Identifying Key Parts of Speech That Differentiate Human and AI-Generated Text Using Eye-Tracking Data

Sonain Jamil
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:20%

Kasem Amnuayrotchanachinda
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:20%

Muhammad Turab Muslim Bajeer
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:15%

Tapio Meriläinen
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:20%

Veikka Vetola
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:10%

Antti Ikonen
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:10%

Manish Saha
University of Eastern Finland (UEF)
Joensuu, Finland
Contribution:5%

Abstract

In this work, we analyze the parts of speech which contribute to the evaluation of text written by human or generated by artificial intelligence (AI) e.g. ChatGPT, we use Tobii Pro eye tracker to record data of the participants. We extract key features such as total fixation count, fixation duration, scanpath to analyze the data. We perform analysis using velocity threshold identification algorithm with two different thresholds i.e. 30 and 100. From the experiment, we observe that the participants focus on nouns in case of human written text, whereas the participants focus on adjectives in case of AI generated text. Our analysis is quite consistent with our observations as AI tools try to make text wonderful and complicated.

Keywords

Eye tracking, fixations, noun, adjective, AI-text

ACM Reference Format:
Sonain Jamil, Kasem Amnuayrotchanachinda, Muhammad Turab Muslim Bajeer, Tapio Meriläinen, Veikka Vetola, Antti Ikonen, and Manish Saha. 2024. Identifying Key Parts of Speech That Differentiate Human and AI-Generated Text Using Eye-Tracking Data. In *Proceedings of Eye tracking project (Eye tracking)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Artificial intelligence (AI) generated text is increasing at rapid pace in the current era. There are several key parts of speech which correspond the most to declare text as AI generated. In this study, we perform an eye tracking experiment using five participants to analyze the part of speech which correspond the most to declare the text as human or AI written text. Initially, we create our data containing ten different stimuli, in each stimuli, there were two text one from books and other from AI i.e. ChatGPT. The participants were asked to identify which text belong to AI and which text is by human. We recorded data of five participants, one participant performed experiment twice once with glasses and once without glasses so there were six recording however, we had to exclude the data of this participant as the eye tracker was unable to properly capture the data as in both cases the fixation during was zero indicating the data as outlier. So we analyze the data of four participants which were our group members and from those four participants, three were normal vision and one was wearing glasses. We also excluded six questions and analyzed only four in the final analysis. We use velocity threshold identification (I-VT) algorithm with two different thresholds for our analysis.

2 Related Work

In study [5], the authors have studied the impact of form, content, and style of text on individuals. For this, they conducted an experiment with different individuals reading various types of texts, including children's stories, random and word-generated texts. These texts were categorized in different categories based on their complexity and coherence. The results showed that the average magnetization of fixation configuration correlates with their complexity. Also, their study found that different texts may induce different cohesive reading activities. The authors emphasized that higher complexity in texts tends to lead to longer fixations, requiring greater cognitive engagement from readers. This suggests that the coherence and structure of a text can significantly influence how individuals process and retain information.

In study [4], the authors used eye-tracking to understand how individuals assess the authenticity of online information, such as fake news. They found that digital knowledge has a major impact

Eye tracking, December 13, 2024, UEF

ACM ISBN XXX-X-XXXX-XXXX-X/XX/XX
https://doi.org/XXXXXXX.XXXXXXX

on how users interact with social media and webpages to assess their credibility. They suggest that paying attention to meta-data is an effective technique to check the credibility, as it is positively associated with attention devoted to meta-data. The study also indicated that individuals with higher digital literacy are more likely to engage with subtle indicators of credibility, such as the source of information and the publication date. This highlights the importance of critical thinking and the ability to evaluate meta-data in online environments, suggesting that strengthening digital literacy could help reduce the spread of misinformation.

In study [3], the author explores how eye-tracking reading times can reflect sentence complexity and be utilized for tasks like readability assessment. It uses machine learning models to predict reading times and demonstrates their effectiveness in modeling text difficulty. The research highlights the interplay between linguistic features and cognitive load, emphasizing how these insights can aid in optimizing content for diverse audiences. Applications include improving educational materials, enhancing user experience in digital interfaces, and refining natural language processing algorithms.

