Hidden Semi-Markov Models to Segment Reading Phases from Eye Movements

Brice Olivier, Anne Guérin-Dugué, Jean-Baptiste Durand

Hidden Semi-Markov Models to Segment Reading Phases from Eye Movements

Brice Olivier, Anne Guérin-Dugué, Jean-Baptiste Durand

RESEARCH REPORT
N° 9398
02/03/2021
Project-Team Statify

ISSN 0249-6399

Hidden Semi-Markov Models to Segment Reading Phases from Eye Movements

Brice Olivier[1] , Anne Guérin-Dugué[2],
Jean-Baptiste Durand[1]
Project-Team Statify

Research Report N°9398 — 02/03/2021 —27 pages.

Abstract:

Textual information search is not a homogeneous process in time, neither from a cognitive perspective nor in terms of eye-movement patterns, as shown in previous studies. Our objective is to analyze eye-tracking signals acquired through participants achieving a reading task aiming at answering a binary question: Is the text related or not to some given target topic? This activity is expected to involve several phases with contrasted oculometric characteristics, such as normal reading, scanning, careful reading, associated with different cognitive strategies, such as creation and rejection of hypotheses, confirmation and decision. To model such phases, we propose an analytical data-driven method based on hidden semi-Markov chains, whose latent states represent different dynamics in eye movements.

Four interpretable phases were highlighted: normal reading, speed reading, information search and slow confirmation. This interpretation was derived using model parameters and scanpath segmentations. It was then confirmed using different external covariates, among which semantic information extracted from texts. Analyses highlighted a good discrimination of reading speeds by phases, some contrasted use of phases depending on the degree of relationship between text semantic contents and target topics, and a strong preference of specific participants for specific strategies. As another output of our analyses, the individual variability in all eye-movement characteristics was assessed to be high and thus had to be taken into account, particularly trough mixed-effects models.

As a perspective, the possibility of improving reading models by accounting for possible heterogeneity sources during reading was discussed. We highlighted how analysing other sources of information regarding the cognitive processes at stake, such as EEG recordings, could benefit from the segmentation induced by our approach.

Key-words:

Eye-movement analysis, Reading, Computational models, Hidden semi-Markov chains, Scanpath segmentation, Decision processes

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Inria Grenoble Rhone-Alpes, 655 av. de l'Europe, F-38335 Montbonnot, France – jean-baptiste.durand@inria.fr

[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 11 rue des Mathématiques, Grenoble Campus BP46, F-38402 Saint Martin d'Hères Cedex

RESEARCH CENTRE
GRENOBLE - RHÔNE-ALPES

Inovallée
655 avenue de l'Europe - Montbonnot
38334 Saint Ismier Cedex France

Modèles de semi-Markov cachés pour la segmentation de trajectoires oculométriques en phases de lecture

Résumé : La recherche d'information textuelle n'est pas un processus homogène dans le temps, que ce soit d'un point de vue cognitif ou de celui des mouvements des yeux, ainsi que l'ont montré des études précédentes. Notre objectif est d'analyser des signaux oculométriques acquis lors de tâches où les participant.e.s doivent répondre à une question binaire : est-ce que le texte est lié ou non à un thème cible donné ? Nous nous attendons à ce que cette activité mette en jeu diverses phases avec des caractéristiques oculométriques contrastées, telle que la lecture normale, rapide, de confirmation et de décision. Pour mettre en évidence des différentes phases, nous proposons une méthode basée sur l'analyse de données fondée sur des modèles semi-markoviens cachés, dont les états latents représentent différentes dynamiques relatives aux mouvements des yeux. Quatre phases interprétables ont été mises en évidence : lecture normale, lecture rapide, recherche d'information et confirmation lente. Leur interprétation découle des paramètres du modèle et de la segmentation des traces oculométriques.
En perspective, nous discutons des possibilités offertes par cette approche pour améliorer des modèles de lecture en prenant en compte de potentiels modes de lecture hétérogènes mobilisés dans ce type de tâche. Nous mettons en évidence comment l'analyse d'autres sources d'information relatives aux processus cognitifs mis en jeu, telles que des enregistrements EEG, pourraient bénéficier de la segmentation induite par notre approche.

Mots clés : Analyse de données oculométriques, lecture, Modèles computationnels, Chaînes semi-markoviennes cachées, Segmentation de traces oculométriques, Processus décisionnels

1. Introduction

The study of the cognitive processes at stake in reading tasks is a major field of investigations in cognitive psychology and educational sciences (Rayner, 1998 ; 2009). To achieve this goal, eye tracking is a particularly useful and powerful source of information (Clifton et al., 2007). Eye trackers provide almost straightforward access to the time sequence of read syllables and thus, words, sentences and full texts. On the one hand, this has become fundamental material to explore and test hypotheses on mechanisms underlying processes at stake in semantic integration [occurring in reading tasks]. On the other hand, all the data and knowledge accumulated has allowed the development of models describing the control of eye movements during reading. The most popular models are EZ Reader (Reichle et al., 1999 ; 2003 ; 2006), SWIFT (Engbert et al., 2005 ; Nuthmann & Engbert, 2009) and Glenmore (Reilly & Radach, 2002 ; 2006). These models provide theoretical frameworks to understand the words identification, i.e., the lexical processing of words by the allocation of attention with the ocular movements. Such models can predict when (fixation duration) and where (fixate or skip the next word) to move eyes. The major difference among them is the early stage of attention allocation assuming a serial lexical process for EZ-reader or a parallel one for consecutive words for SWIFT and Glenmore. But in all cases, it is a reading strategy called "rauding" in the classification proposed by (Carver, 1990 ; 1992), who introduced the terminology.

In addition to eye tracking, electroencephalograms (EEGs) have been used since the 1980’s to characterize cognitive processes. In the one hand, using time-frequency analysis, specific waves with particular frequencies and locations in brains show synchronizations patterns, which are interpreted as markers of high-level cognitive functions and highlight how brain sites are synchronized in response to incoming sensory inputs (Klimesch, 1999). In the other hand, event-related potentials are able to show the time-course of the information process from the early visual input component, like P1 to later components like N400 for semantic integration or P600 for syntactic anomalies at a sentence processing level (Rayner & Clifton, 2009). It is now a complement to eye-tracking with, as main outputs, the possibility of predicting elementary components of reading tasks or inferring, for example, levels of semantic integration.

Until recently, experiments in this domain were restricted to carefully controlled experimental designs, and even more for the textual material. The combination of simultaneously-recorded EEG and eye-movement trajectories now offers new possibilities for analysing more complex reading scenarios, which are closer from real tasks of everyday’s life. Among those are for example, journal reading and web browsing for information search, in which readers have the possibility at every moment to decide to continue or quit reading, to change their focus of interest, etc. In this perspective, the ZuCo database consists of several datasets on natural reading of sentences from Wikipedia with different tasks such as plain reading, or reading combined with evaluating semantic relationships (Hollenstein et al., 2018).

As a consequence, the scientific topic of interest for researchers does not only focus on the reading processes, but also on the intertwined process of associated decisions on where to look next, closely linked to semantic integration and reader’s aims. An additional consequence is the non-homogeneous nature of the reading process: depending on the reader’s current focus and intention, it may go through different phases, such as text scanning, careful reading or in the case of multimedia documents, making a connection between a text and an image or a video. The ability to detect such states to identify which one is currently carried out and what their dynamics are, is thus of significant importance to explain and analyse eye-movement and EEG measurements.

In this work, our hypothesis is that such phase changes exist in poorly constrained experimental reading situations, are latent and can be deciphered by appropriate statistical analysis of eye-movement data. Phases can be obtained using segmentation methods, such as hidden Markov models (HMMs).

HMMs are generally dedicated to modelling processes subject to regime switchings, by associating a state to each possible regime. Not only do they provide signal segmentations but they also offer the possibility to model state dynamics, since in contrast to change-point detection, probabilistic properties of segment durations and transitions to previously-visited states are included into the model. They have been used successfully to model the dynamics of eye movements, both in reading tasks - as previously mentioned - and in exploration of images. For example in (Chuk et al., 2020), pairs of faces were shown to participants who had to indicate which face they preferred. An HMM-analysis of their eye movements aimed at capturing cognitive state transitions during the task, highlighting exploration and preference-biased periods and providing predictions regarding times to decisions.

HMMs were introduced in the context of reading tasks characterized by eye movements by Simola et al. (2008). Regarding EEG analysis, HMMs were used by (Obermaier et al., 2001) and (Kim et al., 2014), respectively in discrimination of imagined left vs. right hand movements and quantification of depth of anaesthesia. In these three studies, HMMs were essentially used in supervised classification of time series, the states accounting for regime switching over time.

In the context of analysing reading experiments, different classes of reading behaviours were defined and studied by (Carver, 1990). These classes were defined a priori in terms of tasks, mostly characterized by associated reading speeds of participants performing the tasks. The comparison with emerging states inferred from eye tracking in free reading experiments is an open question. Here, we address the problem with HMMs, which are relevant to identify in an unsupervised way scanpath segments with homogeneous vs. heterogeneous properties and simultaneously allow the clustering of similar segments. Our approach, although based on the same statistical models, is different from the one by Simola et al. (2008) since they used discriminative HMMs. As a consequence, their inferred reading states were defined so as to maximise discrepancies between models associated with three pre-defined tasks (word search, answering a question and search for the most interesting title within a collection). In our study, we propose to use hidden semi-Markov chains (HSMCs) to infer states that optimize predictions of eye movements in less constrained experimental conditions. These states are primarily defined by reading dynamics characterized by signed number of words crossed in outgoing saccades, interpreted in terms of progression, regression, refixation, etc. Moreover, they are fully characterized by saccade durations, directions and fixation durations, which integrate oculomotor features. This is also a difference with the approach by Simola et al. (2008), who based their HMMs on several variables that depend on text layout, such as saccade directions and lengths, while our approach is based on a single layout-independent variable.

