Effect of Plausibility on Analysis of Garden Paths

A Gaze Tracking Study

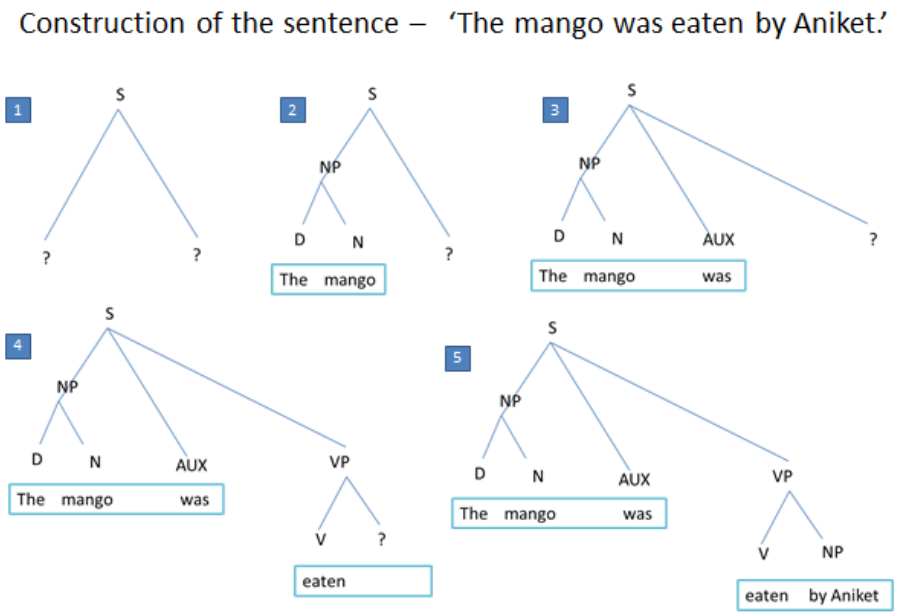
Chirag Gupta
Indian Institute of Technology, Kanpur

1 Introduction

Research on the cognitive science of sentence processing has been driven heavily by the study of resolution of syntactic ambiguity in Garden Path sentences. Recent work has shown that apart from syntax, semantics is another essential component that affects the analysis of sentences. In this study, we analyze the role of semantic plausibility in the resolution of syntactically ambiguous Garden Path Sentences.

1.1 Theory of Incremental Evaluation

The brain processes sentences sequentially, and the most 'probable' parse tree is selected at every word. The probability is predominantly determined by the *syntactical role* of words, rather than their semantic meaning.



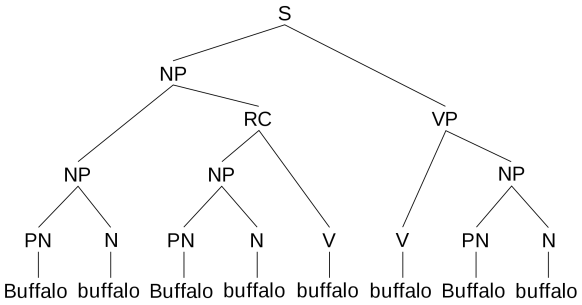
On encountering a parse that is wrong, the information is reanalyzed, and a new parse tree is constructed.

1.2 Ambiguities in incremental evaluation

It turns out that ambiguities often occur in incremental evaluation. A particular class of sentences is **Garden Path Sentebces**. In these sentences, the most likely parse in incremental evaluation turns out to be incorrect. let us look at a couple of examples -

- **The criminal confessed his sins** *which* upset kids harmed too many people.
- **As the woman edited the magazine about fishing** *amused* all the reporters.

Often, sentences may be impossible to parse, such as the following -



The following keywords are enlisted along with their definitions. These will be used in the discussion that follows.

- **Misanalysis** An initial parse that turns out to be wrong.
- **Disambiguating Word** A word that does not fit into the current parse and hence forces reanalysis.
- **Reanalysis** Formation of a new parse tree on basis of the new information, may involve visiting previous parts of the sentence again.

