



HAL
open science

Ocular strategies analyses during reading through a text comprehension model

Etienne Langevin, Anne Guérin-Dugué, Aline Frey, Benoît Lemaire

► To cite this version:

Etienne Langevin, Anne Guérin-Dugué, Aline Frey, Benoît Lemaire. Ocular strategies analyses during reading through a text comprehension model. ECEM 2019 - 20th European Conference on Eye Movements, Aug 2019, Alicante (SPAIN), Spain. ⟨hal-04940077⟩

HAL Id: hal-04940077

<https://hal.science/hal-04940077v1>

Submitted on 12 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



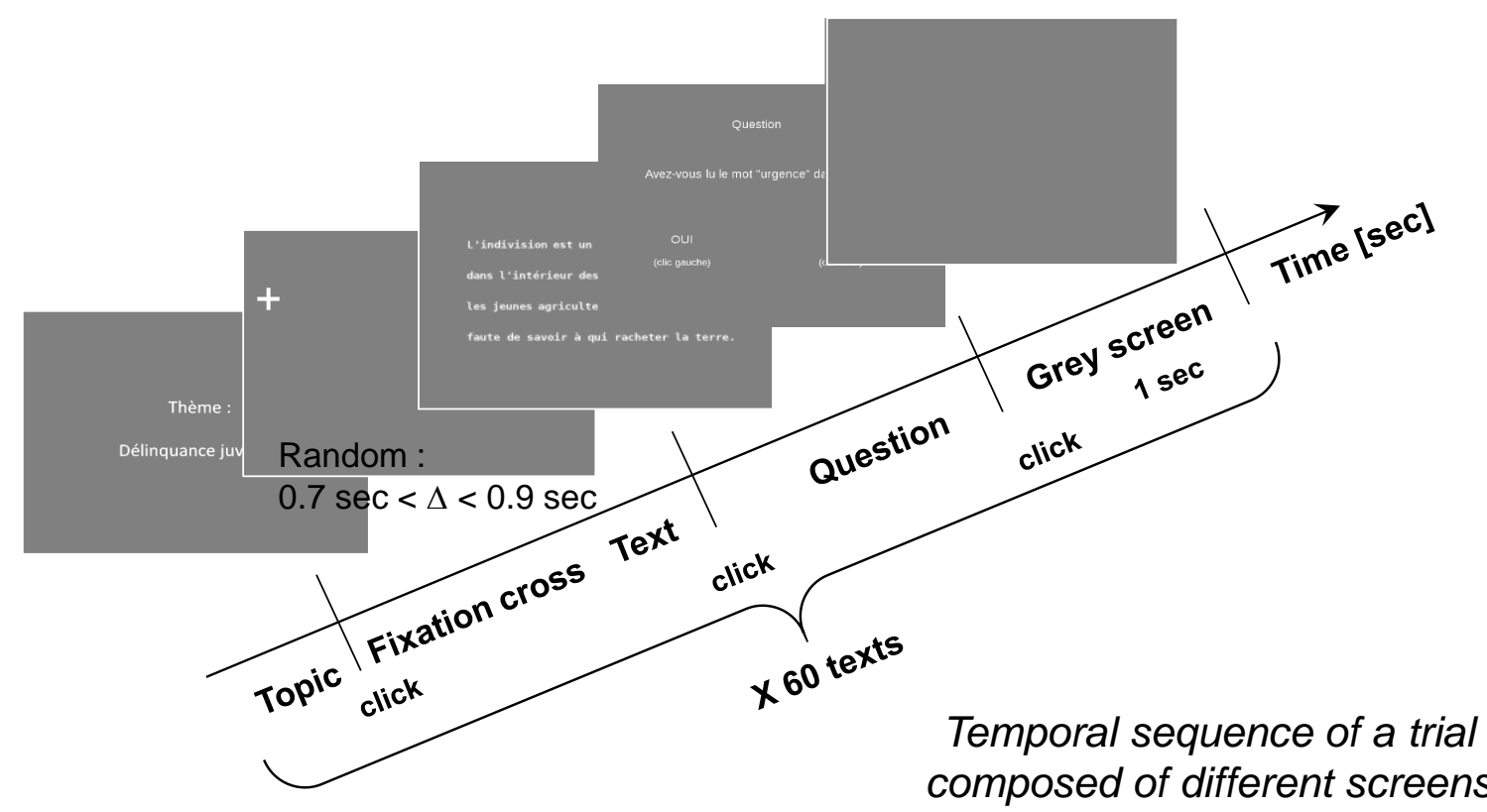
HAL Authorization

Objectives

- Strategies controlling reading saccades throughout a text depend on many factors including semantic ones.
- The aim of this work is to simulate the time course of semantic integration during reading, using:
 - construction-integration model (Kintsch, 1998)
 - semantic memory simulated with Latent Semantic Analysis (Landauer et al., 2007)
- The simulation of the progressive semantic integration of each fixated word is then compared to:
 - a fictional reader processing all words
 - a fictional random reader processing words located at random position