2. Material and methods

2.1 Participants

Twenty-one healthy adults participated in the experiment, all French native speakers. Data of six participants were discarded because they did not follow the rules of the experiment thoroughly or data was too noisy during the acquisition with the eye tracker. The fifteen remaining participants (6 women and 9 men aged from 20 to 32 years, 25 years 9 months ± 7 years 6 months, mean plus or minus standard deviation, *sd*) had normal or corrected-to-normal vision. There were free of any medical treatment or any neurological or psychiatric disorder, past or present. None of them had prior experience with the experimental task. All gave their written and informed consent prior to the experiment and were paid 20€ for their participation. The whole experiment was reviewed and approved by the ethics committee of Grenoble CHU (“Centre Hospitalier Universitaire”) (RCB: n° 2011-A00845-36).

2.2 Material

Textual material was the same as in Frey et al. (2013). 180 short texts were extracted from the French newspaper *Le Monde*, edition 1999. Texts were given a topic and were constructed around three types, those which were highly related (HR) to the topic, or moderately related (MR) to the topic, or unrelated (UR) to the topic. There were 60 texts of each type, hence 180 in total. The semantic relatedness of the text to the topic was controlled by Latent Semantic Analysis (LSA) (Deerwester et al., 1990). To do so, LSA was trained on a French corpus of 24 million words composed of all articles published in the newspaper *Le Monde* in 1999 and a word or set of words (sentence, text, etc.) was represented by a vector in this 300-dimension semantic space. A cosine function was used to compute the similarity between vectors composed for the topic in the one hand and for the text in the other hand. The higher the cosine value, the more related the topic and the text are. For all highly related topics, semantic similarity with the text was above 0.2, while for all unrelated topics, semantic similarity was below 0.06. The moderately related texts were in-between.
All the texts were composed of an average of 5.18 ± 0.7 (mean plus or minus standard deviation) sentences and 30.1 ± 2.9 words. The average number of characters of words was 5.34 ± 3.24. For the screen layout, the average number of lines was 5.18 ± 0.68, and the text was displayed with 40.1 ± 5.4 characters per line.

2.3 Experimental procedure

The goal of the experiment was to assess as soon as possible during reading whether the text was or not related to a given topic.
First the topic was presented to participants and then they clicked to start the trial. Then a fixation cross was presented on the left of the first character at the first line, to stabilize the gaze location at the beginning of the text. The duration of this step was random to avoid anticipation of the reading start. They also did not know whether the text was HR/MR/UR so that they could not plan on a search strategy in advance. The texts were randomly ordered for each participant. When the text was displayed, participants read and had to mouse-click as fast as possible to stop reading and decide during another screen if the text was related or not to the topic. The trial was then repeated for the 180 texts with breaks in-between.

2.4 Apparatus

Each text was displayed at the centre of a 24-inch screen with a resolution of 1 024 by 768 pixels. Participants were seated 68 cm in front of the screen. Thus, texts covered in average 21° x 11° of visual angle and each character covered 0.52°of horizontal visual angle, corresponding to about 3.8 characters in fovea.

Along the experiment, electrical cerebral activity was measured through a 32-channel electroencephalogram (EEG) with 1 000 Hz sampling rate[3] (Brain Products GmbH). X/Y eye positions on screen were collected using a remote binocular infrared eye tracker EyeLink 1000 (SR Research).

2.5 From gaze fixations to words and to reading strategies

During trials, the eye tracker gave the position of each fixation on the screen, and the fixation duration. The minimum fixation duration threshold was set to be 80ms whereas the maximum duration was 600ms. All fixations outside these limits were removed for all analyses.

A posteriori it was necessary to know which word was being processed by the participant. First, the word identification span was defined as the necessary area from which a word can be identified. This span varies according to the direction of the reading, the alphabet, or the language, but can also be micro-context related as it was for several reading models such as EZ-Reader (Reichle et al., 2003) or the SWIFT model (Engbert et al., 2005). For simplicity, we used a fixed span that is considered for most of Latin languages (Rayner, 1998): an asymmetrical window of 4 characters left and 8 characters right to the fixation. Moreover, a word may not entirely be located in the word identification span. Based on Farid and Grainger (1996), we considered a word to be processed if at least 1/3 of its beginning or 2/3 of its end was inside the window. This result was obviously language sensitive, only valid in French, and considers that the important root of the word necessary to its understanding is located at the beginning of the word. Finally, another hypothesis had to be made on the processed word within the window since several words might be captured. For this, we assumed that only one word could be processed during a given fixation and that this word was chosen as the closest to fixation centre, excluding stop words. Consequently, one word per fixation was selected. Thanks to this enhancement, features characterizing the reading strategy were defined.

Each fixation was associated to its outgoing saccade. From now throughout the article, the term “saccade” will be referred to the outgoing saccade of a given fixation. Thus finally, data associated with each fixation were the fixation duration, the fixed word, the saccade amplitude expressed in visual degree, the number of crossed words between two saccades and the saccade duration. The saccade as a marker of the reading strategy was characterized by this number of crossed words, which would be negative for a backward progression, null for a refixation or positive for a forward progression. At the text level, the reading speed is known to be a global marker of the reading (Carver, 1990). This can simply be measured by how far (in words) a reader can go in a text per time unit. While our aim was to segment text according to reading strategies, this feature had to be evaluated at a finer level. At the saccade level, the reading speed was evaluated as the number of crossed words during the saccade plus one (the fixed word during the current fixation) divided by the current fixation duration and the saccade duration. At an intermediate level on a text segment, the speed reading was evaluated as the number of crossed words plus the number of not yet fixed words divided by the sum of the fixation durations and saccade durations. For a text, composed of the different segments with different sizes (number of fixations) but with the same reading strategies, the reading speed was computed as the ratio of the number of words (fixed and crossed over in saccades) summed over all segments divided by the sum of fixation durations and saccade durations over all segments. If some word was crossed several times in a same scanpath, it was counted only once in the total number of words.

2.6 Statistical analysis

To test the assumption of several reading strategies, we used an approach inspired by Simola et al. (2008) based on state-space models. The principle is to associate observations X_t at each time step t (fixations) with some underlying (or “hidden”) categorical random variable (state) S_t representing the current probability distribution of some eye-movement characteristics. Thus, successive identical states represent homogeneous segments in terms of these eye-movement properties (see below). State transitions correspond to marked changes in eye-movement properties. These properties, potentially combined with external variables, may lead to state interpretation as reading strategies.

We proposed two main modifications to the approach proposed by Simola et al. (2008). The first one was another eye-movement statistic to characterize reading strategies independently from text layout. To achieve this goal, we used saccade amplitudes, as Simola and collaborators; however here this feature was not coded in pixels but rather defined as the number of words crossed during a saccade (we use “crossed” instead of “skipped” since in some cases, those words may have been inferred by readers who thus do not need to fix them). In this way, this feature was independent from text layout; thus return sweep saccades could be included in the analysis, as their saccade amplitude did not impact the analysis.

Thus, we used as observed variables a categorical function X_t of the signed number of words W_t crossed in one saccade (positive for forward progression, null for refixation and negative for backward progression). The second modification was the use of hidden Semi-Markov chains (Yu, 2010) to relax the Markovian assumption on states. Specifically, we resorted to particular HSMCs called explicit duration HMMs. These models assume that the hidden process $S=(S_t)_t$ enters an initial state k with probability π_k. Once a

[3] Potential use of EEGs is mentioned in Section Discussion. EEG data are not used in analysis for this report.

new state entered, the process stays in that state during a random number of time steps with distribution p_{θ_k}. When leaving state k, some new state ℓ is entered with probability $A_{k,\ell}$. At time t if $S_t=k$, then $X_t=x$ occurs with probability $P(X_t=x|S_t=k)= B_{k,x}$. Denoting by K the number of possible values for S_t, the set of parameters λ is thus defined by $\boldsymbol{\pi}=(\pi_k)_k$, the transition matrix A, the observation probability matrix B and the parameters $(\theta_k)_k$ for the state duration distributions.
Several choices are possible to define X_t. We tried the following possibilities: outgoing saccade direction (upward, forward, downward, backward as in Simola et al. (2008), indicator variables for reaching last fixation, fixation durations, outgoing saccade amplitudes, numbers of characters or words crossed in outgoing saccades, numbers of words crossed in outgoing saccades, as well as the following proposals for defining X_t (referred to as Read Mode):

- $X_t=W_t$ if W_t = -1, 0 or 1; X_t="<-1" if W_t < -1; X_t=">1" if W_t > 1 (Read Mode 1);
- $X_t=W_t$ if W_t = -1, 0, 1 or 2; X_t="<-1" if W_t < -1; X_t=">2" if W_t > 2 (Read Mode 2);
- $X_t=W_t$ if W_t = -2, -1, 0, 1 or 2; X_t="<-2" if W_t < -2; X_t=">2" if W_t > 2 (Read Mode 3);
- $X_t=W_t$ if W_t = -1 or 2; X_t="<-1" if W_t < -1; X_t=">2" if W_t > 2; X_t="0/1" if W_t = 0 or 1 (Read Mode 4);.

