Analysis of reading behavior based on POS through sentence completion tasks

Jintao Ren, Pter Pribil

February 19, 2018

Abstract

In this paper, we evaluate the reaction of people's behavior towards information from reading incomplete sentence based on their eye movements' metrics while they try to fill the missing word.The experiment has potential meaning for psycholinguistics as well, however, this paper only uncovers the relationship between part of speech(POS) and the missing word in one sentence, or how syntactic works with incomplete grammar dependency of POS. We discovered that fixation metrics perform differently under missing word scenario with ordinary reading tasks. Through an eye-tracking experiment, the eye tracker obtained eye measurements from 19 subjects in 2 groups of missing word sentences of different POS. In the first group of the stimuli, there was 30 individual Noun missing sentences, and in the other group, 30 individual Verb missing sentences are placed. Gaze metrics of each AOIs(tokens of a sentence) were obtained over 19 subjects, a qualitative analysis was conducted afterward based on the mean of total fixation duration on AOIs.

1 Introduction

With the development of natural language processing, multiple shades of techniques have been applied to the field. Including Part-of-Speech tagging and Parsing, one of the key elements for understanding human's perception towards syntax are dependencies and dependency grammar[Kurdi, 2017]. Meanwhile, parsing technology has been developed profoundly in recent decades, [Covington, 2001] developed a fundamental algorithm which could attach each word as soon as it can be attached, corresponding to parser referring the function in the human brain.

However, the syntax of the natural language is always ambiguous. There are many possible analyzes of typical sentences. In fact, perhaps surprisingly, there may be thousands of potential parses for a typical sentence (most of which seem totally pointless to humans)[1]. Which brings up our curious, how people handle syntax of a sentence in real life. What if its an incomplete sentence, how do we figure things out by existing information. Extensively, how people react to complete or repair the fragment of omission

[1]Wikipedia source:`https://en.wikipedia.org/wiki/Natural-language_processing`

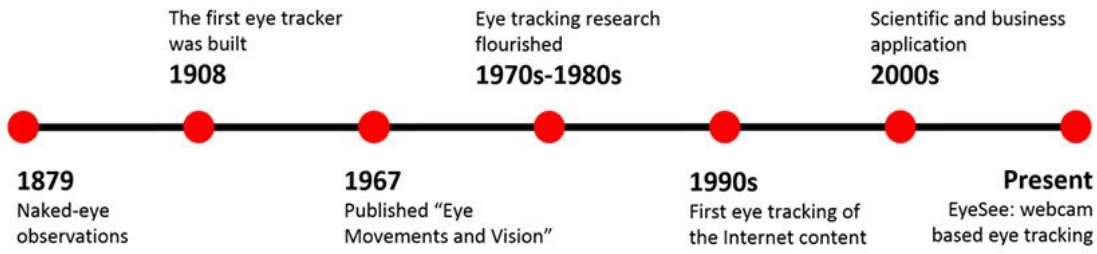

Figure 1: A brief history of eye tracking

word.

It is quite hard to track people's mind while reading, how do the thoughts exactly flow while reading has a profound meaning for the study of natural language. Which is also being studied by Neuroscientists with the help of fMRI images[Ramsey et al., 2001]. Fortunately, with the development of eye-tracking technology1[2], researcher could track down eye movement data while reading[Rayner, 1998]. Upon which, we could analyze the behavior of people's eye movement. Since then, numerous NLP studies based on eye tracking experiment have been published, some of them are quite interesting and relevant to our object. On one hand, [Nilsson and Nivre, 2009] uses a logistic regression model to predict saccadic eye movement behaviors. Which also emphasized that contrary to intuition, readers' eyes may not cross one line or one page of text smoothly. Reading studies are a striking fact that the eye carries out a series of very fast ballistic movements (called glances) from one place to another. Between eye movements, the eyes remain relatively still (fixation) for a short period of time. On the other hand, [Barrett and Søgaard, 2015] showed that gaze information could be used to discriminate grammatical functions by parts of speech(POS).

Moreover, from the previous study by [Duffy et al., 1988]; [Lavelli et al., 2009], we noticed that in normal reading patterns, readers fixate more and longer on open syntactic categories (verbs, nouns, adjectives) than on closed class items like prepositions and conjunctions. Is it the same while people try to complete a context-dependent word omission sentence? How existing POS correlate with eye tracking measures while controlling for oculomotor and missing POS word?