1. Experiment

- 20 healthy adults silently read texts (10 men and 10 women; age range: 20-39; M=27.2; SD=5).
- 120 texts from a corpus of all the articles published in the newspaper "Le Monde" in 1999 :
 - Number of sentences: M = 1.8; SE = 0.13;
 - Number of words: M = 30.1; SE = 0.65;
 - For each word, number of characters : M = 5.34; SE = 0.72;
- Task : read and decide whether to retain or reject the text depending on the presented topic
- Eye-tracker
 - Eyelink 1000 (SR Research)
 - Sampling at 1000 Hz

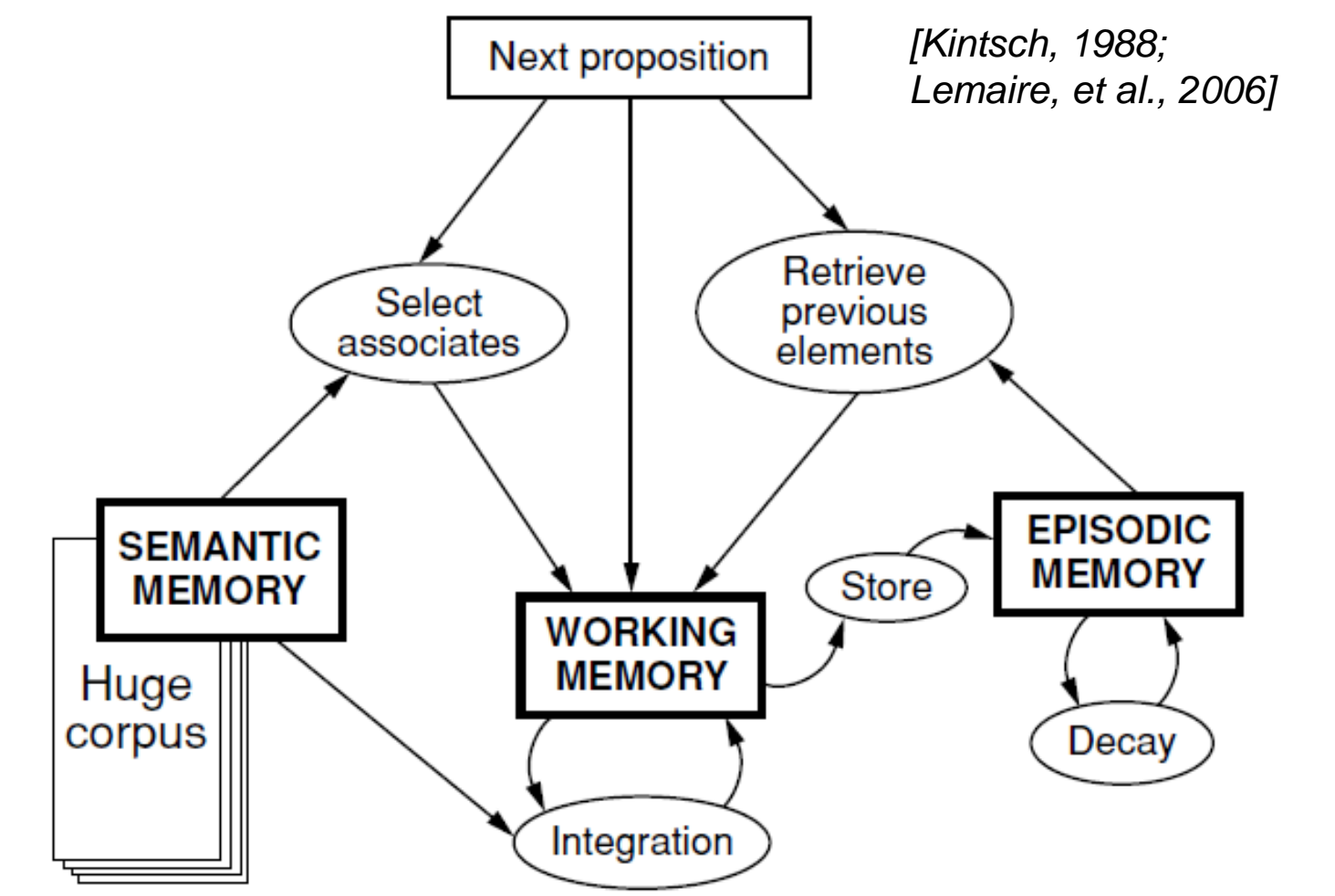


The study of the dynamics of the rings which surround all giant planets orbiting around the Sun allows astronomers to understand the process by which solar system moons were created.

Example of the text about the planet observation, and its English translation.