In study [1], the author systematically reviews how eye-tracking technology assesses text comprehension by analyzing cognitive effort during reading. It examines factors such as media type, task perspective, and instructional strategies that affect comprehension. The findings suggest that eye-tracking can be a reliable indicator of mental engagement, offering insights into how readers process different types of texts. Additionally, the article discusses implications for designing instructional materials and evaluating digital reading tools, highlighting its potential for advancing both educational practices and cognitive research.

3 Selected Fixation Algorithm

We selected I-VT algorithm [2] for our analysis. In this algorithm, the initially an empty list of fixations is created, after that for all the gaze points in the eye tracking data, velocity is computed between two consecutive points using equation 1.

$$Velocity(i) = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i} \tag{1}$$

where x_{i+1} and x_i are x coordinates of two consecutive points, y_{i+1} and y_i are y coordinates of two consecutive points and t_{i+1} and t_i is current time and previous time, respectively.

If the calculated velocity is less than velocity threshold then this point is classified as fixation point otherwise the point is classified as saccade. After this, the consecutive fixations are clustered and for each group centroid points are calculated for fixations. At the end these fixations are stored in a list. The psuedo code of this algorithm is presented in Algorithm 1.

4 Data Creation

We created dataset by finding definitions from different books about various branches of science, such as biology, computer science, mathematics and neuroscience, and definitions of same topic by ChatGPT, we saved the dataset in the form of csv file to make it easier to use in the experiment.

Algorithm 1 Velocity-Threshold Identification (I-VT)

Inputs:
$GazePoints = \{(x_i, y_i, t_i)\}_{i=1}^{N}$: List of gaze data points with coordinates (x_i, y_i) and timestamps t_i
$VelocityThreshold$: Maximum allowable velocity to classify a point as part of a fixation
Outputs:
$Fixations$: List of detected fixation clusters with their centroids and timestamps
Algorithm:
Initialize an empty list for fixations: $Fixations \leftarrow \emptyset$ ▷ This will store all identified fixation clusters
for $i = 1$ to $N - 1$ **do** ▷ Iterate through all gaze points except the last one
  Compute velocity between consecutive points:

$$Velocity(i) = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i}$$

  ▷ The velocity is calculated using the Euclidean distance divided by the time difference
  if $Velocity(i) \leq VelocityThreshold$ **then** ▷ Check if the computed velocity is below the fixation threshold
    Classify p_i as a fixation point ▷ Points with low velocity are part of a fixation
  else
    Classify p_i as a saccade point ▷ Points with high velocity are considered saccades
  end if
end for
Group consecutive fixation points into clusters ▷ Group fixation points that are close together in time and space
for each fixation group **do** ▷ Process each group of consecutive fixation points
  Compute the centroid of the fixation:

$$CentroidX = \frac{\sum x_i \in Group}{|Group|}, \quad CentroidY = \frac{\sum y_i \in Group}{|Group|}$$

  ▷ The centroid is the average position of all points in the fixation group
  Record fixation details:

$$Fixations \leftarrow Fixations \cup$$
$$\{StartTime : \min(Group.t), EndTime : \max(Group.t),$$
$$Centroid : (CentroidX, CentroidY)\}$$

  ▷ Store the start time, end time, and centroid position for each fixation
end for
return $Fixations$ ▷ Output the list of all detected fixations

5 Participants and data recording

The experiment participants consisted of five students from University of Eastern Finland's eye-tracking course. Three of the participants are students in the international master's program Computational Color and Spectral Imaging (COSI), one participant is a Computer Science master's programmer student, and one participant acts as a teaching assistant in the course. None of the participants are native English speakers, which might have had an influence on the results. One of the participants performed the experiment twice: with and without eyeglasses.

The data collection was done by displaying stimuli on 1440×2560 resolution monitor, and Tobii pro eye-tracker with capture rate set to 240 Hz. For each participant, calibration was performed using Tobii Pro Lab software before starting data recording.