"Grammar is autonomous and independent of meaning"
—Noam Chomsky[Noam et al., 1966]

[2]Eye tracking history: `http://eyesee-research.com/blog/eye-tracking-history/`

2 Method

Design

Eye tracking technology is used in various fields, including Artificial Intelligence, Cognitive Studies, Linguistics, Virtual Reality, etc. Our interested field was Linguistics, as we were interested in the gaze duration on words in sentences focusing on syntactic categories strictly. The syntax is the study of sentence structure. Annotations for syntactic analysis include grammatical functions such as thematic and functional words. This table of Grammatical functions in appendix explained with detailedness and integration.7

We were mostly interested in category Noun and category Verb, so we will be putting an emphasis on those two. We wanted to combine syntactic analysis with Eye tracking, as it sounded interesting to see syntactic prominence not only in terms of purely theoretical syntactic analysis but also which words do people look at the most. We decided not to go into semantic analysis too deep, as it would have been a much more greater project, we only discuss observed results on the level of syntax.

To dig up the relation between parts of speech based on human eye movement behavior, we design and conduct the experiment on varies sentence completion tasks. We divide our experiment into two groups: Noun and Verb.

We would like to emphasize that the motivation behind this experiment of omission word is not to break through a new route in computational linguistics to explore eye movements during reading. We only want to testify our hypothesis in order to analyze and solve problems in a different direction.

Stimuli and experiment setup

For stimuli content, We constructed 60 different sentences, each with a missing word. Out of the 60, half of it was concerning only Nouns as missing words, and the other one Verbs. We paid attention not to put too many PPs, as ambiguity is not something with which we wanted to deal. The sentence length differs. There are some 5-word sentences, those are the shortest ones, and the longest one was 15 word long. They are all one sentences, except for a few, where it was necessary for the context to have two, but they are closely related. Some of the sentences had emotion words in them, mostly because we wanted to see how subjectivity or personality changes the gaze duration on these specific words.

For the construction part, we plant a notification page at the beginning of the experiment, briefing the upcoming experiment sections. Participants could press any key to start the experiment. For the stimuli sections, we put each sentence in one page, and form them with two parts, Noun and Verb, thus each part has 30 pages, in each page,

an enlarged sentence is placed in the center of the page. We also manually increased space between each word in the sentence for the convenience of later AOI region selection process. Between each section, we planted a notification screen with highly colored contrast information towards stimuli, indicating the break between each section and the beginning of next section. Users could press any key to swap to next page of sentence or section. While they reading the sentence, a voice recorder is placed to record their voice. Participants are required to speak out the word which they'd like to fill in the missing place, this is also informed by conductor and consent form. After the last sentence of stimuli, a black page of notification was showed to the participant, notifying the ending of the experiment, "Click to finish and please notify the conductor before your leave...".

The voice recording is a method to testify to guarantee that participant is focusing on the task, which is also an authentication for later selection of gaze data and has no functionality towards data collection and analysis. List of our experiment stimuli sentences could be seen at Appendix II - Stimuli5.

Participants

We recruited 19 participants(students) from a variety of master programs in University of Copenhagen. All of them have a strong English communication and reading skills since they all passed the necessity of IELTS score(6.5 or relevant) for entering the master programs. One of the participants is a native English speaker, who obtain a higher English application level comparing to others. All participants got instructions from the conductor and obtaining the necessary using ability towards an eye tracking experiment. All participants are aware of their upcoming experiments before conducting, they all have a brief understanding of the purpose of the experiment. They all signed the consent form which they were obligated to accept, including the confidential terms to protect the content from others who haven't do the experiment yet. Which is a protection method to reduce the false subject and bias of the experiment.

We provide them free coffee, chocolate, cookies and even beers(only after the experiment) as rewards for joining in the experiment.

For each of the participants, we conducted the experiment following the "Experiment Protocol", a calibration was obligated for every participant including pilots. Only when gaze percentage is larger than 80%, the experiment was continued, otherwise, retake the calibration. The participants were required to hold a relaxing, stable sitting posture and relative still distance and position towards eye tracker through the whole procedure of experiment, including calibration. The body position was adjusted in the first step of the calibration.