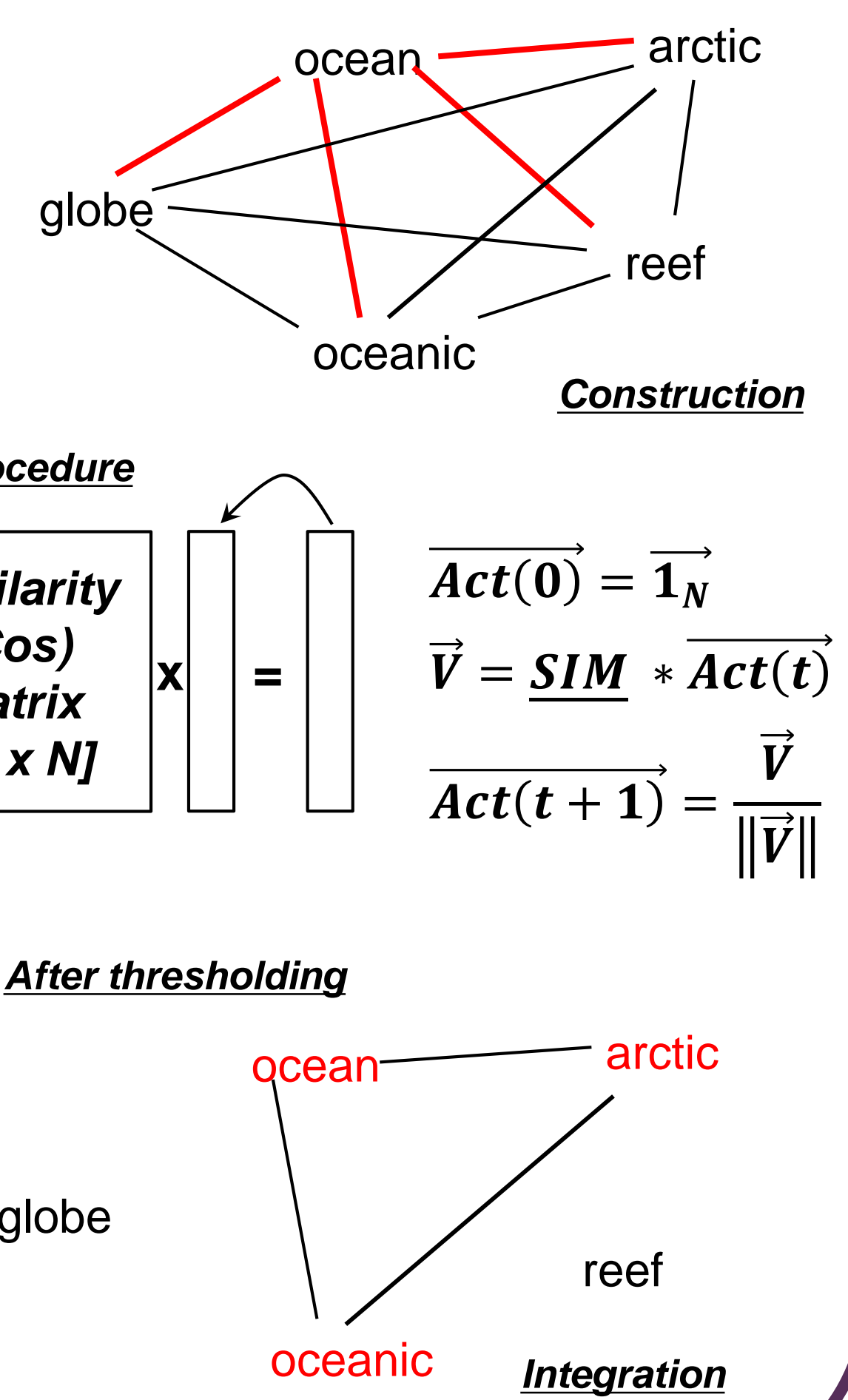
3. Comprehension model

- Input :
 - Proposition = one word at a time
- Semantic memory
 - Corpus : Newspaper "Le Monde" (1999)
 - Latent Semantic Analysis
 - 300-dimensional vector
- Output : Working memory
 - Semantic representation of the text in the working memory
 - Subset of weighted words
 - Each word : 300-dim vector, one weight



4. Mechanism: Construction - Integration

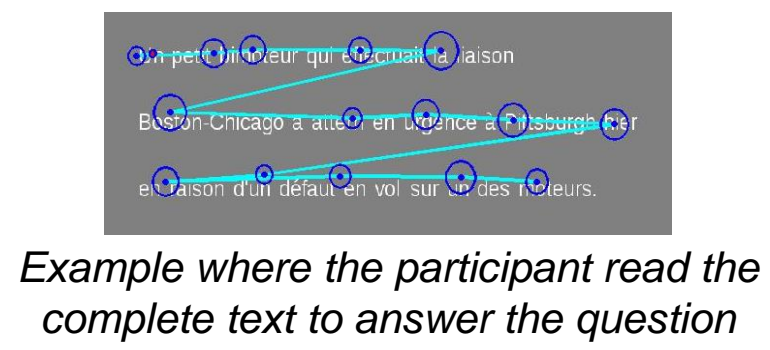
- Text : Dans l'océan il y a beaucoup de poissons. La mer aussi est un endroit accueillant pour la vie. (In the ocean there are many fishes. The sea is also a welcoming place for life)
- First input word: ocean
- Construction with:
 - Input word
 - Its neighbors
 - Working memory
 - Episodic memory
- Construction: Neighbors of océan (ocean)
 - Word: arctique (arctic) .. cosinus value: 0.704