6 Experimental Design

In the experiment, we displayed the stimuli to the observers as shown in Figure 1, the experiment was conducted in eye tracking lab in semi controlled environment. The participants were asked to identify the text whether its human written or AI generated.

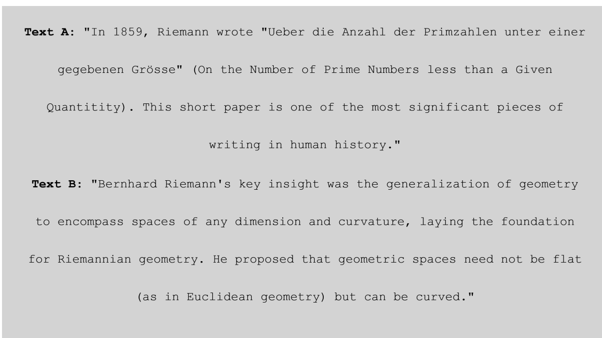

Figure 1: Example stimuli shown to the participants.

7 Experimental Results

In our experiment, we have area of interest (AOI) which are bounding boxes around each word in the sentence as shown in Figure 2, we can assign parts of speech to each word in the stimuli as shown in Figure 3.

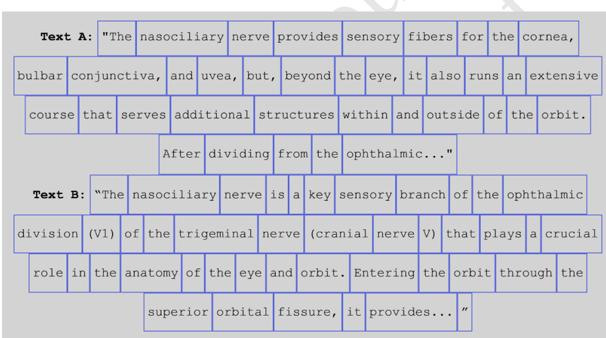

Figure 2: Stimuli showing AOI.

From these stimuli, we first use heatmap to visualize critical words on which all the participants focused. We show average heatmap of all participants for four stimuli in Figure 4, 5, 6 and 7, respectively.

In Figure 4, we can observe that the participants mainly focused on "1859" in Text A whereas in Text B the focus was on different

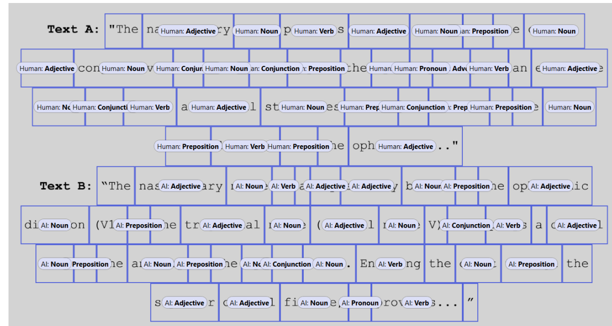

Figure 3: Stimuli showing AOI with parts of speech.

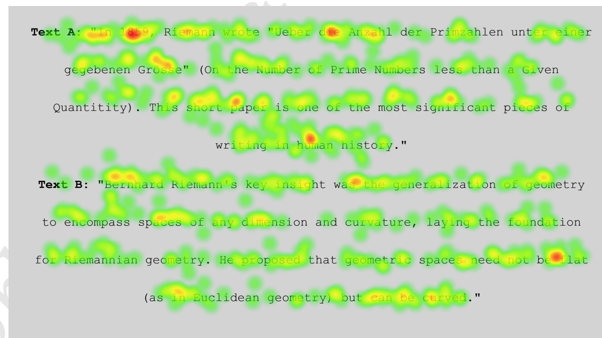

Figure 4: Average heatmap of all participants for question 1.

words such as "Bernhard", "spaces", "was", "be flat". All the participants rated Text A as human written and we can see that the part of speech which corresponds to their decision is a noun. In contrast, "be flat" is combination of verb and adjective making it a predicate which was focused more in case of AI text.