The sampling data extracted from the majority of participants looks perfect and sounded. However, there are 2 participants'(P15 and P16) gaze percentage is lower than 70%(majority above 85%) even though they passed the calibration. Both of them

Figure 2: Choose of AOI regions based on heat map results(points out of AOI region would leads to a lack of data, but their region are barely distinguishable)

were glass wearers and both of their experiments were run only under artificial illuminations(night). The verified guessing was the reflection of screen light from glasses intervened the seize process of the retina from eye tracker cameras. The natural daylight would be a supplement towards illumination, allowing more lights reflected by retina penetrate the glasses. It explains why such a scenario never happened to other glass wearers who's experiments were run at daytime. Unfortunately, we haven't found this issue in our pilot experiments, since the glass wearer pilot did the pilot experiment at morning.

Area of Interests(AOIs)

Selection and choose of AOI regions are always a task needs to be precision and patient. In this experiment, we are interested in gaze information from each POS element - each word. We segment AOI based on their geometry layout on the stimuli pages. Technically, even punctuations have their function towards syntax. However, current eye-tracking technology is not yet precise enough to discriminate punctuations with their nearest neighboring word. Thus we treat them as a single AOI. We would also ignore the weight of punctuations on POS tagging process in further steps.2 From the picture, we could see the gaze heat map from all 19 participants, we choose the region of each AOI referring the heat map results. By excluding some remote points would remove the poorly calibrated results, since which are hardly recognizable to assign into different AOI regions. We draw AOI regions as rectangles, by including the majority of points close to each word of sentences. We set all AOIs to be active since every part of our AOIs are relevant throughout different parts of the timeline.

In this task, we tested couple gaze metrics on the pilot experiment. In the first place, we lock down several eye tracking metrics including Time of first fixation, fixa-

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Figure 3: Correlation coefficient are applied to choose of fixation metrics

tion counts, total fixation and even regression counts. Time of the first fixation might be useful on normal sentence POS analyze, however, it omits the re-gaze duration after people found the missing blank while re-reading and pondering time for filling the blank. Our first trial was failed due to such a choice. The statistic figures on first fixation duration doesn't apply with overall metrics, which barely correlates with our target.

Afterwards, we also found that fixation counts and total fixation duration have a high correlation coefficient3 with each other in inter-subject(same person) experiment(we obtained numpy.corrcoef function from Python numpy library to calculate correlation coefficient[3]). However the fixation counts are also strongly dependent towards individual's habits of gazing, on the other hand, the figure doesn't have much variation in some individual, thus the fixation counts limits the generalization possibility. We decided to exclude this metric.

It is very likely that the global fixation time measure, such as the total time spent on an object during the scene viewing and the fixation duration on the object (all initial fixation time before the object leaves the object), reflects the post-recognition process. Therefore, the fixation duration in scene processing may reflect other processing than object recognition.[Duchowski, 2007] Which might be a chaotic measurement in recognition experiment, but for our word omission experiment, we need to track the duration of thinking period while people try to fill the blank word.

We choose Total fixation duration as our major metrics for statistical usage. This metric measures the sum of the duration for all fixations within an AOI (or within all AOIs belonging to an AOI group)[4] In our case, each individual word. We also treat blank mark "_____" as an AOI, and include punctuations to it's left(closest word) AOI region. For statistic purpose, we calculated the mean total fixation duration from each subject and align them with each POS tags(as labels).

Parsing and POS tagging on stimuli

In this part of the experiment, we choose to use NLTK library for tagging and parsing. Firstly we tokenize each stimuli sentence, POS tagging with existing methods as following:

[3]Python library: `https://docs.scipy.org/doc/numpy/reference/generated/numpy.corrcoef.html`

[4]Tobii Studio, user's manuals: `https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf`

Sample of stimuli from Verb missing group:

Abraham sincerely ____ cognitive experiments.

Tagging:

Abraham/NNP sincerely/RB _/VBZ cognitive/JJ experiments/NNS ./.

Parsing:

```
(ROOT
  (S
     (NP (NNP Abraham))
     (ADVP (RB sincerely))
     (VP (VBZ _)
        (NP (JJ cognitive) (NNS experiments)))
     (.  .)))
```

3 Data collection

In this experiment, we perform both pilots and real experiments with the same equipment in the same room. The experiment uses Tobii T120 eye tracker as the major information-gathering equipment. Subsequently, a supplementary voice recorder was used to verify participants attention while conducting the experiment.

Before each experiment, a well explain guide was introduced by the experimenter. All participants are noticed of the overview, procedures of the experiment, they noticed the potential risk and inconveniences(tired and bored) which may occur while doing or after they done the experiment. They all aware and agree with benefits, costs and especially the confidentiality listed in the informed consent form. All of them agreed and signed the form.