In the sequel, $W_t > 1$, $W_t = 1$, $W_t = 0$, $W_t = -1$ and $W_t < -1$ will be referred to using the following abbreviations respectively: Fwd+, Fwd, Rfx, Bwd and Bwd-.
For each possible choice for X_t, we used BIC (Cappé et al., 2007) to select the number K of hidden states. Then we kept the choice for X_t that minimized joint state entropy, as computed according to Durand and Guédon (2016). Models were discarded if they yielded mean segment lengths shorter than 4 fixations or longer than 25 fixations, since these were either specific of one single subject or could not be interpreted as reading strategies.
The HSMC parameters were estimated by maximum likelihood using the EM algorithm with 100 initial parameter values. The considered parametric distributions for the quantitative variables were uniform, Poisson, binomial and negative binomial with shift parameters in the four cases. We used the HSMC implementation from the VPlants software, which is part of the OpenAlea platform (Pradal et al., 2015).
The states were interpreted using the emission and state duration distributions, together with the transition matrix. Potentially, reading phases could be defined by patterns involving short cycles between several model states; in this case, these cycles were merged into phases for the sake of interpretability.
The most probable states values (maximum a posterior, MAP) were computed for each scanpath to obtain their segmentations in homogeneous zones with respect to the distribution of X_t.
To highlight subject variability in scanpaths, correspondence analysis (CA) (Greenacre, 1984) and an independence test were performed on the contingency table defined by the number of fixations in each phase for each subject, merging fixations from all scanpaths.
The effects of text type were assessed by tests in regression models. Three families of models were considered: linear mixed models (LMMs), binomial generalized linear mixed models (BGLMMs) and multinomial generalized linear mixed models (MGLMMs) with Gaussian random individual effects. LMMs were used to assess effects of covariates on quantitative variables (reading speed and fixation durations computed for each scanpath) using the *lmer* package of the *R* software (Venables & Ripley, 2002). Significance of fixed effects within a given model was determined by ANOVAs. BGLMMs were used to assess effects of covariates on binary variables (occurrence of phase transitions or not) using the *glmer* package in *R*. Model selection regarding fixed effects was achieved by computing BIC on the whole collection of models built from all subsets of covariates and their interactions. BIC for mixed models was defined as in (Delattre et al., 2014). Significance of individual effects was assessed by comparing BIC values of the models with the best set of covariates, considering in turn variants with and without individual random effects. This was complemented by the use of confidence intervals on the standard deviation of random effects, using profile likelihood as described in (Bates et al., 2014). MGLMMs were used to assess the effect on nominal categorical variables (Read Mode and Phase) using Bayesian estimation with the *MCMCGlmm* package in *R* (Hadfield, 2010). Significance of individual effects was assessed using credibility intervals at level 0.995 on variance parameters while significance of fixed effects was assessed by comparing DIC values (Spiegelhalter et al., 2002). The model minimizing the considered criterion was selected.
It is assumed is this study that subjects take their decisions by detecting semantically related words to target topics (in HR texts) or incongruent words (in UR texts). It is thus expected that such words trigger phase changes. This was addressed by first detecting these words called "trigger words" and then, assessing the effect of distance to trigger words and text types, on the probability of phase transitions. Thus, the HR category was refined depending if at least one word of the target topic ("target word") appeared or not in the text. In the positive case, these texts were referred as "HR+" and the "trigger words" would include "target words". The categorization including HR+ texts is referred to as "extended text type" hereafter.
"Trigger words" were detected using a FastText[4] representation of words (Joulin et al., 2017). Only two "trigger words" per text were defined. This consists in embedding words into Euclidean spaces, allowing for computing semantic proximities between words using Euclidean metrics (here, the cosine distance). For HR and HR+ texts, we considered as "trigger words" the two words that had the highest cosine with the presented target topic. For UR texts, these were the two words with lowest cosine with topic weighted by inverse word frequencies, to account for the fact that incongruent words are both unrelated to the topic but also specific of other

[4] FastText representation was preferred to LSA representation because the former is more accurate regarding acronyms.

topics. In MR texts, the concept of trigger word is not clearly defined, since texts may be more or less related to the topic and may contain both incongruent and target words. Thus, the two considered words where those with the highest and the lowest cosine.
Since HSMC states are random and hidden, the times of transitions are uncertain. Thus, instead of considering transition or not at “trigger words”, the effect of distance of transitions to “trigger words” was measured in numbers of fixations, focusing on “trigger words” with lowest distance to transitions. Its effect of transition probabilities was assessed using regression models. Firstly, frequencies for the distances associated to each incoming phase (among every possible distance for that phase) were modelled with linear mixed regressions, using distance, text type and phase as predictors, with subjects as random effects. Secondly, the binary random variable corresponding to occurrence or not of a transition at each possible distance of a fixation to “trigger word” was modelled with generalized linear mixed regressions. Binomial distributions were considered, using the canonical link function and the same three predictors as above. In both approaches, models with interactions of order 2 and 3 between predictors were estimated, in addition to models without interaction. Models were compared using BIC. The model with minimal BIC (referred to as M1) was then used to assess the significance of random subject effects, by comparing BIC with that of a model without random effects. M1 was also compared with the model obtained by removing distance as a predictor (referred to as M0). The justifications for using both approaches (linear models on frequencies or GLMMs on binary variables) were twofold: firstly, GLMMs easily suffer from lack of convergence for high-order interactions and thus some of these models cannot be compared. Secondly, the linear assumptions on frequencies seemed reasonable given the shape of the cloud of points (See Figure 6).
Since texts were rather short and total numbers of fixations were rather low (see Sections 2.1 and 3.1), the effect of small distances increasing transition probabilities could be credited to distances being necessarily small, even if transitions were drawn at random and independently from the positions of “trigger” words. To assess this possible bias, randomized procedures were developed. The first one consisted in sampling transition locations (uniformly without replacement) and permuting the order of phases, thus constraining the number of transitions to remain the same within each scanpath. The second one consisted in sampling the number of transitions with replacement within their empirical distribution and drawing phase values uniformly, with the constraint that successive phases must be different. In both cases, the whole data set was resampled 1 000 times. Each time M1 and M0 were estimated again, as well as their difference in BIC. The percentage of differences obtained by resampling was compared to the true difference, thus assessing the significance of the distance effect.
Independently of the HSMC model, the effect of text type (categorical variables HR, MR and UR) on saccade amplitude, fixation duration, reading speed, number of fixations per scanpath and Read Mode frequencies were assessed using MGLMMs and the *MCMCGlmm* package in *R* as explained hereinbefore. The effect on quantitative variables was assessed using LMMs. In both cases, random subject effects were included and their significance was tested. Effects of categorical predictors were tested using analyses of variance (ANOVAs). Normality of residuals in LMMs was assessed using Shapiro-Wilk normality tests complemented with histograms of empirical residuals.
The values of Read Mode were coded as follows: Bwd- / 0 / $W_t < -1$; Bwd / 1 / $W_t = -1$; Rfx / 2 / $W_t = 0$; Fwd / 3 / $W_t = 1$; Fwd+ / 4 / $W_t > 1$.

3. Results

3.1 Summary statistics on observed data

After visual inspection of all scanpaths, some of them were discarded if the drifts on gaze positions were too large, making it impossible to assign a word at each fixation, typically when the gaze positions were in between text lines. Moreover, scanpaths with less than four fixations were removed (assumed to be non-characteristic of the task). Globally, HSMC models were run on 2 390 scanpaths and 39 564 fixations.

Scanpaths number	Fixations number per text	Fixation duration [msec]	Saccade amplitude [°]	Saccade amplitude [w]	Reading speed [wpm]
159.3 ± 22.4	16.6 ± 4.7 [7.1]	184.0 ± 23.1 [62.3]	5.3 ± 0.67 [3.9]	1.9 ± 0.5 [2.5]	404.9 ± 119.8 [155.8]

Table 1: Average ± between-participant standard deviation [within-participant standard deviation] for the number of texts, the number of fixations per text, the fixation duration, the saccade amplitude in [°] and in number of words [w], and the reading speed [wpm]

Table 1 summarizes the average individual statistics per participant, on the number of scanpaths, the number of fixations per text, the fixation duration, the saccade amplitude expressed in visual degree [°] or in number of crossed words during each saccade [w], and the reading speed expressed in words per minute [wpm].

3.2 Effect of text type on scanpath characteristics

3.2.1 Statistics before segmentation by HSMC

The scanpaths were characterized by the number of fixations, the fixation duration, the saccade amplitude expressed in degrees [°] and in words [w] (cf Table 2). Text type had a strong effect on the number of fixations per text (ANOVA highlighting significance at level 10^{-16}), with a strong individual variability (BIC difference of -877 with null model ignoring individual effects). MR and HR texts did not show significant differences while UR texts have quite lower numbers of fixations. As illustrated in Figures S1 and S2 in Appendix, text type showed no effect on mean fixation duration and saccade amplitude.

Text	Number of fixations per text	Fixation duration [msec]	Saccade amplitude [°]	Saccade amplitude [w]	Reading speed [wpm]
HR	15.4 ± 4.4 [6.6]	185.8 ± 22.9 [63.0]	5.3 ± 0.7 [3.9]	1.8 ± 0.5 [2.2]	381.9 ± 117.7 [124.8]
MR	20.1 ± 5.0 [7.0]	184.2 ± 22.8 [62.6]	5.4 ± 0.7 [3.9]	1.9 ± 0.5 [2.6]	365.2 ± 99.1 [116.5]
UR	14.3 ± 5.1 [6.0]	181.9 ± 23.9 [61.0]	5.2 ± 0.6 [3.8]	2.0 ± 0.5 [2.5]	466.8 ± 151.6 [183.9]

Table 2: Average ± between-participant standard deviation [within-participant standard deviation] for the number of fixations per text, the fixation duration, the saccade amplitude expressed in degree [°] and in number of words [w] and the reading speed [wpm], depending on the type of text

Text type had also a strong effect on reading speed (ANOVA highlighting significance at level 10^{-16}), with a strong individual variability (BIC difference of -1098 with null model ignoring individual effects). MR and HR texts did not show marked differences while UR texts had quite larger reading speeds.
Normality tests indicated lack of normality of empirical residuals in both models for reading speed and number of fixations at level 10^{-16}, presumably due to skewness in their distribution (see Figures S3 and S4 in Appendix).
The average numbers of fixations, fixation durations, saccade amplitudes and reading speeds per text type are summarized in Table 2.

	Long regression (Bwd-)	Regression (Bwd)	Refixation (Rfx)	Progression (Fwd)	Long progression (Fwd+)
HR texts	0.06	0.02	0.26	0.23	0.42
MR texts	0.07	0.02	0.26	0.21	0.43
UR texts	0.06	0.02	0.25	0.22	0.45

Table 3: Average per text for the five Read Mode frequencies depending of the type of text

GLMMs modelling the effect of text type on Read Mode showed significance of random individual effects, with 99.5% credibility intervals of (0.2, 2.3) for the variance for individual effect. This is in accordance with the difference in DIC values between models with and without random individual effects (-125). The difference in DIC with the null model was positive, indicating absence of effect for text type. This result is somehow counter-intuitive given the significance of all individual parameters at level 0.01, particularly regarding overrepresentation of Fwd+ in UR texts. This does not seem either in accordance with empirical distributions depicted in Figure S5 in Appendix. However, this lack of significance could be explained by predominance of individual variability. The Read Mode frequencies per text type are summarized in Table 3.