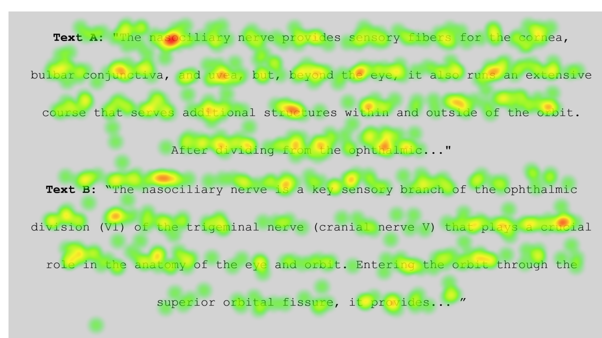

Figure 5: Average heatmap of all participants for question 3.

Similarly, in another stimuli in Figure 5, we can see that the participants focused on "nasociliary", "runs", and "structures" in case of Text A whereas in Text B the participants mainly focused on "crucial" which is an adjective and this word is very common in AI generated texts.

In the Figure 6, we can see that the participants focused on "vital" which is adjective and its synonym of crucial. This shows that

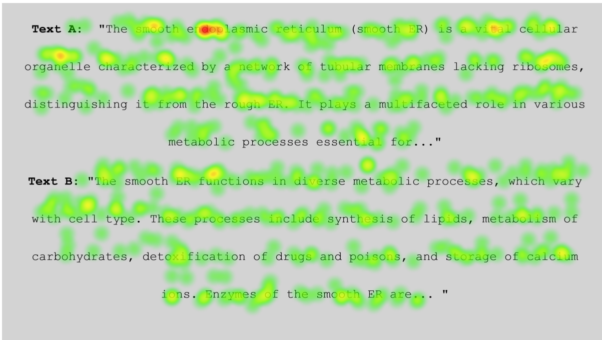

Figure 6: Average heatmap of all participants for question 4.

the participants mainly focused on adjectives to declare text as AI generated.

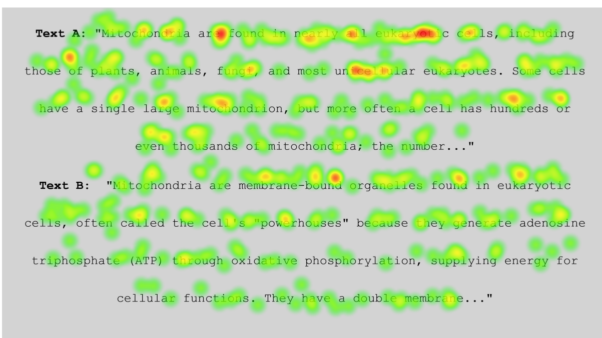

Figure 7: Average heatmap of all participants for question 5.

In Figure 7, we can observe that the participants focused on "membrane-bound" which is adjective to declare Text B as AI generated.

The trend is similar even if we visualize the scanpaths. We can take example scanpaths of each participant for question 1 as shown in Figure 8, 9, 10 and 11, respectively. From these scanpaths, it is clear that the participants focused on "1859" which is coherent with heatmap visualization. From these scanpaths we can also observe that all the participants spent significant time on "1859" before reading other words.

In our analysis, the most interesting feature was re-reading duration. We visualize re-reading duration for the parts of speech for each participant for Human and AI generated texts as shown in Figure 12 and 13, respectively.

Figure 12 shows that all the participants re-read noun and preposition to decide that the text is human written. From this we can conclude that noun and preposition contribute most to declare text as human written.

On the other hand, Figure 13 shows that the participants re-read the adjective and verb frequently which means that these two parts of the speech correspond most to declare text as AI generated.

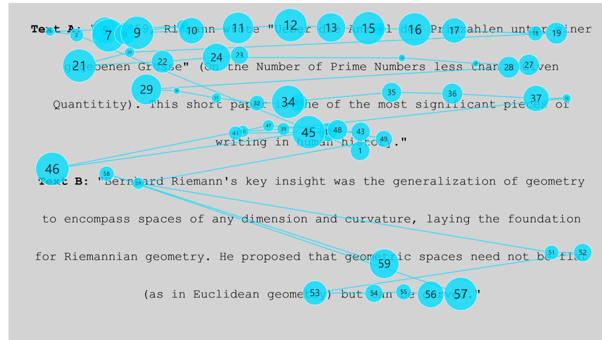

Figure 8: Scanpath of participant 1 for question 1.