Another problem we found related to participants is, even though we choose the gaze metric from variety evaluation directions, fixation duration is still strongly related to an individual. From the previous study, we know that most of the fixation time is about 250 milliseconds, but there are considerable differences from participants to participants. As a result, some gaze durations are pretty low, while other gaze durations may more than 500 milliseconds.[Rayner, 1998] This may be caused by reading hobby of each individual. This factor may also influence our choice to manipulate the data.

Collection of data based on AOIs

From the last section, we explained the selection of the AOIs regions and the gaze metric. In this part, we obtained total fixation duration data for each individual AOI(word)

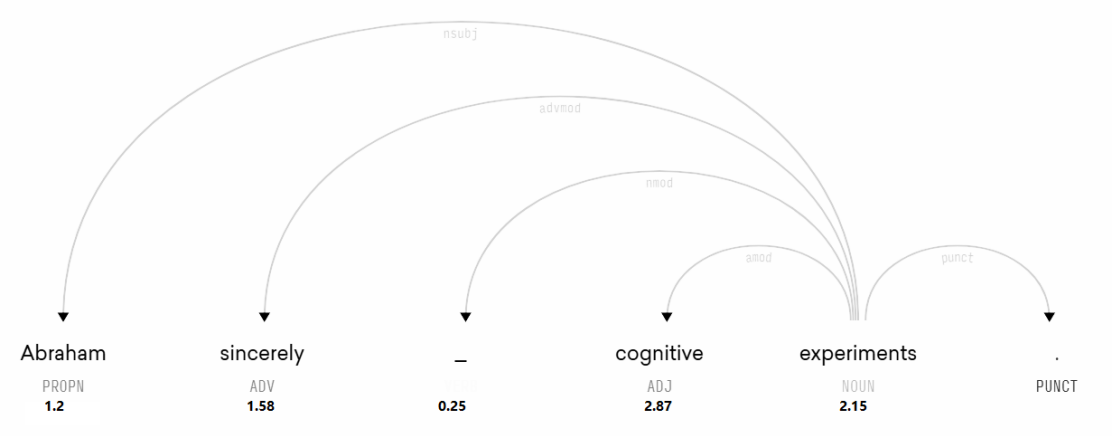

Figure 4: Dependency structure with mean total fixation duration(second) over 19(17 valid) participants of each AOI(punctuation is included to it's closest word)

from 60 different sentences.

In each of the sentence(stimuli), we planned to obtain 19 participant's total fixation duration data towards every AOI. However, as we have already noticed the low gaze ratio problem from participant P15 and P16. After evaluation of their data, we found that lots of their fixation data were absent or incomplete for total 60 stimuli. Some of the whole sentence data were missing, in most other cases, we miss part of words in one sentence. After evaluating the practical function, we decided to exclude their all data entries for later procedures.

Besides, we also noticed that some of the participants are missing data on the certain region of AOIs. Since we've already foreseen such a problem on the pilot experiment, the missing data problem is caused by the selection of the AOIs, we only choose the majority data which have an obvious distance meaning towards each AOIs. Thus the excluded points lead to an empty figure in certain AOIs for certain participant. We decided to exclude the specific sentence of their data since these cases are relatively dispersed.

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1}$$

Afterwards, a simple statistic method: Arithmetic mean was applied1. We calculated the mean value of total fixation duration from 17 valid participant's valid data in each AOI. In conclusion, the majority of the data which we obtain are in solid condition, since we only choose the meaningful close data points. The existing data are more representative and sound towards each POS.