1.3 Gaze Tracking

The effect of Garden Paths can be studied using Gaze Tracking data. This data can be analyzed by looking for the following -

- **Regressions** Moving of the gaze from the current word to a previous word.
- **First Pass time** The total time spent to read the sentence increases.
- **Fixation** Subjects fixate gaze on the disambiguating word, since it is unexpected.

The set up we used also provided videos of gaze tracks, which could be observed. In our eventual analysis, the analysis of the data was difficult and we reverted to these videos itself to prove our theories.

2 Proposition

Pickering and Traxler (1998), in a benchmark study, formally studied the precise role that semantics plays in identifying the most probable sentence. (for original study, see [4])
We know that the syntactic role of elements in a sentence predominantly controls its parsing. However, Pickering and Traxler proved that the **semantic plausibility** of the parse defines the extent of **attachment** towards the sentence. Essentially, although the final parse is decided by the syntax, the difficulty encountered in 'letting go' of a wrong parse (during a garden path effect) is proportional to the semantic plausibility of the initial parse. Greater semantic plausibility of the inital parse implies a larger time required for reanalysis alongwith increased regressions.

3 Experimental setup

(see Section 6 for a link to the test data)
Two experiments were performed simultaneously, on two different types of clauses. Both related to garden path sentences having plausible and implausible interpretations for the post-noun phrase. The two types of clauses studied were -

- **Experiment 1 : Subordinate Class**
  A clause dependent on a main clause. Subordinate clauses are typically introduced by conjunctions.
  → **When the ambassador negotiated the treaty** *about* arms upset many of the civilians.
- **Experiment 2 : Complement Clause**
  A clause introduced by a complementizer such as that or whether.
  → **The sailor reads the chart from London** *described* new routes around the world.

3.1 Types of sentences

4 types of sentences can be formed for each of the clauses. These are,

- **Type I.** Initial parse plausible, syntactic disambiguation absent.
  As the artist paints the picture of the roses pleases the critics greatly.
- **Type II.** Initial parse plausible, syntactic disambiguation present .
  As the artist paints, the picture of the roses pleases the critics greatly.
- **Type III.** Initial parse implausible, syntactic disambiguation absent.
  As the artist sings the picture of the roses pleases the critics greatly.
- **Type IV.** Initial parse implausible, syntactic disambiguation present.
  As the artist sings, the picture of the roses pleases the critics greatly.

Types I and III are garden path sentences, with plausible and implausible pre-verb phrases. These are the sentences we compare amongst. Types II and IV are control sentences for types I and III resp. All data for a test sentence is taken with respect to its appropriate control. Each of the 8 sentences have different base meanings, along with different structures. In particular, the control of a sentence matches the sentence only with respect to its structure, and not the meaning.

3.2 Methodolgy

Certain specifications of our experiment are listed below -

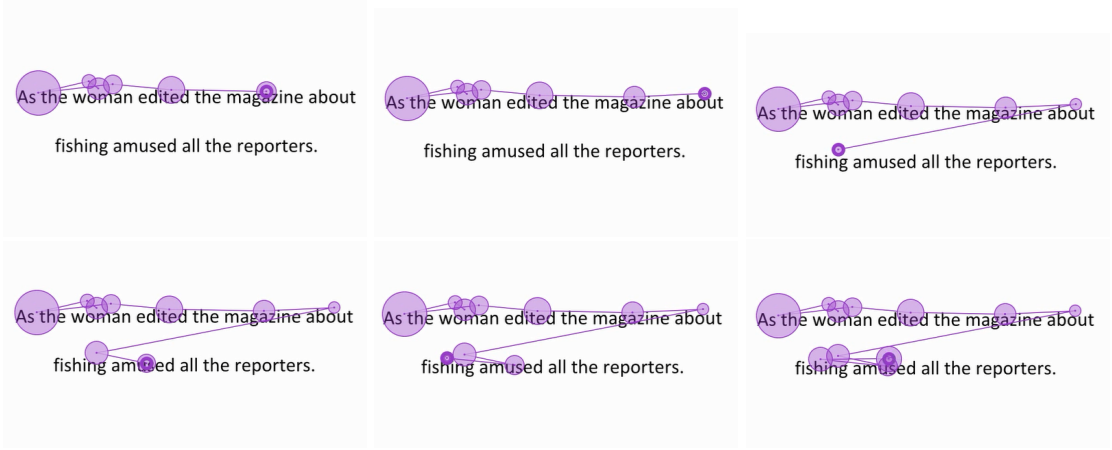
1. Gaze Tracking apparatus and software designed by **SensoMotoric Instruments (SMI)** was used.
2. 4 test sentences (as specified above) + 4 controls sentences + 18 fillers. No two test sentences were consecutive.
3. 4 sentences for each of the two experiments → types I, II, III, or IV.
4. All sentences written over two lines, sufficient distance (1") between the two lines. (essential for better gaze tracking results)
5. Instructions to participants
   - Do not waiver focus from the sentence. Read sentence sequentially, word by word.
   - You can go to a previous word to understand the meaning, but always focus on the current word you are reading.
   - Close your eyes once you are done reading the sentence. *(This is an indicator for the experimenter to move to the next sentence)*
   - Understanding the meaning of the sentence is essential to the experiment. Close your eyes only after understanding the meaning.