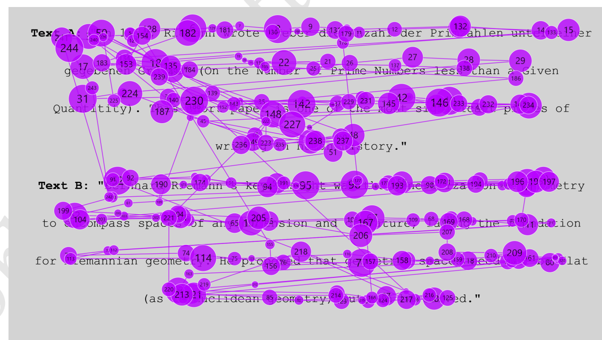

Figure 9: Scanpath of participant 2 for question 1.

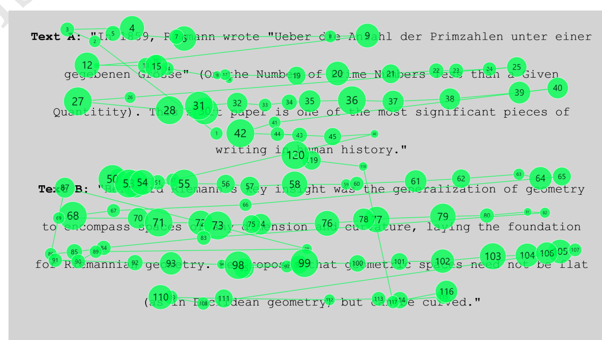

Figure 10: Scanpath of participant 3 for question 1.

7.1 Analysis of I-VT with two different thresholds

We also extract other features like average fixation duration, and number of fixations by changing threshold of the I-VT algorithm, we used two thresholds i.e. 30 and 100. We can these features for the human written and AI generated text to identify key parts of speech which contribute most.

Figure 14 shows that based on average fixation duration noun and pronoun contribute most for the text to be declared as human written when the threshold of I-VT algorithm is 30.

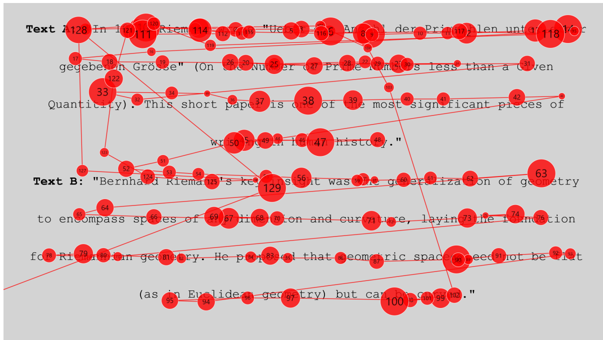

Figure 11: Scanpath of participant 4 for question 1.

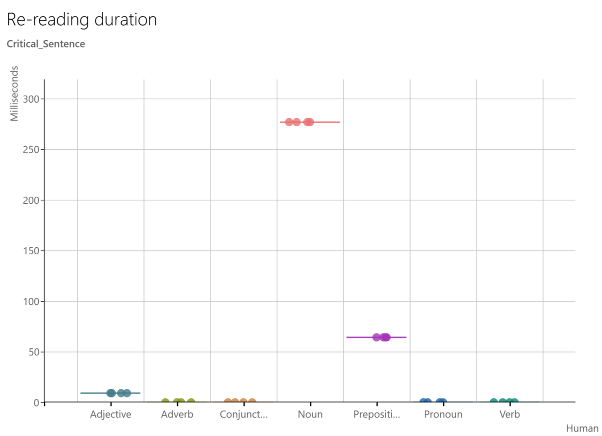

Figure 12: Re-reading duration for human written text for all participants.