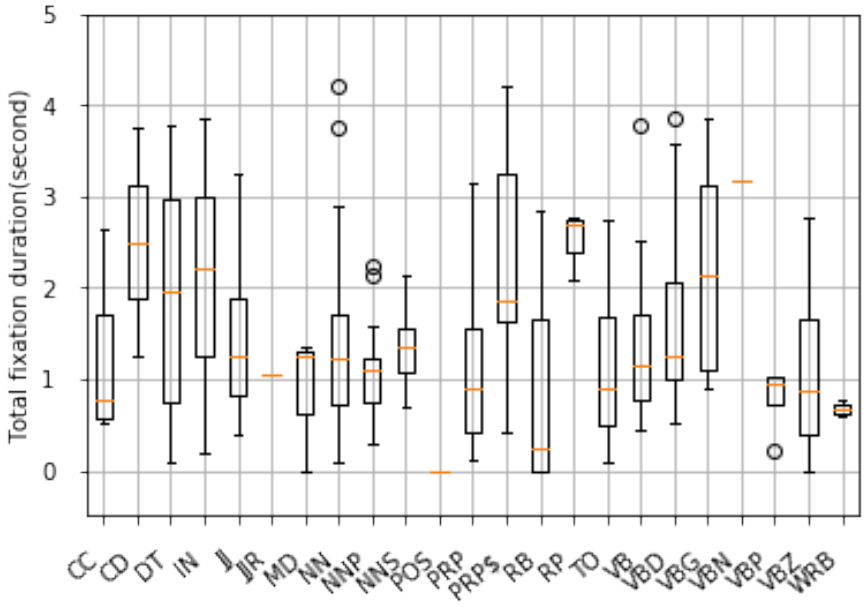

Figure 5: Boxing plot of mean total fixation duration over all POS in Noun omission set

4 Results

Data Analysis

This section intends to find out the coefficient behind different parts of a sentence with inefficient information(missing words) syntactically. Thus a linguistically driven study and analysis based on extracted fixation figures are used.

We didn't differentiate between the Grammatical Functions. Objects, Subjects, obliques, and so on, were all just Nouns without any specifications. Verbs were thematic, not functional. We did not care about which words the participants actually used in the missing spots. It would have had little to do with our experiment. This is also one of the reasons we didn't differentiate between Noun categories. We were just interested in their way of finding out what the missing word might be.

Set of missing Nouns

In the first 30 sentences, we analyzed the outcome where the Nouns were missing. The shortest fixation duration was 4.01 seconds for the sentence "The _ asked for money". The longest fixation duration was 23.08 seconds for the sentence "The bird flew past the _ before it got shot."

Surprisingly, the length of the sentence doesn't always correspond to the fixation

duration. The longest sentence had the fixation duration of 12.59 for "The kid always said to her mother if the _ wants something, he'll get it." Also, there were examples of short sentences with surprisingly long fixation duration: "John put _ on his pizza." with 17.22 seconds.

Depending on the length of the sentence, each missing word has a scope of 0-3 words that were taken into account when trying to figure out the missing word. By that, we mean that those words took the longest to read, or they went back to them with their eyes for double checking before answering.

All the other words were running through quickly, except for some cases: "John decided to stop eating meat after he saw a _ being slaughtered."

"Mary didn't answer the phone because she knew it's her _."

"Dancing makes you feel like a _."

"meat" in the first case had 2.25-second fixation duration, even though it was outside the scope of 3 words. Everything else around it ranged from 0.23 to 1.52 seconds.

"didn't" in the second case took 2.79 seconds to read, was way outside the scope, and every word around it had 1.2 or 1.57-second fixation duration.

In the third case, "dancing" resulted in 2.15, again not correlating with the hypothesis that if a word is outside the scope, it's read quickly.

In all three cases, the irregularity is not surprising at all. They are relevant parts of the sentence in order to fill in the missing word. As these sentences are not long, not too much information is given for the context, whereas in longer ones, participants could afford to look only at the surrounding words of the blank space.

In the case of shorter sentences with long fixation durations, each word took surprisingly long to read. The missing space's fixation duration was considered short except for one case (2.78 seconds). The duration ranged between 0.1 and 1.2 seconds mostly. If it was more, we considered it to belong. This is why the 2.78 was outstanding in this case.

Let's take a look at the words in the scope now. The scope is usually the determiner of the noun, the subject/object, and the verb. Prepositions might also occur, but they usually have shorter fixation duration.

There isn't seem to be an exact reason why each word is looked at for so long. We were expecting the verb to be the most prominent one, as it is the predicate of the sentence. However, whatever that was immediately next to the missing word, it was