 - Word: récif (reef) .. cosinus value: 0.682
 - Word: océanique (oceanic) .. cosinus value: 0.682
 - Word: globe (globe) .. cosinus value: 0.631
- Integration: Iterative mechanism
 - Iterative extraction of words highly connected
 - Activities after thresholding
- For the next input word: poissons (fishes)
 - ocean, arctic, oceanic: in the working memory
 - globe, reef: in the episodic memory



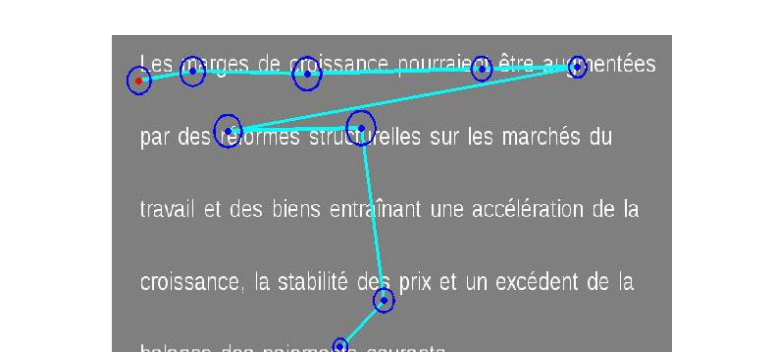
2. Behavioral results

- Rate of correct answer: 97.8% (SE = 0.49)
- Reading time: 3.17 s (SE = 0.25)
- Number of fixations: 12.31 (SE = 0.83)
- Fixation duration: 185.53 ms (SE = 6.08)
- Reading speed: 488.15 wpm (SE = 25.27)
- % of fixations depending on the fixated words
 - Long progression (#1): 13.7% (SE = 1.0)
 - Short progression (#2): 24.5% (SE = 1.6)
 - Refixation (#3): 2.6% (SE = 0.5%)
 - Re Reading fixation (#4) : 4.2% (SE = 0.7%)