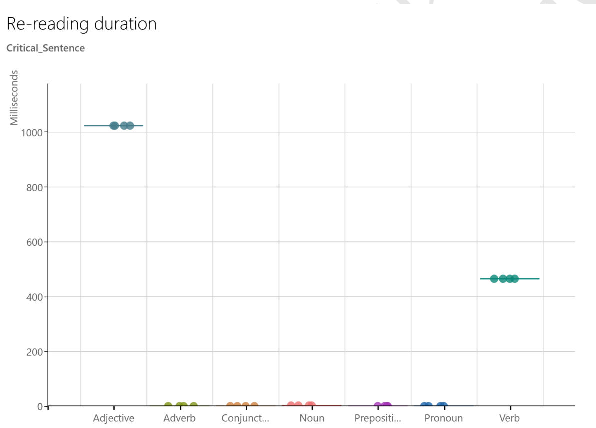

Figure 13: Re-reading duration for AI written text for all participants.

On the other hand, as shown in Figure 15 if we increase the threshold to 100 then adjective and noun are two prominent parts of

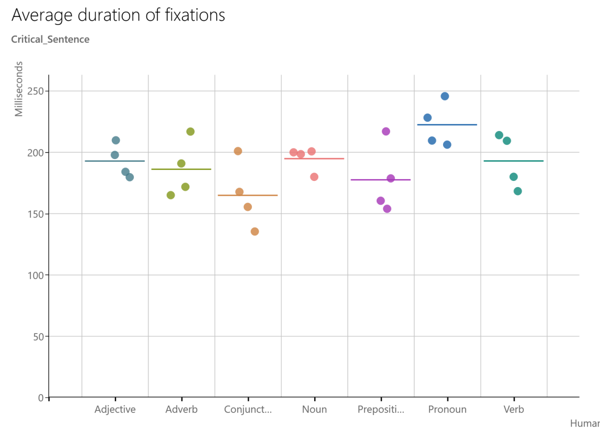

Figure 14: Average fixation duration for human written text using I-VT threshold=30.

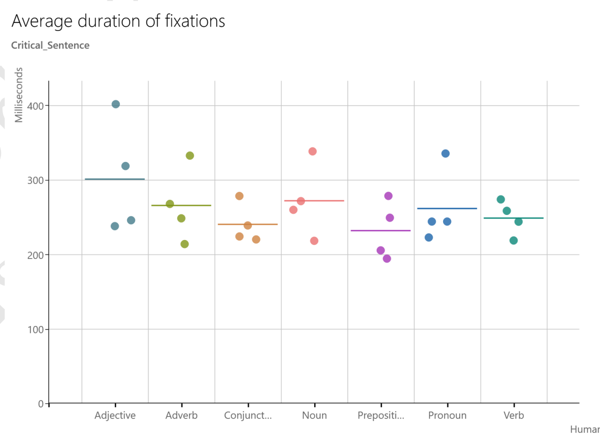

Figure 15: Average fixation duration for human written text using I-VT threshold=100.

speech to declare text as human written. So, if we take the common part of speech from both thresholds, i.e. 30 and 100 then it will noun which coherent with other analysis that it contributes most for the human generated text.

We also analyze average fixation duration for AI generated text using these two thresholds for I-VT.

Figure 16 shows that based on average fixation duration adjective, conjunction, noun and verb contribute most for the text to be declared as AI written when the threshold of I-VT algorithm is 30.

Other other hand, Figure 17 shows that adjective is the prominent part of speech which contributes most to be declare text as AI generated. Similarly, if we pick one common part of speech in two different settings of I-VT then it will be adjective for AI generated text.

We also analyze number of fixations for the two different thresholds for human written and AI generated text.

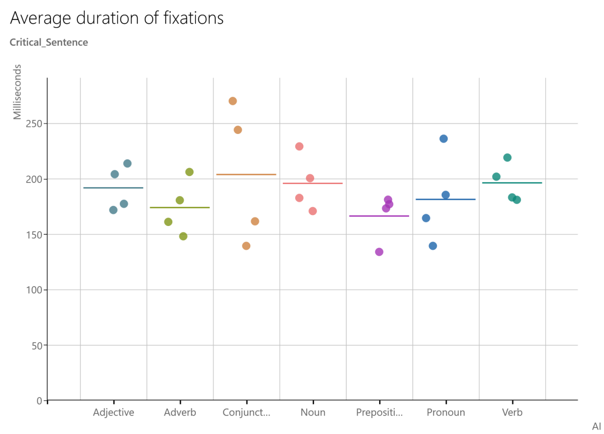

Figure 16: Average fixation duration for AI written text using I-VT threshold=30.