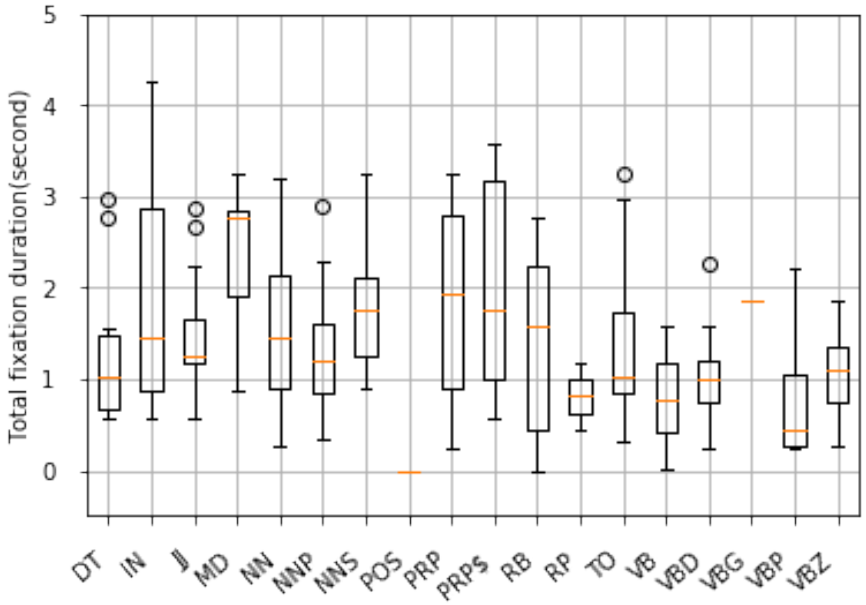

Figure 6: Boxing plot of mean total fixation duration overall POS in Verb omission set

highlighted as long fixation duration, and they are usually the longest ones, except if there is a negation in the scope of 2, or an infinitive + verb, where the verb gets a longer duration.

Also, referencing ("his", "her", its") took a lot of time. This shows that it takes time to connect the reference and the referee in the sentence, which is time-consuming. Its position was taken into consideration, but statistics show that the ones close to the missing words took the longest to read. The longest one was in the sentence "John put _ on his pizza." , resulting in a 4.21 fixation duration. Only that one word took more time on average to read than the shortest sentence.

Interestingly, if the verb is outside the scope, it has less than 1-second fixation duration usually. Exceptions are those that are negated. All the negated verbs, auxiliaries have higher fixation duration, even if they are outside the scope, and not only the verbs themselves, but the negation as well.

Set of missing Verbs

The sentence with the shortest fixation duration was 8.05 seconds, which was "Abraham sincerely _ cognitive experiments". The longest one-sentence example was 18.08 seconds, "Andrew is the type of man who would _ his puppy." We included the longest one-sentence example, and not the example for the longest fixation duration, as it was 2 sentences. Relative to the sentence length, the longer ones are not interesting usually.

We have seen that most of them get filtered out quickly, and only the scope is interesting. The same happens with multiple sentence input. The one in which there are no missing words, the participants read through them quickly. As in the missing Noun examples, the length does not necessarily correspond to the fixation duration. But again, the scope does.

The scope was again 0-3 words to the right or to the left. The ones closest to the missing word had the longest duration. The scope is usually an infinitival "to", a preposition, or a noun. Contrary to the Nouns, the prepositions always had long fixation durations. Prepositions are more important for verbs, as they are necessary for completing argument structures. In cases where the missing word was accompanied by a "to" on its left, and the missing word's right part was not a period, in all the cases, except for two, the infinitival "to" was read in significantly less time than the word on the right. The two exceptions had the difference of 0.1 and 0.16 seconds, whereas the rest had 2+ seconds difference. All the infinitival "to" not next to the scope had less than 1.0 gaze length. Everything else seems to be the same as for the Noun category words.

There were some exceptions to the scope of missing verbs also.

"Harry _ his child with extreme happiness"

Here, "happiness" has a rather high fixation duration, that is 2 seconds.

In case of "Elizabeth became suicidal after her mother told her that she _ her", "suicidal" had a duration of 2.25, even though it was at the beginning of the sentence.

Another notable example is "Peter _ Mary to take birth control pills.", where "birth" and "control" were high on the list with 2.59 and 3.21 seconds.

Referencing was important in this category also. It behaves the same way as in the missing Noun sentences. They had longer fixation duration on average than any other grammatical functions.

Evaluation of the two sets

The two sets give similar results: The scope of the missing words is 0-3 words (some end or start with the missing word). There were only two examples in the noun category where the sentences were short, and their gaze duration was surprisingly high, leading to the whole sentence as the scope ("Josh doesn't like following _" and "Mary received a letter with a lot of _").

