The MG model in short

A brief reminder: the MG model derives the complexity of a sentence by evaluating different notions of memory usage (metrics) for that sentence's derivation. Thus, we can do minimal comparisons across derivations: e.g. we can predict between two sentences which one is more burdensome from a processing perspective. While we have tons of metrics to do that, I'm focusing here on just 2, for consistency with some previous work from the lab: Max Tenure and Sum Size:

- Tenure measures how long a node must be kept active in memory by the parser. Max is just the highest tenure over the derivation
- Size intuitively measures the hierarchical distance between the base position of a mover and its landing site. So Sum is the total length of movement dependencies in a tree

The linking hypothesis: ***less memory burden –> greater acceptability***.

On the empirical side:

- I'm only looking at subject and adjunct islands. For whether islands, I'm assuming no structural differences between the island/non-island conditions, so the parser would make no predictions
- I'm comparing the parser predictions to the experimental results in (Sprouse, Wagers, & Phillips) and those in (Sprouse, Caponigro, Greco, Cecchetto)

Tl/DR: Results are (surprisingly) encouraging but mixed?

Subject Island in (Sprouse, Caponigro, Greco, Cecchetto)

That paper compares 4 sentences across 2 conditions: subject vs object extraction & island vs non-island structure. The parser predicts (x > y indicates that x is more acceptable than y):

- Subj/Non Island > Obj/Island
- Subj/Non Island > Subj/Island
- Subj/Non Island > Obj/No Island
- Obj/No Island > Obj/Island
- Obj/No Island > Subj /Island

These match the results in the paper. However, the parser also predicts:

- Subj/Island > Obj/Island

Because it picks up on the length of the extraction in the object case. However, it should really be Obj/Island > Subj/Island, as the Subj/Island condition is the violating one. This is not a bad results for me though, as I'd argue that the

reason the parser cannot capture this contrast is because the lower acceptability of the Subj/Island condition is not due to processing factors. Crucially, the parser correctly captures the gradient of acceptability for those conditions that, grammatically speaking, should all be equivalent. So, success?

Subject island in (Sprouse, Wagers, & Phillips)

The previous results led me to wonder what happens when the grammatical violation and a processing factor like length of the dependency coincide. In such cases, the parser should make the right predictions all over. (Sprouse, Wagers, & Phillips) compare matrix vs embedded sentence extraction (here short dependency vs long dependency, respectively).

The parser predicts:

- Short/No Island > Long/No Island
- Short/No Island > Short/Island
- Short/No Island > Long/Island
- Long/ No Island > Long/Island
- Short/Island > Long/Island
- Short/Island > Long/No Island

Now, Ex.1 and Ex. 2 in that paper report different results for the short vs long dimension in the non-island case. Ex.1 shows that Long/No Island > Short/No Island; in which case the parser would make the wrong prediction However, Ex.2 shows the opposite, and the parser would make the right prediction.

Adjunct Island

A similar problem re:which results do we use as a baseline comes up with adjunct islands.

The Parser predicts:

- Short/No Island > Long/No Island
- Short/No Island > Short/Island
- Short/No Island > Long/Island
- Short/Island > Long/Island
- Long/ No Island > Long/Island

Which matches results both in (Sprouse, Wagers, & Phillips) and in (Sprouse, Caponigro, Greco, Cecchetto) But also predicts:

- Long/ No Island > Short/Island

Which only matches the results in (Sprouse, Wagers, & Phillips), while (Sprouse, Caponigro, Greco, Cecchetto) show the opposite.

As the parser has several degrees of freedom, it would be important to understand how significant these differences are in the experimental results. I might be misinterpreting the stats. But if they are not significant, the problem for the parser remains, as we should probably expect a Tie). Do you have any thoughts on this?

Open questions

I do think that these early results work as a proof of concept, but there are a few things to think about.

- If you think the differences in results reported across experiments are interesting/worth investigating, one hypothesis could be different structural assumptions on the subjects' part? At that point, the parser could give us good insights into what/why this correlates with acceptability differences. It could be good to look at individual differences maybe?
- Relatedly, right now the metrics only account for the surface geometry of the trees. In my dissertation I'm working on incorporating feature differences in the way memory is computed. Depending on the theory of Islands one adopts, including features might change some results (and, depending on how much we want to push it, might also allow us to look at whether/if differences maybe).
- Finally, you mentioned you had a corpus looking more closely at minimal differences across conditions. If we think these results are an interesting start, it might be good to look at such data for a more precise characterization of the effects.