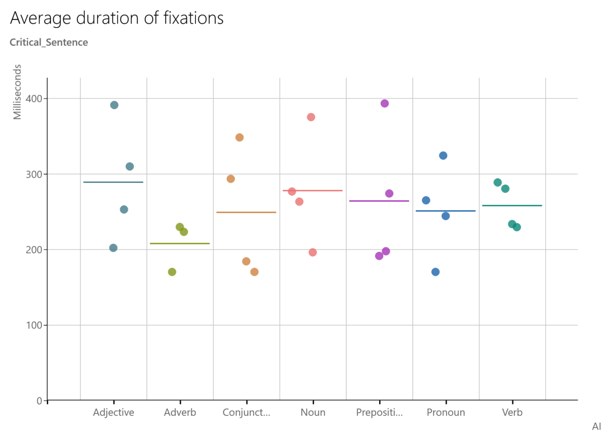

Figure 17: Average fixation duration for AI written text using I-VT threshold=100.

Figure 18 shows that the number of fixations for adjective and noun are higher than other parts of the speech to declare text as human written text when the threshold of I-VT is 30.

Figure 19 shows that the number of fixations for adjective and noun are higher than other parts of the speech to declare text as human written text when the threshold of I-VT is 100 which is similar to the threshold 30. From this we can also observe that like other analysis noun also plays important role even the feature is different.

We also analyze number of fixations for two thresholds for the AI text. We observe the prominent part of speech using this feature.

Figure 20 shows that the number of fixations for adjective are higher than other parts of the speech to declare text as AI written text when the threshold of I-VT is 30.

Figure 21 shows that the number of fixations for adjective, noun and verb are higher than other parts of the speech to declare text as

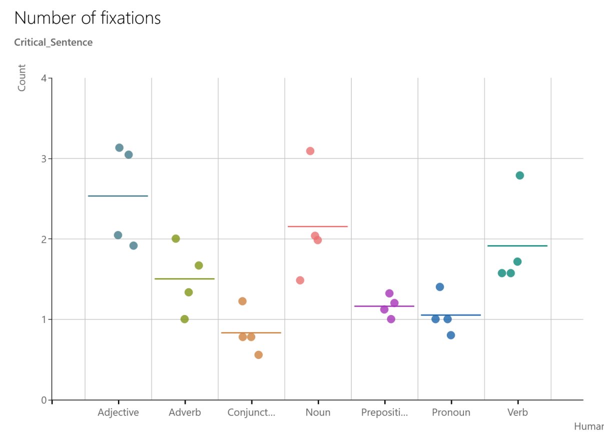

Figure 18: Number of fixations for human written text using I-VT threshold=30.

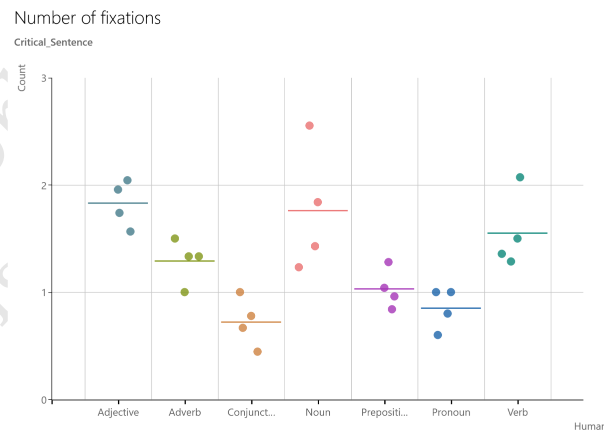

Figure 19: Number fixations for human written text using I-VT threshold=100.

AI written text when the threshold of I-VT is 100. From this we can also observe that like other analysis adjective also plays important role to declare text as AI written even the feature is different.