The missing words were only looked at for a notable period of time when the sentence itself took longer to finish. Out of the 60 examples, this happened 3 times, all of them were shorter sentences. Referencing was the same in both sets, taking up most of the

participants' fixation time.

Fixation duration cannot be predicted from the length of the sentence. In both sets, there were shorter sentences with longer gaze than longer sentences. This further proves that the scope is limited in most cases, and redundant information is looked over quickly.

All the irregularities but one are strongly emotional, which might be the reason why the participants were looking at them for long even though they were out of the scope of the missing word. The only one that is not is the negation "didn't". Negation attracted the attention surprisingly well in both sets.

5 Discussion and Conclusion

There is a scope for the missing words in both sets. For this reason, longer sentences might not take too long to read, as redundant information is filtered. This might not be the case for words displaying strong emotions. Also, referencing takes up a lot of time, most probably to identify the referee.

We indeed found types of words that are more prominent than others (emotion words, PPs in case of guessing the verbs), but since the data we worked with is limited, we can only conclude from this amount of sentences. Besides the main purpose, we also could conclude that the most neighboring words have the most fixation durations comparing to others. This also explains the human nature towards mystery, by seeking closer information to uncover the mask.

However, If we had more data, especially, well organized and categorized data we could cluster the types of words in the scope of the missing words more accurately.

Further research:
As the fixation duration is strongly individual independent, we could introduce a standard deviation tool for those strong biased individuals, and re-scale their data to match global metrics. Another improvement could be, we simply extract all major gaze metrics, and run correlation coefficient based on SVM training results.

On the other hand, strong emotion words affecting the sentences and the gaze duration are present in some of the sentences. Due to the lack of data, we could not conclude how they work precisely. More data would allow us to do extensive research on how the referencing works. The use of machine learning techniques would also be prominent if we have the potentiality of digging up extensive sentences. They were chaotic, but there seems to be some kind of a connection to its position. Whether it's put after or before the missing word or other specific words yields longer duration remains a mystery.

References

[Asudeh and Toivonen, 2009] Asudeh, A. and Toivonen, I. (2009). Lexical-functional grammar. *The Oxford handbook of linguistic analysis*, pages 425–458.

[Barrett and Søgaard, 2015] Barrett, M. and Søgaard, A. (2015). Using reading behavior to predict grammatical functions. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pages 1–5.

[Covington, 2001] Covington, M. A. (2001). A fundamental algorithm for dependency parsing. In *Proceedings of the 39th annual ACM southeast conference*, pages 95–102. Citeseer.

[Duchowski, 2007] Duchowski, A. T. (2007). Eye tracking methodology. *Theory and practice*, 328.

[Duffy et al., 1988] Duffy, S. A., Morris, R. K., and Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of memory and language*, 27(4):429–446.

[Kurdi, 2017] Kurdi, M. Z. (2017). *Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications*, volume 2. John Wiley & Sons.

[Lavelli et al., 2009] Lavelli, A., Hall, J., Nilsson, J., and Nivre, J. (2009). Maltparser at the evalita 2009 dependency parsing task. *Proceedings of EVALITA Evaluation Campaign*, 31.

[Nilsson and Nivre, 2009] Nilsson, M. and Nivre, J. (2009). Learning where to look: Modeling eye movements in reading. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 93–101. Association for Computational Linguistics.

[Noam et al., 1966] Noam, C., Duncan, L. R., Eugene, G., et al. (1966). Syntactic structures.

[Ramsey et al., 2001] Ramsey, N., Sommer, I., Rutten, G., and Kahn, R. (2001). Combined analysis of language tasks in fmri improves assessment of hemispheric dominance for language functions in individual subjects. *Neuroimage*, 13(4):719–733.

[Rayner, 1998] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372.

