer disease detection using magnetic resonance images: an updated review. *Measurement* **226**, 114100 (2024).
- Gurevich, P., Stuke, H., Kastrop, A., Stuke, H. & Hildebrandt, H. Neuropsychological testing and machine learning distinguish Alzheimer’s disease from other causes for cognitive impairment. *Front. Aging Neurosci.* **9**, 114 (2017).
- Garcia-Gutierrez, F. et al. Diagnosis of Alzheimer’s disease and behavioural variant frontotemporal dementia with machine learning-

- aided neuropsychological assessment using feature engineering and genetic algorithms. *Int. J. Geriatr. Psychiatry* **37**, (2022).
16. Battista, P., Salvatore, C., Berlinger, M., Cerasa, A. & Castiglioni, I. Artificial intelligence and neuropsychological measures: The case of Alzheimer's disease. *Neurosci. Biobehav. Rev.* **114**, 211–228 (2020).
 17. Hason, L. Speech Features for Monitoring Alzheimer's Disease Using Random Forest Classifier. (Toronto Metropolitan University, 2022).
 18. Qonita, Z. & Indah, R. N. Semantic and pragmatic impairments of person with Alzheimer's disease. *LUNAR J. Lang. Art.* **7**, 1–17 (2023).
 19. Ahmed, S., Haigh, A.-M. F., de Jager, C. A. & Garrard, P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain* **136**, 3727–3737 (2013).
 20. Malcorra, B. L. C. et al. Low speech connectedness in Alzheimer's disease is associated with poorer semantic memory performance. *J. Alzheimers Dis.* **82**, 905–912 (2021).
 21. Jack, C. R. Jr. et al. Amyloid-first and neurodegeneration-first profiles characterize incident amyloid PET positivity. *Neurology* **81**, 1732–1740 (2013).
 22. Buckner, R. L. et al. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *J. Neurosci.* **29**, 1860–1873 (2009).
 23. Protas, H. D. et al. Posterior cingulate glucose metabolism, hippocampal glucose metabolism, and hippocampal volume in cognitively normal, late-middle-aged persons at 3 levels of genetic risk for Alzheimer disease. *JAMA Neurol.* **70**, 320–325 (2013).
 24. Jones, D. T. et al. Tau, amyloid, and cascading network failure across the Alzheimer's disease spectrum. *Cortex* **97**, 143–159 (2017).
 25. Lee, H. *Langage et maladie d'Alzheimer: analyse multidimensionnelle d'un discours pathologique*, 3 (Montpellier, 2012).
 26. Trognon, A. Computational diagnosis of Shwachman-Diamond syndrome through cognitive and dialogical investigations. Doctoral dissertation, Lorraine University (2022).
 27. Trognon, A. et al. Self-beneficial transactional social dynamics for cooperation in Shwachman-Diamond syndrome: a mixed-subject analysis using computational pragmatics. *Front. Psychol.* **15**, (2025).
 28. Trognon, A., Humeau, C., Mahdar-Recorbet, L., Verhaegen, F. & Musiol, M. A physical framework to harmonize human interaction analysis across disciplines. *Curr. Psychol.* <https://doi.org/10.1007/s12144-025-07297-x> (2025).
 29. Carney, M. et al. Teachable machine: approachable web-based tool for exploring machine learning classification. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* 1–8 (Association for Computing Machinery, 2020).
 30. LeCun, Y. & Bengio, Y. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* **3361**, 1995 (1995).
 31. Trognon, A., Cherifi, Y. I., Habibi, I., Demange, L. & Prudent, C. Using machine-learning strategies to solve psychometric problems. *Sci. Rep.* **12**, 18922 (2022).
 32. Trognon, A. & Richard, M. Questionnaire-based computational screening of adult ADHD. *BMC Psychiatry* **22**, 401 (2022).
 33. McKhann, G. et al. 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**, 939–944 (1984).
 34. MacWHINNEY, B. & SNOW, C. The child language data exchange system: an update. *J. Child Lang.* **17**, 457–472 (1990).
 35. Chhipa, V. & Poonia, R. Teachable machine: a web based machine learning tool for user voice biometric authentication system. *J. Discret. Math. Sci. Cryptogr.* **27**, 1345–1355 (2024).
 36. Malahina, E. A. U., Iriane, G. R., Belutowe, Y. S., Katemba, P. & Asmara, J. A grid-search method approach for hyperparameter evaluation and optimization on teachable machine accuracy: a case study of sample size variation. *J. Appl. Data Sci.* **5**, 1008–1025 (2024).
 37. Grari, M., Yandouzi, M., Mohammed, B., Boukabous, M. & Idrissi, I. Comparative study of teachable machine for forest fire and smoke detection by drone. *Bull. Electr. Eng. Inform.* **13**, 1970–1979 (2024).
 38. Anh, T. T., Nhi, N. T. H. & Anh, T. H. Detecting dental caries through captured images using the machine learning technology teachable machine. *Asian J. Health Res.* **3**, 20–23 (2024).
 39. Kim, C. et al. Assessment of stem cell viability through visual analysis coupled with teachable machine. *Int. J. Stem Cells* (2025).
 40. Beyeler, A. et al. Divergent routing of positive and negative information from the amygdala during memory retrieval. *Neuron* **90**, 348–361 (2016).
 41. Namburi, P. et al. A circuit mechanism for differentiating positive and negative associations. *Nature* **520**, 675–678 (2015).
 42. Johnson, P. D. & Besselsen, D. G. Practical aspects of experimental design in animal research. *ILAR J.* **43**, 202–206 (2002).
 43. Ghosh, R., Gilda, J. E. & Gomes, A. V. The necessity of and strategies for improving confidence in the accuracy of western blots. *Expert Rev. Proteom.* **11**, 549–560 (2014).

Author contributions

A.T. conceptualized the study. A.T., C.D., and G.V. curated the data and conducted the formal analysis. A.T. and C.D. carried out the investigation. A.T. and H.A. developed the methodology. Project administration and resource provision were handled by A.T. Software development was performed by A.T., H.A., and L.M.-R. Supervision was provided by A.T. and H.A. Validation was carried out by A.T. and H.A. A.T., H.A., and L.M.-R. created the visualizations. A.T. wrote the original draft. A.T., C.D., G.V., H.A., and L.M.-R. reviewed and edited the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41514-025-00267-4>.

Correspondence and requests for materials should be addressed to Arthur Trognon.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025