We also calculated mean fixation duration of each participant for four questions for two different threshold of I-VT which are presented in Table 1.

Table 1: MFD of each participant

Participant ID	MFD (Threshold =30)	MFD (Threshold =100)
1	198.71 ms	257.74 ms
3	170.26 ms	206.48 ms
4	201.50 ms	323.31 ms
5	177.28 ms	241.53 ms

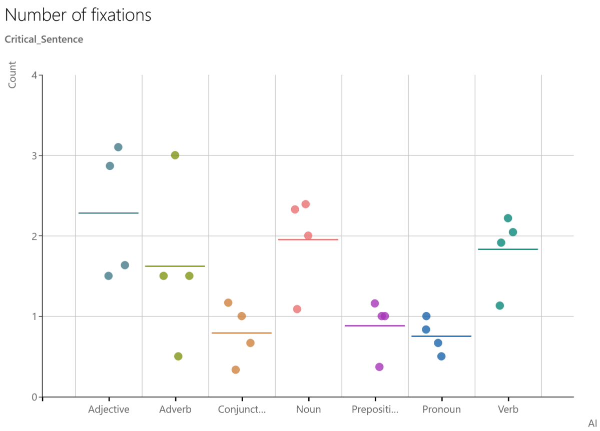

Figure 20: Number of fixations for AI written text using I-VT threshold=30.

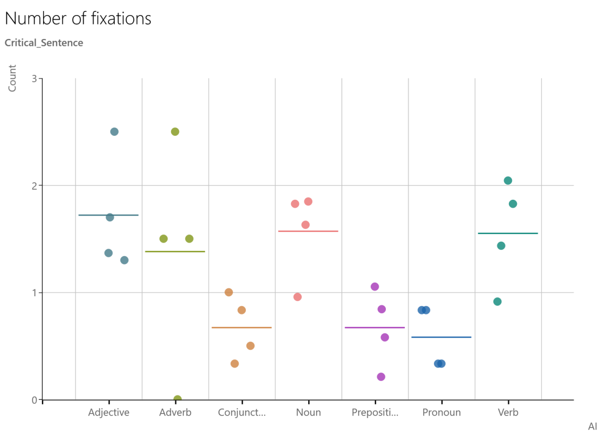

Figure 21: Number fixations for AI written text using I-VT threshold=100.

8 Conclusion

In this study, we performed eye tracking experiment to identify which parts of the speech contribute most to declare text as human written or AI generated. We performed analysis using I-VT algorithm with two different thresholds for four participants data. From our analysis, we can conclude that the participants focused more on noun when declaring text as human written whereas the participants focused more on adjective in case of AI generated text. The analysis are quite coherent with our real life observation as AI tools try to make sentence complex and wonderful. This experiment can be extended to more participant to make general observation and analysis.

9 Acknowledgments

The authors would like to thank Mohammadhossein Salari and Aadya Menon for their support during data recording session.

.

References

[1] Georgia Andreou and Maria Gkantaki. 2024. Tracking Adults' Eye Movements to Study Text Comprehension: A Review Article. *Languages* 9, 12 (2024), 360.
[2] Dario D Salvucci and Joseph H Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71–78.
[3] Abhinav Deep Singh, Poojan Mehta, Samar Husain, and Rajakrishnan Rajkumar. 2016. Quantifying sentence complexity based on eye-tracking measures. In *Proceedings of the workshop on computational linguistics for linguistic complexity (cl4lc)*. 202–212.
[4] Nili Steinfeld. 2023. How Do Users Examine Online Messages to Determine If They Are Credible? An Eye-Tracking Study of Digital Literacy, Visual Attention to Metadata, and Success in Misinformation Identification. *Social Media+ Society* 9, 3 (2023), 20563051231196871.
[5] Débora Torres, Wagner R Sena, Humberto A Carmona, André A Moreira, Hernán A Makse, and José S Andrade Jr. 2021. Eye-tracking as a proxy for coherence and complexity of texts. *PloS one* 16, 12 (2021), e0260236.

Received 13 December 2024; revised 13 December 2024; accepted 13 December 2024