4 Results

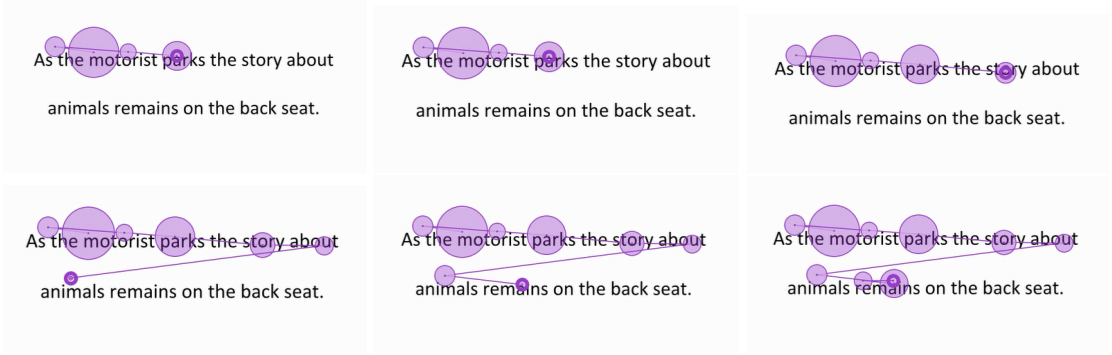
The first subsection shows some real videos we obtained during the gaze-tracking studies, along with their analysis. The subsequent subsection contains graphs plotted by Pickering and Traxler, since our own data could not be collated. Analysis of the graphs, with references to our snapshots is also done. (see also Section 6 for a link to some of the videos)

4.1 Snapshots of videos

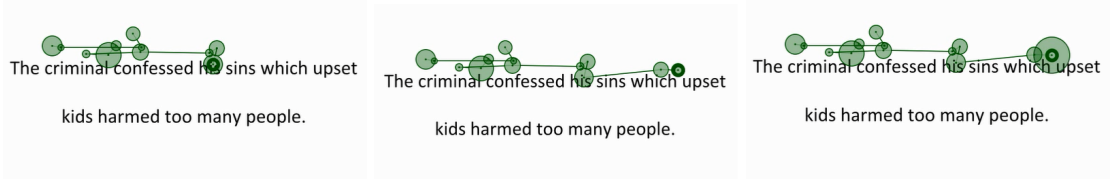
1. Subordinate clause, plausible noun phrase. Observe the sharp regression on encountering the unexpected word, *amused*.