Content and individual responsible

*Appendix I

SUBJECT (SUBJ)	Some people with no shame walked in and wrecked the party. The party was wrecked by some people with no shame.
OBJECT (OBJ)	*First object.* Ricky trashed the hotel room. Ricky gave John a glass. Ricky gave a glass to John.
OBJECT$_\theta$ (OBJ$_\theta$)	*Second object. Thematically restricted object.* Sandy gave John a glass. Tom baked Susan a cake. #Tom baked a cake Susan. (OBJ$_\theta$ *in English restricted to theme, cannot be beneficiary*)
OBLIQUE$_\theta$ (OBL$_\theta$)	*A complement (non-subject argument) that has oblique case or is a PP.* Julia placed the vase on the desk. Ricky gave a glass to John.
COMPLEMENT (COMP)	*Closed (saturated) complement: A clausal argument which has its own subject.* Peggy told Matt that she had won the prize.
XCOMP	*Open (unsaturated) predicate complement: A predicative argument with no overt subject of predication.* I told Patrick to quit. Peggy-Sue seems to be a complete fraud.
ADJUNCT (ADJ)	*A modifier, a non-argument.* Mary read a good book. Mary counted the cars very quickly. Sally killed a bug in the yard. Since she had no money, Mary was forced to get a job.
XADJ	*Open predicate adjunct.* Having no money, Mary was forced to get a job.
SPECIFIER (SPEC)	*Possessor or quantificational determiner phrase.* John's book's cover is red. At least three books are red.
TOPIC (TOP)	*Grammaticalized discourse function.* *Must be identified with or anaphorically linked to another grammatical function.* Mary met the author whose books __ annoyed Peggy. (TOPIC = SUBJ) Bagels, Mary loves __. (TOPIC = OBJ) As for bagels, Mary loves them. (TOPIC anaphorically linked to OBJ)
FOCUS (FOC)	*Grammaticalized discourse function.* *Must be identified with or anaphorically linked to another grammatical function.* Which author do the critics praise __? (FOCUS = OBJ) Cén t-údar a molann na léirmheastóirí é? (FOC anaphorically linked to OBJ) Which author COMP praise the critics him (Irish; McCloskey 1979: 53)

Figure 7: Table of Grammatical Functions[Asudeh and Toivonen, 2009]

*Appendix II - Stimuli sentences

Noun group

The _ asked for money.
John asked Mary to slap her _.
The young, handsome guy took advantage of the _.
The kid always said to her mother "if the _ wants something, he'll get it".
Peter picked up the _ from the dirt and ate it.
Megan introduced her _ to her parents.
Obama got into the car and saw three _.
Mary spat on Peter's face because he cheated on her with her _.
I wish I had an _ to grant all my wishes.
_ turned out to be gay, but the father didn't mind.
I spent 1000 dollars on _ just to show the world how rich I am.
Resistance is futile.Said the _ to the victim.
The white lab coat guy invited the _ to his old truck.
John decided to stop eating meat after he saw a _ being slaughtered.
Mary didn't answer the phone because she knew it's her _.
Josh doesn't like following _.
He's so violent.Peter asked God how to be a good boy.
He said, "trust your _"."I did nothing wrong", repeated the _ several times.
The bird flew past the _ before it got shot.
Stop eating _ every day.It will make your life shorter.
Bob is a rebel.He told his _ he won't eat spinach.
John put _ on his pizza.
Programming makes you feel like a _.
Dancing makes you feel like a _.
When I ordered my first _, I was quite nervous.
Mary received a letter with a lot of _.
Bob opened a jar with _.
Peter gave some money to his friend with _.
Megan kissed her boyfriend and gave him a _.
I gave the world _ and made it a better place.

Verb group

Peter asked the nice lady to _ with him.
Mary got interested in psychology so she _ a friend to help her out.
Bob wanted to _ with Susan, but she refused.
Josh told her angry mother that he didn't _ his homework.
Joe became really paranoid after his friend _ him.
Abraham sincerely _ cognitive experiments.

David wanted to _ someone after he realized he had a week left.
Italians often _ about the quality of their food.
The famous actor _ with his children.
Jonathan really enjoyed the dinner.He said, "I will _ you".
Harry _ his child with extreme happiness.
Andrea never _ about her negative traits.
Peter always _ about his negative traits.
The chocolate factory _ children.
Peter _ Mary to take birth control pills.
Travis _ about having kids to his new girlfriend.
Patric never told his mother that he _ John.
Barbara couldn't find clothes that would fit her, so she decided to _.
Elizabeth became suicidal after her mother told her that she _ her.
John loves Christmas because he can _ people.
Johanna wanted to be rich, so she _ her boyfriend.
Video games are great to _ your life.
I can't _ my mother.
Kate's dog is a very good boy.She _ (at) strangers.
Bob studied hard to become famous so he could _ with people.
Andrew is the type of man who would _ his puppy.
Rebecca has no idea how to _ a psychopath.
John asked for a cheeseburger, but his mom _ him.
Peter hates cats.Whenever he sees one, he _.
John got into a love triangle.The only way out is to _ them.