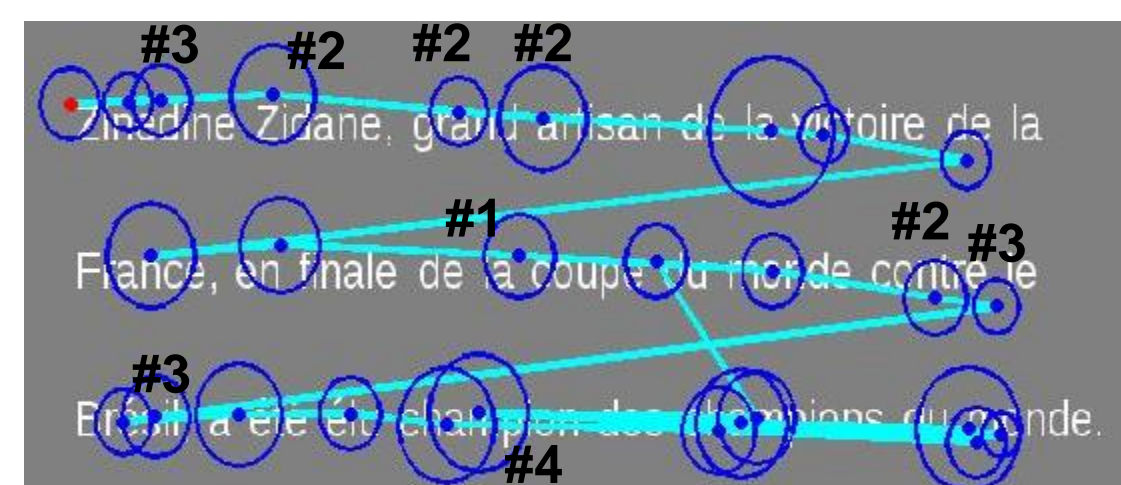
Examples of scanpath



Example where the participant read the complete text to answer the question



Example where the participant didn't read the complete text to answer the question



Le traitement des données est

4 char. 8 char.