Output process defined as:	Entropy
Simola et al. (2008)	10 104
Read Mode 1	10 771
Read Mode 2	11 581
Read Mode 3	11 165
Read Mode 4	8 027

Table 4: State entropy at optimum number of states for the five considered output processes.

3.2.2 HSMC modelling

Entropies of HSMC models associated with different possible definitions of the output process are summarized in Table 4. It turned out that Read Mode 4 and the choice of Simola et al. (2008) led to models with the lowest entropies. However, these models yielded mean segment lengths shorter than 4 fixations or longer than 25 fixations. These segments were either specific of one single subject or could not be interpreted as reading strategies and the models were thus discarded. Entropy minimization among the other possibilities led to using the following output process: Read Mode defined as $X_t = W_t$ if W_t = -1, 0 or 1; X_t =”<-1” if W_t < -1; X_t =”>1” if W_t > 1 (Read Mode 1).
Using the Read Mode 1 outputs, BIC selected a 5-states model. The estimated parameters and distributions are represented in Figure 1. States 0 and 1 are characterized by short sojourn lengths and quasi-systematic alternation, which is typical of a macro-state, here referred to as “phase”. Thus, phases were defined as Phase 0 = {State 0, State 1} and Phase i = {State i+1} if i > 1. Initial phase probabilities are 0.75 for Phase 0, 0.012 for Phase 1, 0.24 for Phase 2 and 0 for Phase 3. Phase 0 has intermediate probabilities for Fwd+, Fwd and Rfx; thus it can be interpreted as the normal reading phase (abbreviated as NR). Phase 1 usually separates two runs of Phase 0 and its duration is short. Its interpretation is not obvious but since it breaks normal reading phases and has high Rfx, Bwd- and Fwd+ probabilities, we named it an information search phase (IS). Phase 2 is transitory, meaning that once left it cannot be returned to. It has the highest Fwd+ probability and thus can be interpreted as a speed reading phase (SR). Phase 3 is absorbing, meaning that no other phase can be accessed from it. It has the highest Bwd and Bwd- probabilities and is followed by no other phase; thus it is interpreted as a slow confirmation phase (SC). Phase interpretation is also consolidated by state restoration, as illustrated in Figures 2 and 3. Other scanpaths are provided in Figures S6 and S7 in Appendix to illustrate other typical behaviours. Figures 2 and S7 highlight that in SR, fixations tend to be sparser than in NR while in SC, backward fixations are more frequent. Note that every scanpath does not necessarily end in Phase 3 since decisions can be reached in any other phase and even after before reading every clause in the sentence, as illustrated in Figure S6.
To validate state interpretations systematically in terms of reading speed, the latter was computed for each phase. Mean reading speed was 304 words per minute (wpm) in NR, 183 wpm in IS, 509 in SR and 263 in SC, which is consistent with our former interpretation. Linear mixed models were used to test the effect of phase and individual variability, accounting for already confirmed text type effects (see Section 3.1). The phase effect was assessed as significant by ANOVA at level 10^{-16}, while with a BIC difference of -875 with the null model, individual variability was assessed as quite significant. The estimated standard deviations were 96 (individual) and 235 (residual), the 95% confidence interval for the individual standard deviation being (66, 140). The normality test indicated lack of normality of empirical residuals at level 10^{-16} (see Figures S8 in Appendix).
The phase sample distribution is represented in Figure S10 (see Appendix). The effect of extended text type on phase distribution (Figure 4) was assessed using GLMMs. The credibility interval at level 99.5% for the variance of individual effect was (0.21, 2.02), indicating significant variability. This is consistent with the high difference in DIC values between models with and without random individual effects (-1 730). The difference in DIC values between the null model without text type effect was moderate (-24), indicating that the visible effect in Figure 4 is somewhat masked by individual variability. MR texts lead to more frequent use of phase 3 (SC) and less frequent use of phase 2 (SR), UR texts to less frequent use of phase 0 (NR) and more frequent use of phase 2 (SR), HR texts to more frequent use of phase 0 (NR).
An independence test between phase and subject yielded a test statistic of 2.1×10^5 for 42 degrees of freedom, corresponding to very clear rejection of independence (the p-value cannot be computed since this is too close to 0). The first CA plane is represented in Figure 5. The ratio of preserved inertia is 99% in this plane. Three clusters of individuals were highlighted: 1) individuals using phases 0 and 1 (normal reading and information search) at the detriment of the other phases, e.g., Subject 2 at the left-hand part of the figure; 2) individuals using phase 2 (speed reading) primarily at the detriment of phase 0 and secondarily 3, e.g., Subject 19 at the bottom-right corner of the figure (fast readers); 3) individuals using phase 3 (slow confirmation) primarily at the detriment of phase 0 and secondarily 2, e.g., Subject 4 at the upper-right corner of the figure (careful readers).

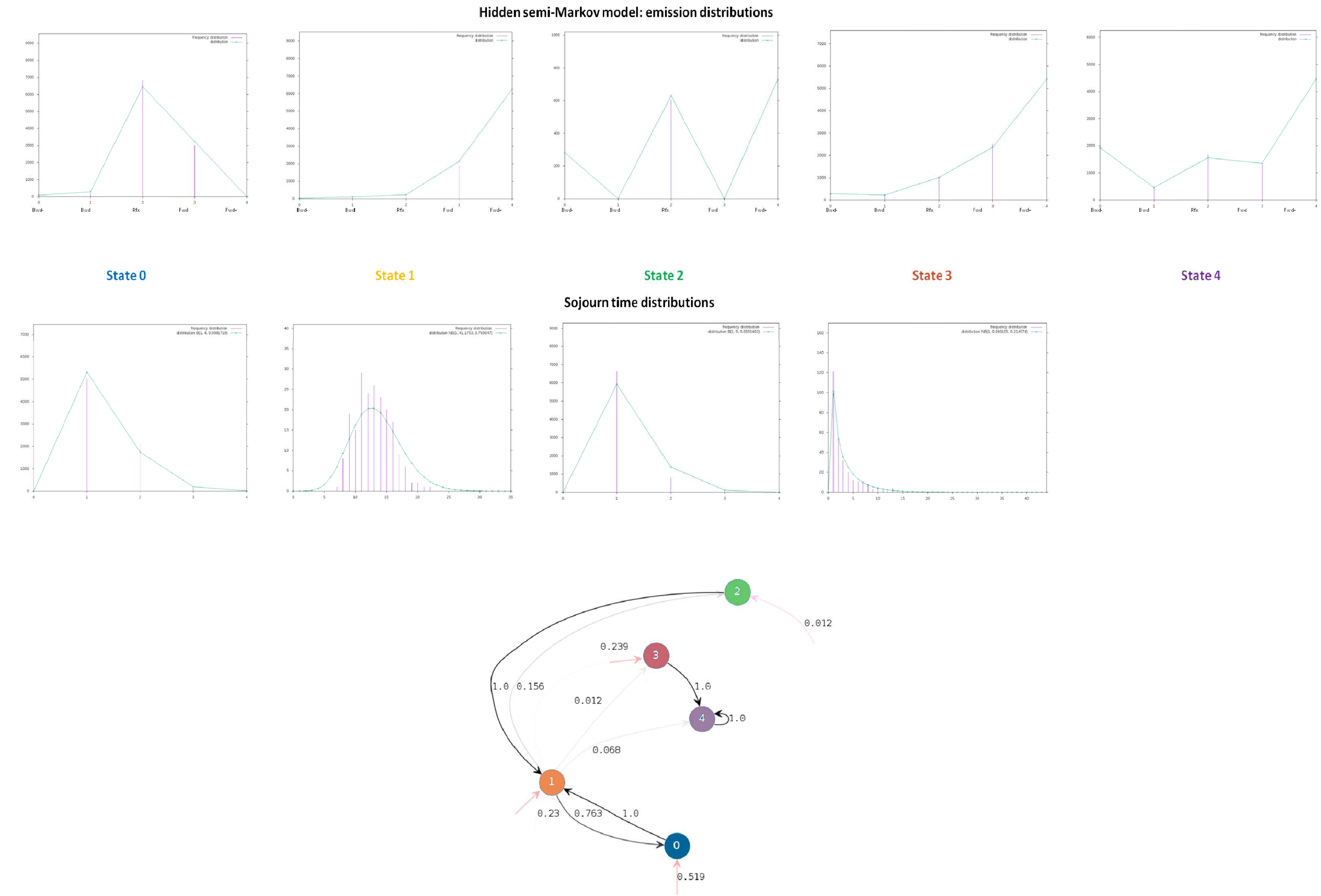

Figure 1. Estimated HMSC parameters. Lines 1 and 2: estimated emission and sojourn state distributions for each state. States are in columns. State 4 is absorbing and has not estimated sojourn time distribution. Line 3: estimated transition diagram between states. The initial state distribution is representing by pink arrows pointing to possible initial states but issued from no other state.

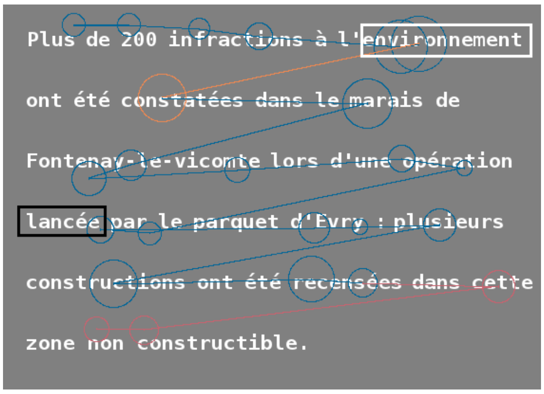

Figure 2. Scanpath of some MR text with phase restoration. Target topic is "Nuclear waste". Phase 0 (normal reading) is in blue, phase 1 (information search) in orange and phase 3 (slow confirmation) in red. Translation: "More than 200 violations to environment were recorded in the swamp next to Fontenay-le-vicomte during an operation launched by Evry's prosecution service; several buildings were recorded within this non-buildable land." The word framed in white is the closest to target topic, that framed in black is the farthermost to target topic.