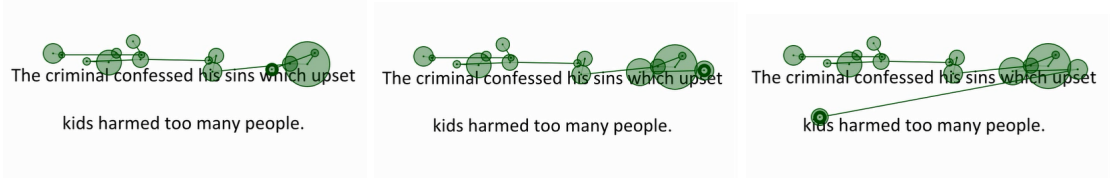


2. Subordinate clause, implausible noun phrase. Observe that there is hardly any regression in this case.

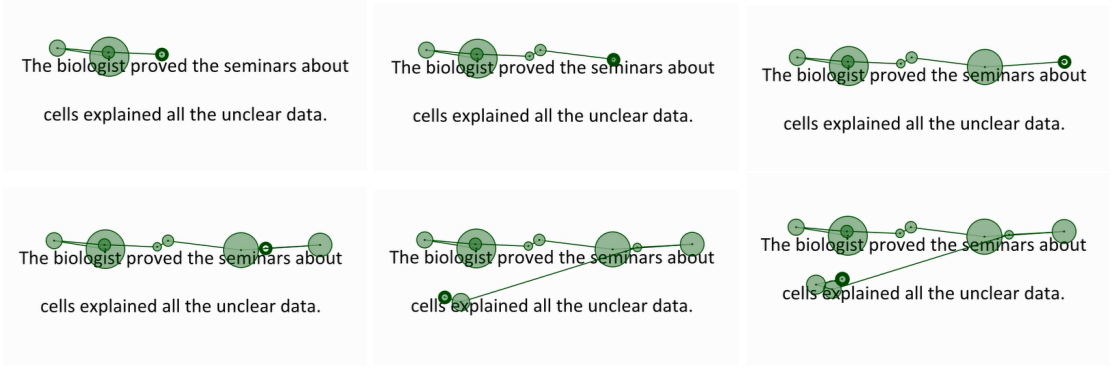


3. Complement clause, plausible noun phrase. Although it is much less clear in this case, and in particular no explicit regression is observed, a high increase in first pass time was observed, as is clear on the amount of time spent on the unexpected part, *which upset*.



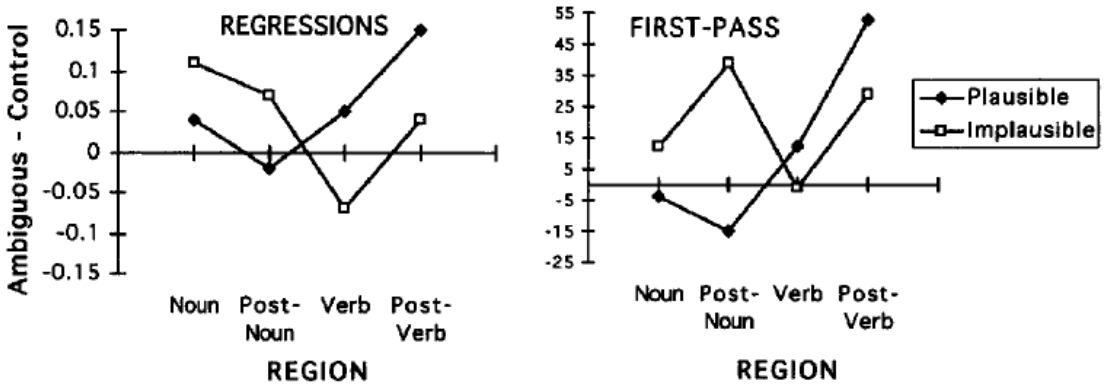


4. Complement clause, implausible noun phrase. As in 2 above, there is a very short regression (although a clear regression can be seen in this case). Very interesting is to note the increase in first pass time on the semantically unexpected (implausible) post-noun phrase, *seminar*.