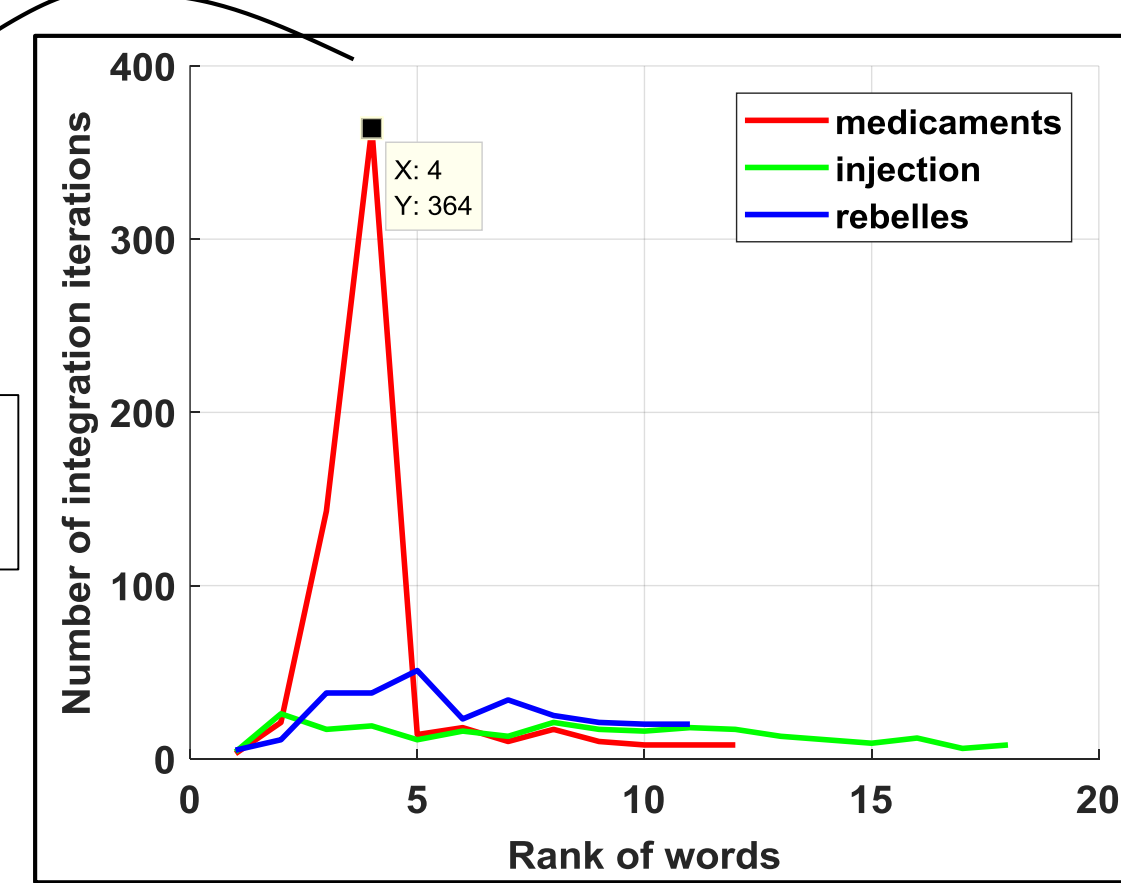
5. Iterative integration

- Example : Text "médicaments" (medications)
 - Input : each word after lemmatization
 - Stop words suppression
 - #w (w = 1, 2, 3, ...) : input words

#1 Les laboratoires ont ainsi réalisé qu'au premier coup dur, leur existence même était menacée. Les médicaments nouveaux n'étaient plus assez nombreux pour compenser la chute de leurs brevets dans le domaine public.

#2 #3 #4 "coup": polysemous word (shot, hit, stunt)

#5 The laboratories realized that at the first hard time, their existence was threatened....