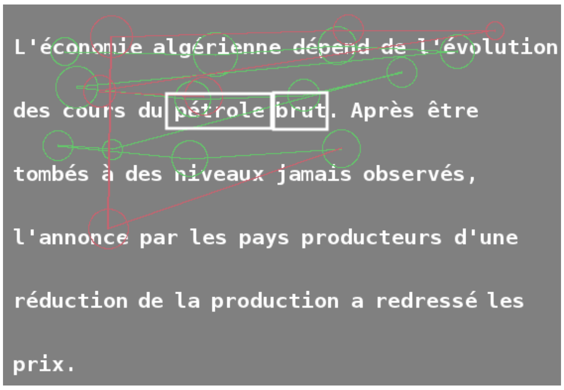

Figure 3. Scanpath of some HR+ text with phase restoration. Target topic is "Oil Price". Phase 2 (speed reading) is in green and Phase 3 (slow confirmation) in red. Translation: "Algerian economy depends on the evolution of crude oil exchange rates. After they fell down to historically low levels, the announcement by producing countries of production reduction led to price recovery". The words framed in white are the closest to target topic (the first one being "Oil").

Figure 4. Phase sample distribution per extended text type. Phase 0 (normal reading) is in blue, Phase 1 (information search) in orange, Phase 2 (speed reading) in green and Phase 3 (slow confirmation) in red.

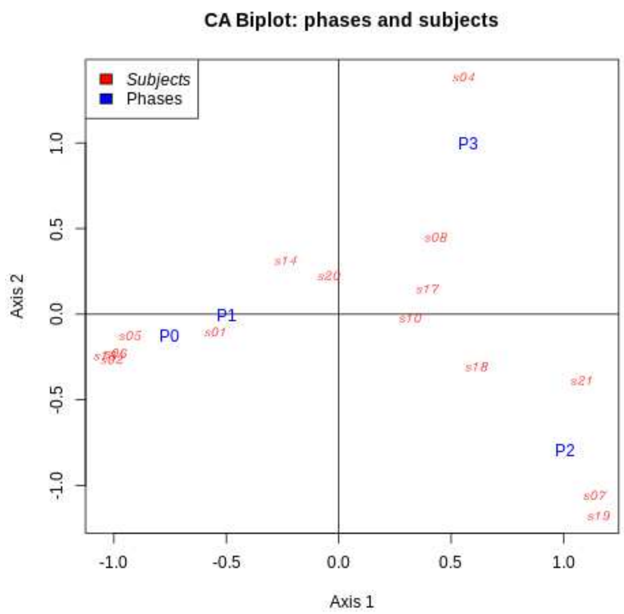

Figure 5. First principal plane of correspondence analysis. Phases 0 to 3 are represented by blue points labelled as P0 to P3. Subjects 1 to 21 are represented by red points labelled as s01 to s21.

The results related to the effect of trigger words are presented in Figure 6. Each diagram represents the distance (in number of fixations) between a transition and the closest trigger word (x-axis) together with the associated transition frequency (y-axis). The three diagrams correspond to different incoming phases (value following a transition). The regression line is shown for each extended text type. Transitions to phase 2 (SR) are too rare (See Figure S12 in Appendix) and thus their frequencies cannot be reliably estimated. Examples of trigger words in specific scanpaths are illustrated in Figures 2 (MR text), 3 (HR+) and S6 in Appendix (UR). Regression lines with low negative slopes correspond to transitions occurring more frequently around keywords. Lines have more negative slopes in UR texts, which shows that incongruent words tend to induce immediate changes in reading strategies. They probably also have a strong effect on the decision to stop reading and proceed to the answer, although this has not been assessed here. The slopes in MR texts have intermediate values between those of UR and HR/HR+ texts, suggesting that even the concept of trigger words in MR texts is ill-defined (notion of MR text being even vague itself): Subjects may base their decisions on either incongruent words or words that are related to target-topic, to decide how to explore texts. The slopes in HR and HR+ are also negative and cannot be assessed as different, showing that reading words from the target topic has no stronger effect on strategy changes than reading words only close to the target topic.

It can be seen from Figure 6 that linear models are relevant to explain dependencies between distances and frequencies. Thus the effects of distance, extended text type and phase on frequencies were considered through LMMs. The model with order-3 interactions between distance, phase and text type minimized BIC (-787). The second lowest value of BIC was -780 (dist + phase * text type). The best model was compared with a linear model with the same structure of effects but no individual random effect, yielding a BIC difference of -42 (significant individual variability). The 95% confidence interval on sd parameter was (0.05, 0.1). The effect of each factor was tested individually, highlighting some very strong effects of phase and extended text type, as well as some significant effect of distance (BIC difference of -32, p-value in ANOVA $3x10^{-5}$), suggesting once again some strong individual variability masking the effect of distance possibly highlighted by Figure 6. By way of comparison, GLMMs applied on binary variables corresponding to occurrence or not of transitions at a given distance to trigger words did not converge for several combinations of interactions; however, they led to conclude to very strong marginal significance of the three effects.

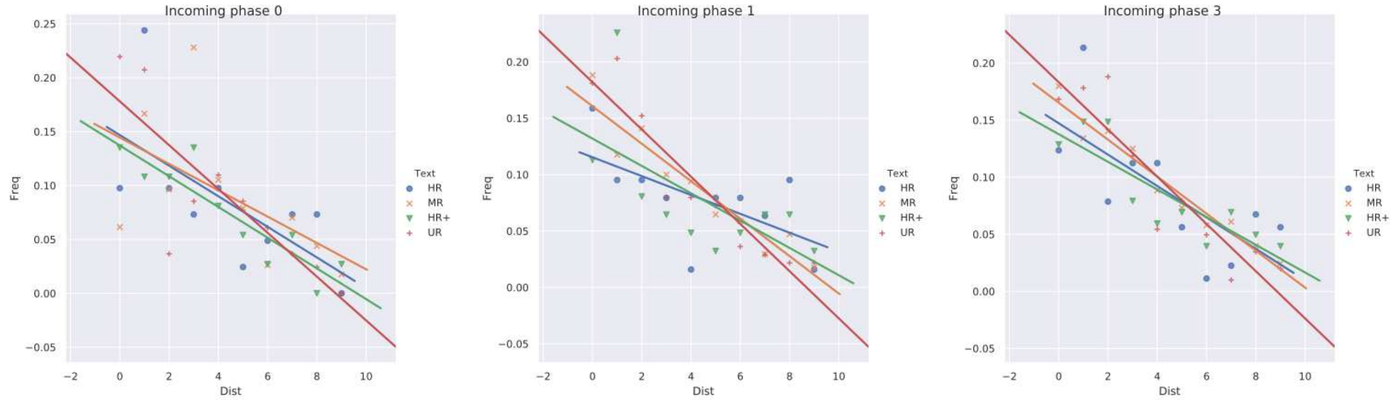

Figure 6. Relationship between distance to trigger words and frequencies of transitions arriving into phase 0 (normal reading, left), 1 (information search, middle) and 3 (slow confirmation, right).

Randomized tests showed that random allocations of transitions yield some lower difference in BIC with the null model than the true difference in 54% of simulated sequences in the case of constrained phase permutations and in 72% of simulated sequences in the case of free phase permutations (See BIC difference histograms in Figures S13 and S14 from Appendix). This suggests that the effect of distance could be partly due to scanpath shortness. However, the same procedure applied to GLMMs on absence / presence of transition led to in 0% BIC difference that are smaller in simulated sequences than the BIC difference in true data, both in constrained and free phase permutations settings (See BIC difference histograms in Figures S15 and S16 from Appendix).

4. Discussion

Our methodology confirmed the importance of modelling phase changes for accurate interpretation of eye movements in loosely-controlled information search tasks. State interpretation was supported by contrasted characteristics in terms of sequencing during the task, Read Mode frequencies, reading speeds and text semantics, summarized here with text types and trigger words.

Particularly, reading strategies were interpreted in terms of reading speeds using the Read Mode variable, which is directly connected to HSMC parameters. It is however interesting to compare reading speeds obtained in each phase to those associated with Carver's reading "gears" (1990): learning, rauding, scanning and skimming. The mean speed of 304 wpm in NR corresponds to rauding, the speed of 509 wpm in SR is intermediate between skimming and scanning, 263 wpm in SC is intermediate between rauding and learning, while the speeds of 183 wpm in IS is comparable with learning.

Although our study has somewhat different focuses and aims compared to the study by Simola et al. (2008), the latter addressed related questions with related tools. In particular, our state interpretations can be compared: they highlighted three states, which were stable in the three different tasks they considered. Three distinctive HMMs (one for each task, "word search", "question-answer", "true interest") were embedded into a unique discriminative model in order to classify each observed trial, from eye-movement features, into one of the three classes, i.e. one of the three tasks. The model selection based on classification highlighted three hidden states for each model. Those states, called scanning, reading and decision were interpreted on the basis of the distribution of the input observed data, specifically on the fixation durations, saccade amplitudes and saccade directions. In our model, five states gathered into four phases were observed. When comparing the interpretations made by Simola et al. (2008) and ours, there is a very simple matching between them: scanning can be assimilated with SR, decision with SC and reading with NR, while IS seems more specific to our experiments.

Another significant output of this study is the predominant individual variability, which can be observed at virtually every level of the analysis. Here again, HSMC models led to a precise characterization of this variability in terms of favoured use of contrasted reading strategies depending on each individual, leading to some clustering of the population. The individual variability was so high that probably, it partially or totally masked the effects of other factors, such as text type or distance to trigger words, to phase-related features. This suggests, on the one hand, that accounting for individual variability in modelling is of uttermost importance and on the other hand, that some additional participants may have to be involved in the experiment so as to confirm the effects of text type and distance to trigger words on transition probabilities.

Comparisons between the three text types based on different indicators (reading speed, phase distribution and transitions, effect of trigger words, scanpath lengths) highlighted that UR (unrelated) texts are easy to process (more speed reading SR, less normal reading NR) whereas MR (moderately related) texts are more difficult, as expected (more slow confirmation SC, less SR). The difficulty of HR (highly related) texts is intermediate and no significant difference was found between HR and HR+ texts. In fact, for UR texts, it is less the semantic construction of the text as such that matters, than the elaboration of the semantic similarity of the text with the displayed topic. This semantic similarity estimated in the LSA space was always very low whatever the scanpaths, because these texts contain words with low frequencies that are unrelated to their target topic. However, for MR texts the semantic construction must contribute to their comprehension so as to be able to answer regarding the link with the theme. Finally, for HR texts, there is a strong variability in the construction of the semantic link between the topic and the read words, because this link depends on the presence or absence in the text of a word belonging to the topic. As a result, our study shows the possibility to obtain such characterization of the different text types by using just eye movements and a very rough description of the text semantic contents (summarized by distances to trigger words).