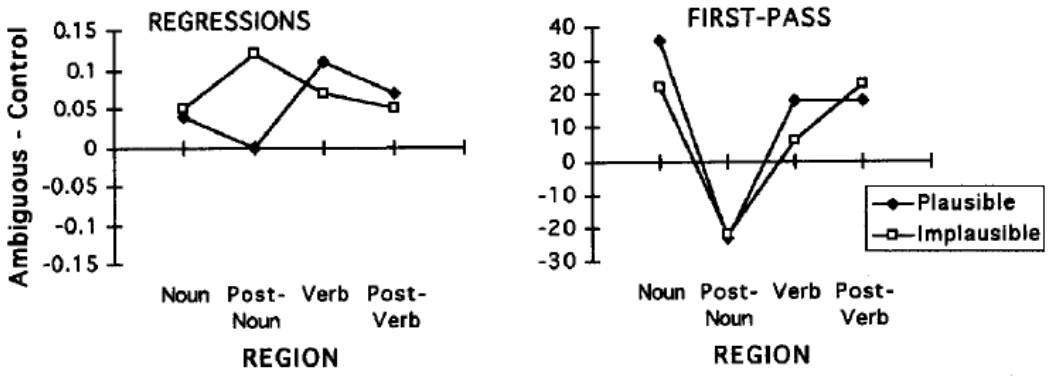
4.2 Plots and analysis

Experiment 1 (Data from authors [4])



Let us call the noun and post-noun region, Region I (R1), and the verb and post-verb region Region II (R2). It is clear that in the plausible case, R1 is acceptible and hence takes less time for analysis compared to the implausible case. More related to our hypothesis is what happens in region R2. Again, it can be seen that total regressions and first pass times are increased in the plausible case, since the subjects attach themselves to the plausible (but incorrect) analysis of R1.
In the videos we obtained, it was clear that often the implausible garden path sentences did not even show regressions. (see images 2 in the previous section)
It is also interesting to note some unexpected but consistently oberved results. Regression and first pass times increase for the post-verb region rather than the disambiguating verb region. A possible eplanation is that on encountering an unexpected strucuture, subjects *delay* analysis, and await more information from following words, before making a decision. This is an expected cognitive reply to an unexpected situation.

Experiment 2 (Data from authors [4])



The results are correlated with those of Experiment 1, as discussed above. The differences are less prominent however; this could simply be a statistical error.

5 Issues experienced and suggested improvements

Due to inaccuracies and errors in the setup, unfortunately the data was in very bad form, and impossible to analyze. We suggest two improvements that we believe should improve the data significantly.

Issue : The apparatus required the head to be kept stationery. Violation drastically affected calibration, and consequently results.
Solution : An adjustable chin-rest could be provided to assist the subject in keeping his head stationery.

Issue : The gaze of subjects often diverted from the sentence despite clear instructions.
Solution : A dark background around the sentence, with the sentence embedded over a white background could be used to make sure that subjects focus on the sentence.

It is important to note here that the instructions provided (see section 3.2) were critical in obtaining interesting videos, if not numerical data, and should also be followed.

6 Impending research

The gaze-tracking data is in csv format, containing X-Y coordinates, times, and pupil dimensions for fixations, saccades, and blinks. Analysis of this data would result in detailed numerical relations between each sentence and the total pass time, the regressions and fixation times on particular interesting elements of the sentence. As noted in Section 5, we faced some issues. Future replications of this experiment by students should keep in mind these bottlenecks and take appropriate measures to obtain better results.

On obtaining the data, we also intend to use the statistical Stanford parser [2] to obtain probabilities of the first part of the sentence. Plotting these probability values against the first pass and regression data would give concrete computational results about how plausibility affects analysis. Consequently an appropriate model can be fit to the data and theories formulated, such as those by Levy 2011 [3].

References

[1] Julie E Boland, Michael K Tanenhaus, Greg Carlson, and Susan M Garnsey. Lexical projection and the interaction of syntax and semantics in parsing. *Journal of Psycholinguistic Research*, 18(6):563–576, 1989.

[2] Danqi Chen and Christopher D Manning. A fast and accurate dependency parser using neural networks. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, 2014.

[3] Roger Levy. Integrating surprisal and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 1055–1065, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.

[4] Martin J Pickering and Matthew J Traxler. Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4):940, 1998.

[5] Patrick Sturt. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48(3):542–562, 2003.

Links

The 26 sentences used during the experiment can be found on the link -
`http://home.iitk.ac.in/~chiragvg/se367/project/TestSet.pdf`
Some of the sample videos havi been uploaded in avi format on the following link. All videos from which snapshots were provided in this report are uploaded on the link -
`http://home.iitk.ac.in/~chiragvg/se367/project/sample_videos/`