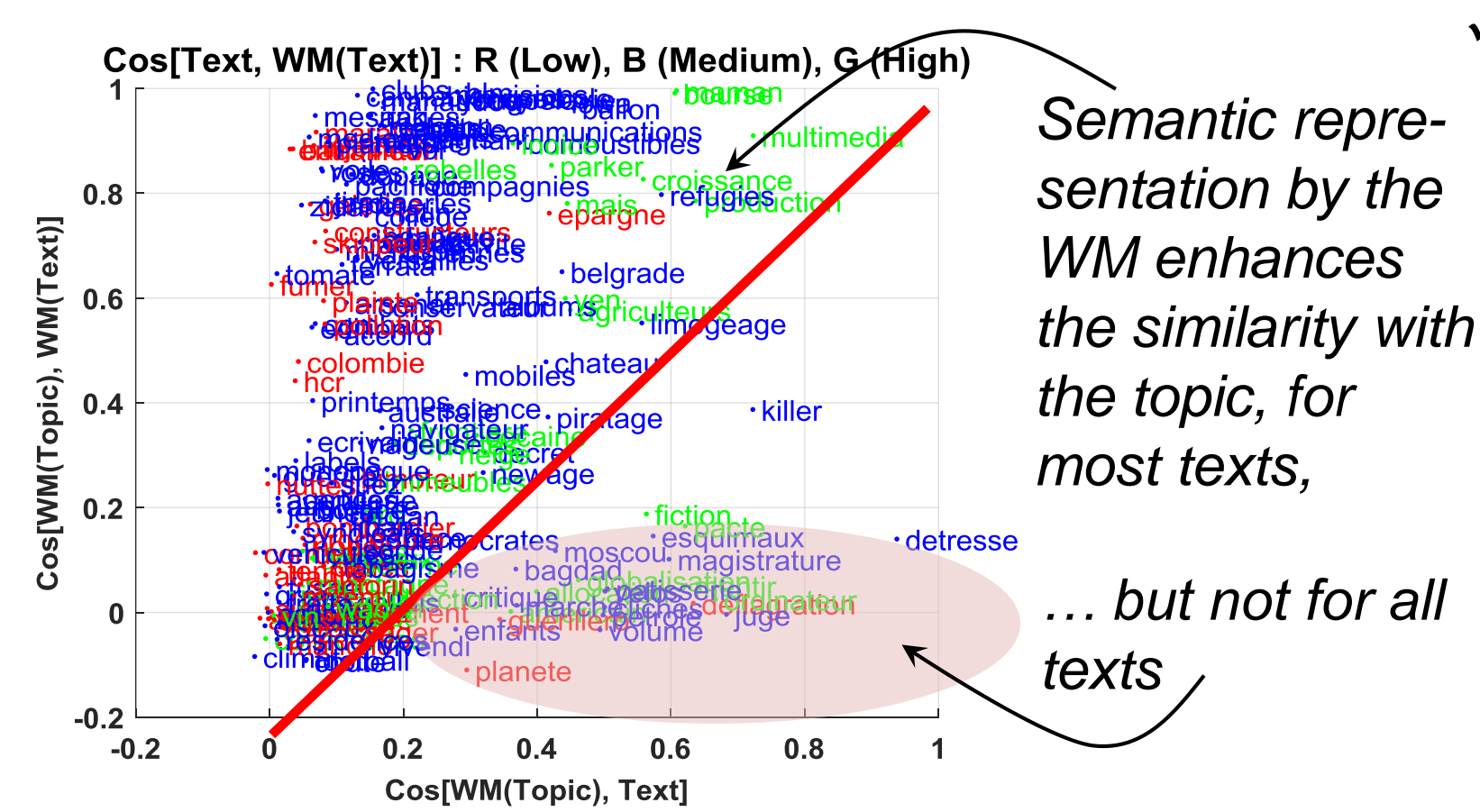


- Locally higher number of iterations
 - for polysemous words with another context
- Globally homogeneous number of iterations

Model	All words	Fixated words
Number of iterations	240.6	180.7 (SE = 7.3)

6. Similarity with the topic

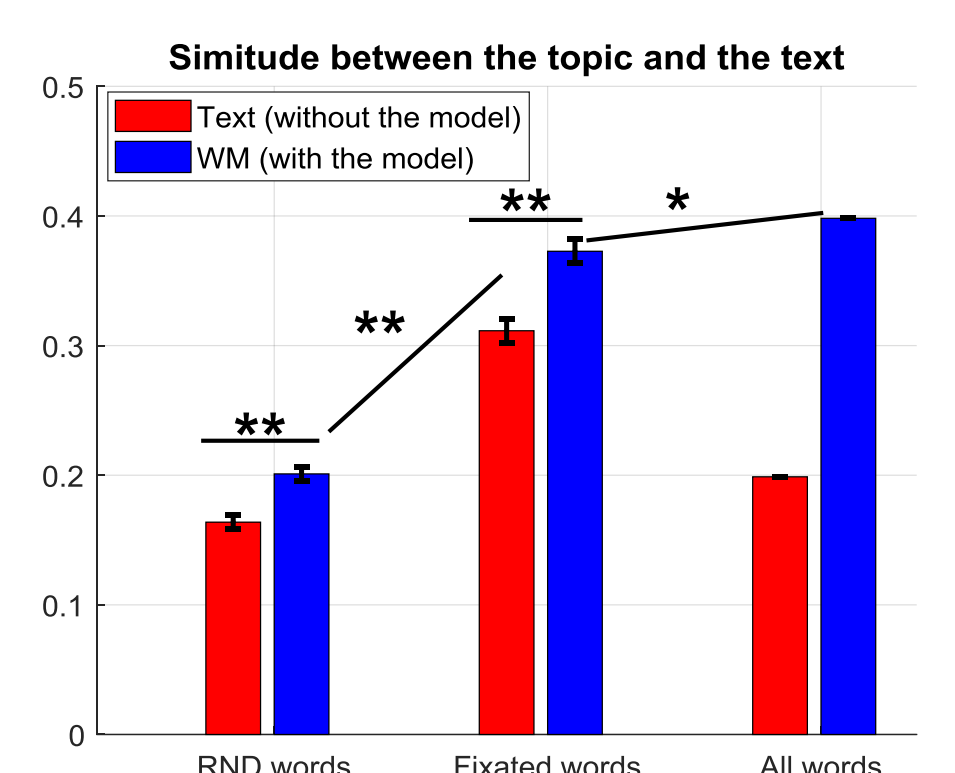
- Joint distribution of the similarity between the text and the topic
 - With the model vs without the model



All words, Model	Without	With
Cos[WM(Topic), Text]	0.2	0.4

7. Assessment

- Methodology
 - All words with vs without the model
 - Fixated words with vs without the model
 - Random words with vs without the model



Model	All words	Fixated words
Cos[Text, WM(Text)]	0.28	0.37 (SE=0.01)

Conclusions

- On the semantic text representation by the working memory
 - Compact representation on less than 15 items (WM size);
 - Higher similarities between the input words and their semantic representation in WM for a real readers (participants) than for a fictional reader processing all words;
 - But, the model sometimes stuck on highly specific wrong meaning.
- Globally higher similarities between the topic and the working memory
- Human-like fixations result in a slight loss of meaning compared to full fixations
 - but with less iterations for a faster reading
- Random fixations result in a large loss of meaning compared to human-like fixations
 - More optimal reading strategy