The quantitative results of our study could be used to improve existing reading models such as EZ-Reader or SWIFT. Indeed, these models try to identify, through eye movements, the different phases in the reading process such as overall attention shifting and lexical decoding. Considering the EZ-reader model, there are two main assumptions. The first hypothesis states that attention is allocated serially on one word at a time and that attention is intrinsically linked to lexical processing. The second hypothesis states that eye-movement control and saccade control are decoupled. The model assumes that the lengths and the frequencies of words have a great importance for the lexical steps, from the earlier step, called “familiarity check” to the last step, called “completion of lexical access”. It is well known that the fixation duration on a word is a function of a range of linguistic factors and among these, word length and frequency are lexical variables with a large effect on fixation duration (Rayner, 1998). For each word of the text, these two variables and the word predictibility in the context of the text sentences, are the core variables of the model (Rayner et al., 2004). From these input data, the model will provide for each word, the probability to be fixed and the fixation duration. But to estimate all the parameters of the model from known scanpaths during reading, it is necessary to assume that they come from the same reading strategy in the sense of the Carver's classification. Let us illustrate this idea for two configuration parameters of the EZ reader model, the minimum duration of the “familarity check” and the “systematic error”. The first one is the fixed part of the estimation of the duration of the “familarity check”. The variable part is indexed on the frequency and the predictibility of the words. It is expected that this minimal duration should depend on the level of comprehension depth induced by readers’ intentions, and also their linguistic expertise or their reading skills (Blythe & Joseph, 2011). This is also the case for the systematic error parameters determining the probability for eye movements to undershoot or overshoot their intended targets. As a consequence, scanpath segmentations obtained from our approach leading to piecewise homogeneous statistical properties could be used for calibrating specific parameters in EZ-reader. Both models could then by coupled, so that the HMSC model could trigger parameter switches in EZ-reader when changing reading strategy.

Our approach considers mixed models to characterize the effect of eye-movement- or semantic-related covariates on phases. This in an improvement compared to Simola et al. (2008); however from a methodological point of view, a possible weakness of this work is due to the separate use of HMSC models on the one hand, and GLMM modelling of the effect of phases including individual effects on the other hand. Indeed, individual variability was highlighted in state dynamics and emission distributions, so ideally, this would have to be accounted for in parameter estimation by maximum likelihood. Inference of state-based parameters based on MAP restoration is likely to cause biases since uncertainty on the state values is not accounted for in *post hoc* analyses. In the same spirit, including effects of covariates (e.g., distance to trigger words or type of text in transition probabilities) could be integrated in HMSC models directly, by using GLMMs instead of plain distributions in the transition matrix, state sojourn duration and emission distributions. Inference in such models was studied in particular cases by Altman (2007). Another possibility to account for individual heterogeneity is to resort to mixtures of HSMMs, but this would lead to some significant increase in the number of model parameters, whereas mixed effect models have tied parameters.

Although we developed some methodology to connect reading phases or strategies to text semantics, the latter is here summarized to two trigger words. However, the semantic progression in the different texts is not homogeneous: in some of them, relevant information with respect to target topic is brought linearly while in some others, it is brought abruptly in one or two major steps. Some text clustering could reveal helpful to investigate connections between the dynamics of accumulated information, as quantified by FastText and the use of particular strategies.

Regarding joint analysis of eye movements and EEGs, our approach opens new avenues to characterize which brain connections are activated or not in each reading strategy. From a general point of view, performing analyses based on EEGs only is particularly difficult in free reading tasks. This is partly due to the high level of noise, related to both inter- and intra-individual variability. Another source of difficulty is the lack of synchronization of different individuals reading the same text using different strategies.

Here, eye-movement based segmentation acts as a medium to resynchronize portions of scanpaths coming from different individuals and trials. The reason for this is that segments of the same nature, with definite dates of beginning and ending, associated with synchronized EEG signals, may be assumed to have common features due to inherent homogeneity in a given phase. Performing within-segment analyses is thus expected to reduce heterogeneity and to facilitate identification of specific EEG patterns characterizing cognitive steps leading to decisions.

Acknowledgements

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir.
The authors acknowledge useful contributions by B. Lemaire regarding eye-movement and natural language processing. They acknowledge helpful remarks by S. Achard to improve the discussion.

References

Altman, R. M. (2007). Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association*, *102*(477), 201–210.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1–48.

Blythe, H., & Joseph, H. (2011). Children's Eye Movements during Reading. In *The Oxford Handbook of Eye Movements*. Liversedge, S., Gilchrist, I., & Everling, S. (Eds.). Oxford University Press, UK.

Cappé, O., Moulines, E., & Rydén, T. (2007). An Information-Theoretic Perspective on Order Estimation. In *Inference in Hidden Markov Models* (pp. 565–602). Springer Series in Statistics. Springer, New York, NY.

Carver, R. P. (1990). Reading rate: A review of research and theory. In *Reading rate: A review of research and theory.* Academic Press.

Carver, R. P. (1992). Reading Rate: Theory, Research, and Practical Implications. *Journal of Reading*, *36*(2), 84–95.

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. H. (2020). Eye movement analysis with switching hidden Markov models. *Behavior Research Methods*, *52*(3), 1026–1043.

Clifton, C., Staub, A., & Rayner, K. (2007). Chapter 15 - Eye movements in reading words and sentences. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye Movements* (pp. 341–371). Elsevier, Oxford, UK.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Delattre, M., Lavielle, M., & Poursat, M.-A. (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics*, *8*(1), 456–475.

Durand, J. B., & Guédon, Y. (2016). Localizing the latent structure canonical uncertainty: entropy profiles for hidden Markov models. *Statistics and Computing*, 26 (1–2), 549–567.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813.

Farid, M., & Grainger, J. (1996). How initial fixation position influences visual word recognition: A comparison of French and Arabic. *Brain and Language*, 53(3), 351–368.

Frey, A., Ionescu, G., Lemaire, B., López-Orozco, F., Baccino, T., & Guérin-Dugué, A. (2013). Decision-making in information seeking on texts: An Eye-Fixation-Related Potentials investigation. *Frontiers in Systems Neuroscience*, 7, 39.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London Academic Press, UK.

Hadfield, J. D. (2010). MCMCglmm: MCMC Methods for Multi-Response GLMMs in R. *Journal of Statistical Software*, *33*(2), 1–22.

Hollenstein, N., Rotsztejn, J., Troendle, M., Pedroni, A., Zhang, C., & Langer, N. (2018). Data descriptor: ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific Data*, *5*, 1–13.

Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). *Bag of Tricks for Efficient Text Classification*. In Lapata, M., Blunsom, P., & Koller, A. (Eds), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Volume 2, Short Paper, 427-431.

Kim, J., Hyub, H., Yoon, S. Z., Choi, H.-J., Kim, K. M., & Park, S.-H. (2014). Analysis of EEG to quantify depth of anesthesia using Hidden Markov Model. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4575–4578. IEEE

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, *29*(2), 169–195.

Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: An analysis based on the SWIFT model of eye-movement control. *Vision Research*, *49*(3), 322–336.

Obermaier, B., Guger, C., Neuper, C., & Pfurtscheller, G. (2001). Hidden Markov models for online classification of single trial EEG data. *Pattern Recognition Letters*, *22*(12), 1299–1309.

Pradal, C., Fournier, C., Valduriez, P., & Cohen-Boulakia, S. (2015). OpenAlea: Scientific Workflows Combining Data Analysis and Simulation. In Gupta, A., & Rathbun, S. (Eds), *Proceedings of the 27th International Conference on Scientific and Statistical Database Management*. Association for Computing Machinery, New York, NY.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. In *Quarterly journal of experimental psychology* , 62(8), 1457–1506.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the E-Z Reader model. *Journal of Experimental Psychology. Human Perception and Performance*, 30(4), 720–732.

Rayner, K., & Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, 80(1), 4–9.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7(1), 4–22.

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, 39(26), 4403–4411.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: comparisons to other models. *The Behavioral and Brain Sciences*, 26(4), 445–526.

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, 7(1), 34–55.

Reilly, R. G., & Radach, R. (2002). Glenmore: an interactive activation model of eye movement control in reading. In Wang, L. (Eds) *Proceedings of the 9th International Conference on Neural Information Processing, 2002 ICONIP '02., 3*, 1194–1200 vol.3. Nanyang Technological University, School of Electrical and Electronic Engineering, Singapore.

Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9(4), 237–251.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(4), 583–616.

Venables, W.N., & Ripley, B.D. (2002). Random and Mixed Effects. In *Modern applied statistics with S* (271–300). Springer, New York, NY.

Yu, S. Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, 174(2), 215–243.

Appendix

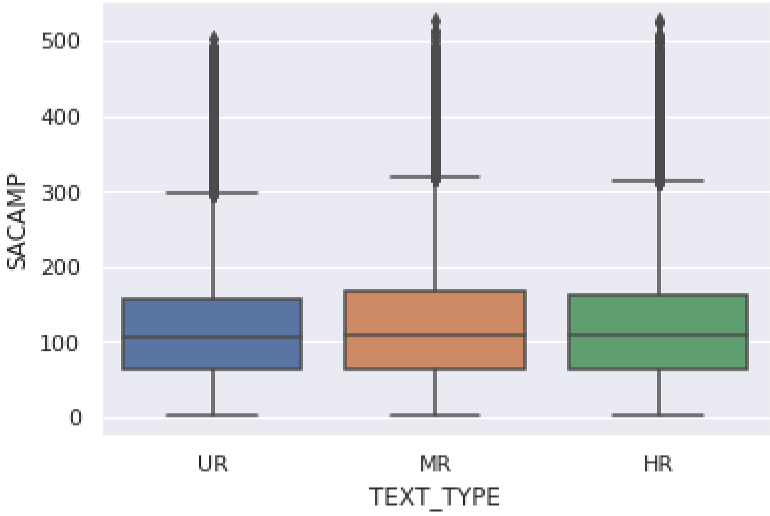

Figure S1. Boxplot of Saccade Amplitude (SACAMP) for each text type.

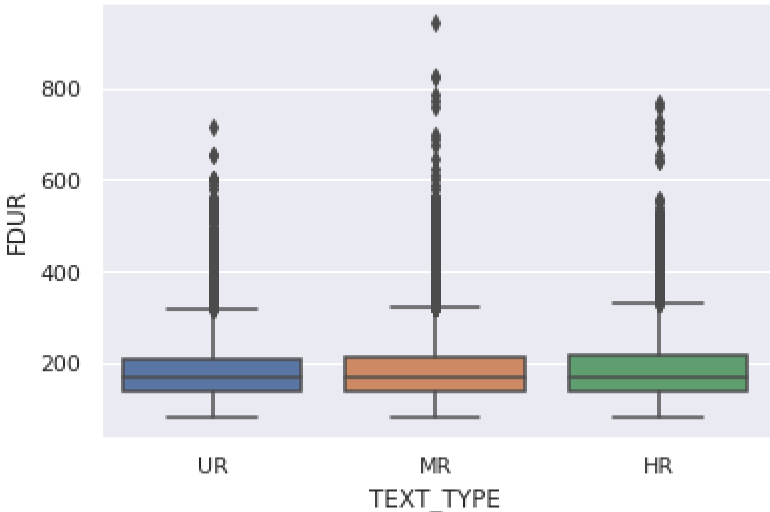

Figure S2. Boxplot of Fixation Duration (FDUR) for each text type.

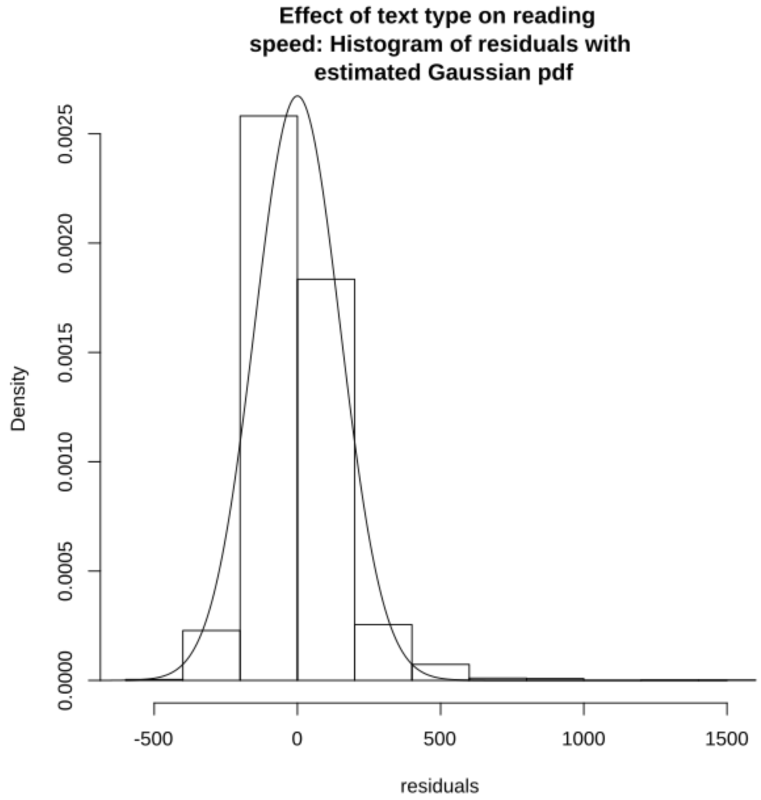

Figure S3. Histogram and estimated normal pdf for the LMMs residuals in modelling the effect of text type on reading speed.

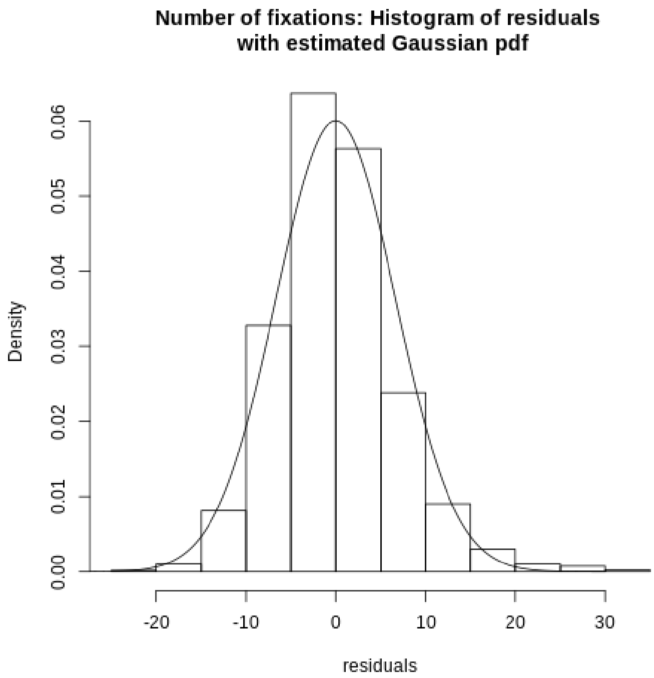

Figure S4. Histogram and estimated normal pdf for the LMMs residuals in modelling the effect of text type on the number of fixations per scanpath.

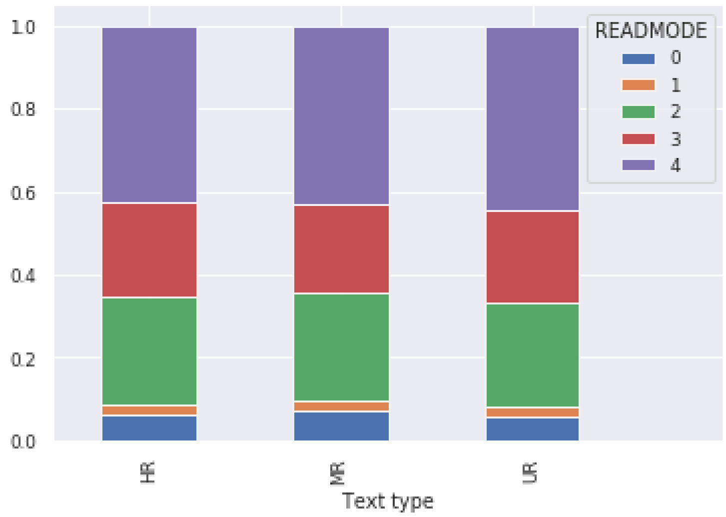

Figure S5. Sample distribution of Read Mode for each text type. Readmode 0 corresponds to Bwd-, 1 to Bwd, 2 to Rfx, 3 to Fwd and 4 to Fwd+.

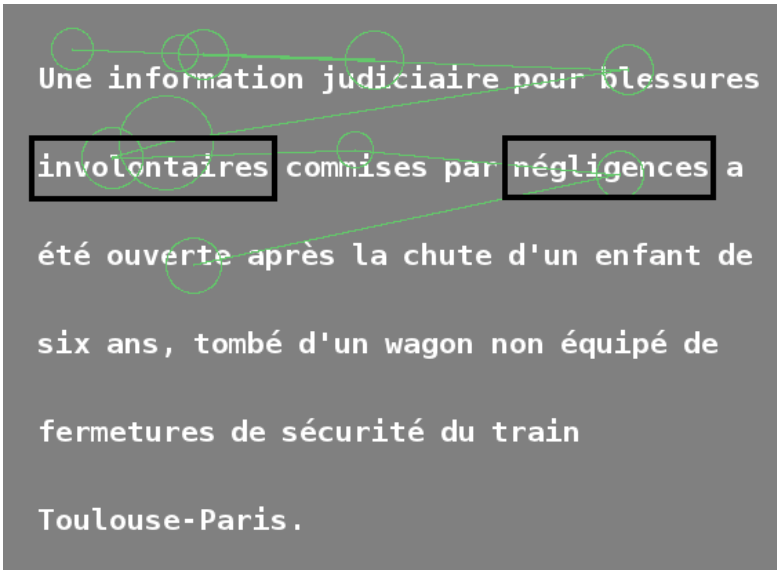

Figure S6. Scanpath of some UR text with phase restoration. Target topic is “Contemporary art”. Phase 2 (speed reading) is in green. Translation: “Judicial investigation for accidental injury due to negligence was opened after a six-year-old boy fell from a train joining Paris from Toulouse and lacking of secure locking mechanism”. The words framed in black are the farthermost to target topic.

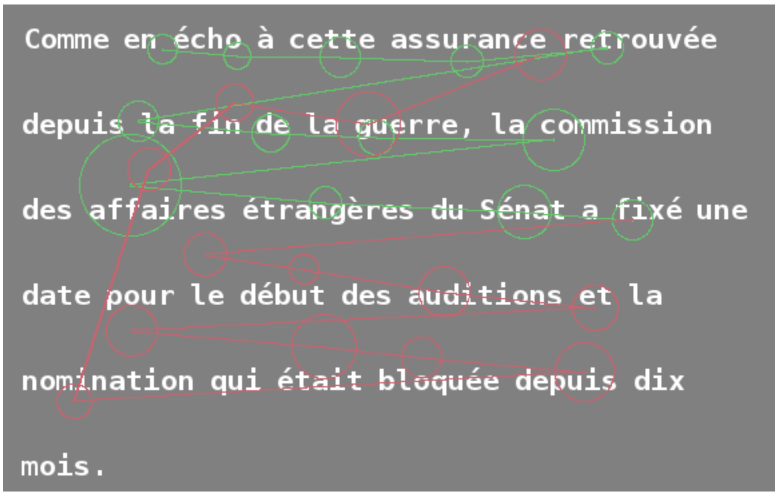

Figure S7. Scanpath of some MR text with phase restoration. Target topic is “Conflict in Irak”. Phase 2 (speed reading) is in green and Phase 3 (slow confirmation) in red. Translation: “As if echoing the self-confidence regained since the end of the war, the Foreign Affairs committee of Senate set up a date to begin auditions and to proceed to nomination, which had been blocked for ten months”.

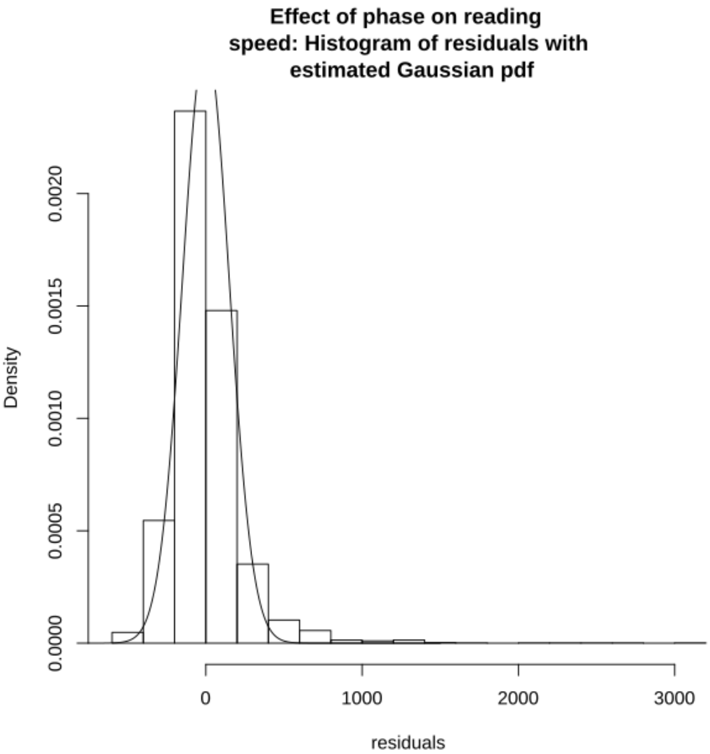

Figure S8. Histogram and estimated normal pdf for the LMMs residuals in modelling the effect of phase and text type on the number of fixations per scanpath.

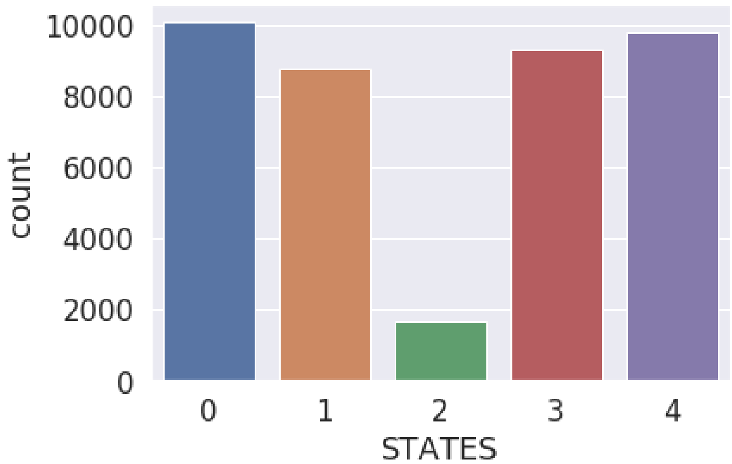

Figure S9. State sample distribution.

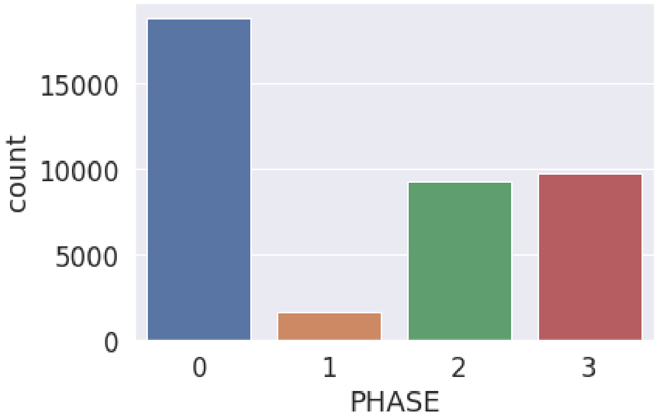

Figure S10. Phase sample distribution. Phase 0 (normal reading) is in blue, Phase 1 (information search) in orange, Phase 2 (speed reading) in green and Phase 3 (slow confirmation) in red.

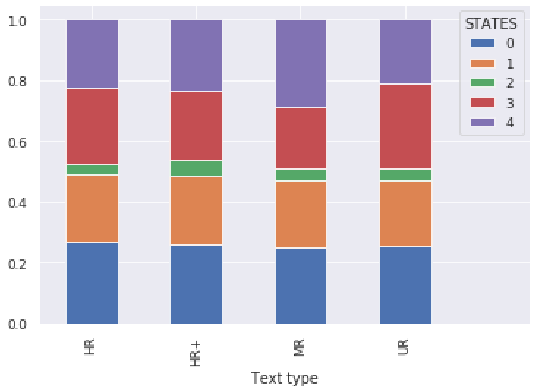

Figure S11. State sample distribution per extended text type.

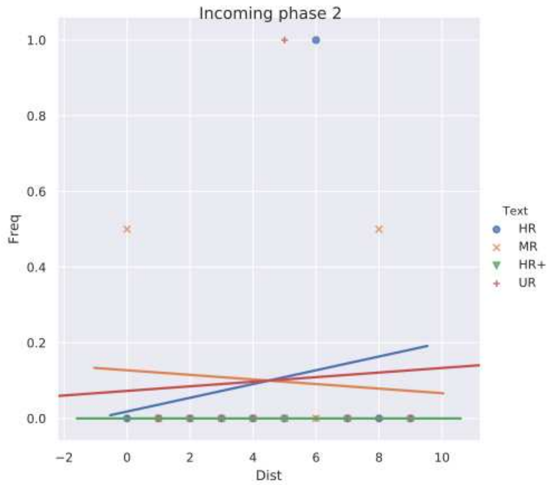

Figure S12. Relationship between distance to trigger words and frequencies of transitions arriving into phase 2 (speed reading).

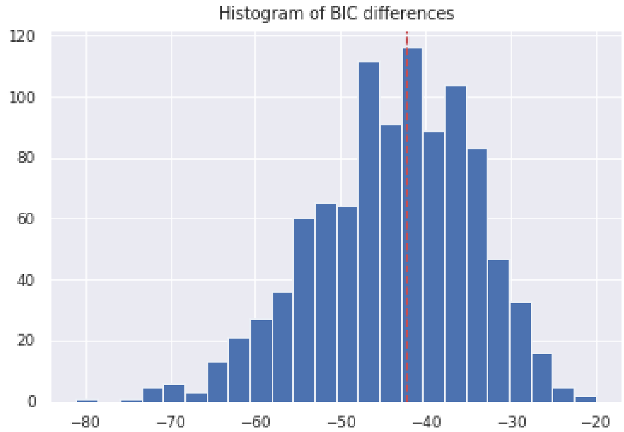

Figure S13. Histogram of BIC differences with null model obtained by constrained permutations of phases under linear mixed models. The BIC difference corresponding to true data is represented by a dotted vertical line.

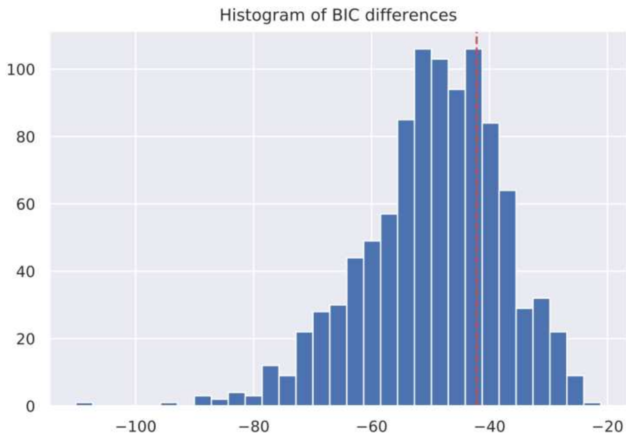

Figure S14. Histogram of BIC differences with null model obtained by free permutations of phases under linear mixed models. The BIC difference corresponding to true data is represented by a dotted vertical line.

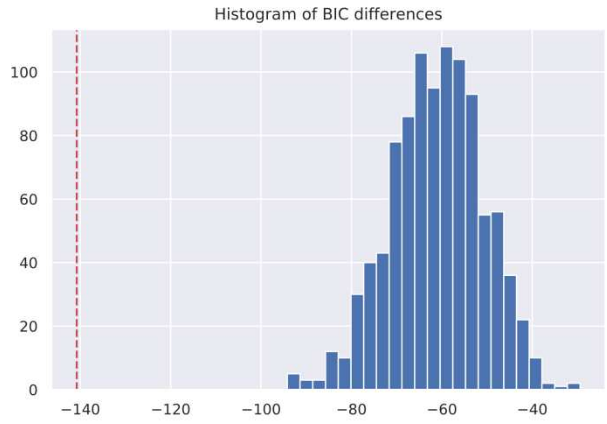

Figure S15. Histogram of BIC differences with null model obtained by constrained permutations of phases under generalized linear mixed models. The BIC difference corresponding to true data is represented by a dotted vertical line.

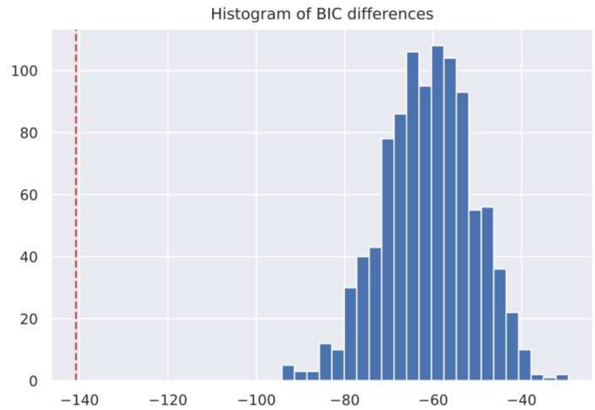

Figure S16. Histogram of BIC differences with null model obtained by free permutations of phases under generalized linear mixed models. The BIC difference corresponding to true data is represented by a dotted vertical line.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe - Montbonnot
38334 Saint Ismier Cedex France

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr
ISSN 0249-6399