

## SUPPLEMENTARY MATERIALS

## A. MY MESSAGE TO GEOFF

From potts@ling.ucsc.edu Tue Dec 24 05:33:49 2002  
 Date: Tue, 24 Dec 2002 05:33:48 -0800 (PST)  
 From: Christopher Potts <potts@ling.ucsc.edu>  
 To: Geoff Pullum <pullum@ling.ucsc.edu>  
 Subject: Long-distance left extern

Geoff!

Although he sometimes retreated to a stance of pure practicality, Feynman gave answers to these questions, philosophical and unscientific though he knew they were. —James Gleick. 1992. *Genius: The Life and Science of Richard Feynman*. New York: Vintage Books (p. 13).

In addition to sporting this rarity, it's a terrific biography. I have an extra copy if you'd like it for the trip back from Atlanta. I read it many years ago, many years before I knew to look for long distance fronted PP left externs.

—Chris

## B. GEOFF'S MESSAGE

From: "Geoffrey K. Pullum" <gpullum@ling.ed.ac.uk>  
 Subject: a case from speech  
 Date: April 12, 2011 at 00:55:05 PDT  
 To: cgpotts@stanford.edu

Just a minute ago I heard someone speaking on BBC Radio 4 (not reading from a script) say something about the European Court of Human Rights that began:

Unpopular though I can well see that it might be, ...

At last, confirmation of the unboundedness from speech! The predicate preposing of the "happy though I am" PP construction is, indeed, an unbounded dependency.

I've been looking for good attested examples of that sort for about ten years, as you well know. You found the first one, in Gleick's biography of Feynman, when you were a puppy. But from spontaneous speech! This is a red letter day for evidence-based linguistics.

Best wishes,  
 GKP

### C. INITIAL REGULAR EXPRESSION

The following is the regex I used to create initial samples to annotate (see Section 4.2, step 1):

```

1 import re
2 main_regex = re.compile(r"""
3     (\S+)
4     \s+
5     (? :though|as)
6     \s+
7     (? :\S+\s+)+""", re.VERBOSE | re.I)

```

I add an additional step of filtering off examples where the case-normalized initial matching group is in the set {as, even, but, and}. As far as I can tell, the only risk this runs is in filtering off cases like *Even though the odds were, we still lost*. This seems preferable to ending up with samples that are totally dominated by phrases headed with *even though*.

### D. PIPP CLASSIFIER

I ended up annotating and collecting a lot of examples as part of doing the work for Sections 3–4. At a certain point, I realized I had collected enough examples that it seemed plausible that I would be able to train a classifier to help me find more useful examples. This turned out to be extremely productive.

To build this classifier, I began with BERT-base-cased parameters. I had 7,043 annotated examples: 6,598 negative cases and 445 positive cases. To get a sense for how effective these models could be, I first split this dataset randomly into 80% train and 20% test examples, which resulted in a test set containing 1,314 negative cases and 95 positive cases – very close to the overall distribution. I trained the model using a HuggingFace transformers training protocol (Wolf et al. 2019), with the only departure from the defaults being that I assigned the positive class twice the weight of the negative class to help make up for the class imbalance. I trained the model using a Google Colab notebook with GPU support. This took about 15 minutes.

This model achieved an F1 of 0.99 for the negative class and 0.92 for the positive class, for a macro F1 of 0.96. This suggested to me that this classifier approach could be successful, so I retrained the model on all my labeled examples using the same protocol. As a check, I evaluated this model on every PiPP and PP case from my Section 5.1.3 materials, in every prepositional head and embedding variant. This led to a global accuracy of 0.99 (two mistakes in the entire test). On the stress-test case used in Appendix H, the model also got 17 of 18 examples correct. Overall, the results struck me as excellent, so I began using the model to find more examples.

The code for the model is included in the project repository, and the model is available on the HuggingFace Model Hub as `cgpotts/pipp-finder-bert-base-cased`. This may be the most obscure and specific model on the

HuggingFace Model Hub right now.

E. NATURALLY OCCURRING PIPPS SPANNING FINITE CLAUSE BOUNDARIES

1. <sup>B</sup> Honourable though I am sure his intentions were, he betrayed you, Ruben.
2. <sup>B</sup> Hold thy explanation, excellent though I'm sure it is!
3. <sup>B</sup> "they 're right here," she told him and, unlikely though she knew it was,  
she couldn't help wishing he'd squeeze her hand – or give her some other  
small token on which to hang all her hope.
4. <sup>B</sup> Eriks [sic] reassurance, heart-felt though she knew it was, did little to ease  
her anxiety over the impending day.
5. <sup>B</sup> Impossible as she thought it would be for anyone else, she swallowed the  
cold coffee and began once again to type.
6. <sup>B</sup> Celibate as he wished they were
7. <sup>B</sup> Weak though I thought they were, when she stamped closer the ground  
shook beneath us.
8. <sup>B</sup> Strange as you may think it is, the soup shortage was part of one . . .
9. <sup>B</sup> Busy as you say you are, I thought you'd be happy I saved you the extra  
work.
10. <sup>B</sup> Guilty though I believe Mars to be
11. <sup>B</sup> Prepared as he thought he was for this confrontation, his knees still  
buckled in anticipation for what Antone said next.
12. <sup>B</sup> Oddly, as wrong as she knew it was for Noah to assume his behavior was  
acceptable, a very small part of her was thrilled by it.
13. <sup>C</sup> We were agog for the Memories of Max Miller and his Life in the  
Theatre, risqué though we feared the Cheeky Chappie might be, from the  
peerless Mr Bill Pertwee, who spared not our blushes and who appeared  
in a Dressing Gown of sorts that almost beggared belief, even though  
redeemed in full by the familiar hat & patter.
14. <sup>C</sup> I will remember that the collective wisdom of gardening and knowledge  
of plants is much bigger than my knowledge of gardening, vast though  
I think it might be, and therefore, it is possible for a new gardener to  
encounter some new wisdom or knowledge that I know nothing about.
15. <sup>C</sup> That, for me, is what the whole thing is about, anachronistic though I fear  
it may be.
16. <sup>C</sup> They are many and, complex though I like to think I am, I am not legion.
17. <sup>C</sup> In my judgment the Secretary of Statement made it quite sufficiently clear  
that what she was doing was simply reaching a different judgment about  
the degree of harm, significant though she agreed that it was.
18. <sup>C</sup> Amen to that, optimistic though I fear it may be.
19. <sup>C</sup> He was trying to make a noise; to ward something off or drown something  
out—what, I could not imagine, awesome though I felt it must be.
20. <sup>C</sup> You must understand how embarrassing it was to discover this melon  
felony, inadvertent though I assure you it was!
21. <sup>C</sup> We reached for our cameras, inadequate though we knew they would be.

22. <sup>c</sup> I am hoping that there will be a second referendum, tedious though I know this will be.
23. <sup>c</sup> So, foolish though I think you are, marginal though I know you to be, you people do indeed manage to do far more real damage than any sensible person would suspect on a superficial looksee.
24. <sup>c</sup> However, early though we thought our arrival at the gate was, many had thought to come much earlier and the line was already halfway up the Mall.
25. <sup>c</sup> You are not like him, Elijah, as much as you seem to think you are.
26. <sup>c</sup> As inept as Giovanni and Carwyn seemed to think Lorenzo was about technology, why did he have a financial guy who had online access in his super-secret bad guy lair?
27. <sup>c</sup> Prepared as we think we are, we're defeated.
28. <sup>c</sup> Beautiful as many people may think it is, it is such a literary design, obviously made for text use, that it always seems out of place in display sizes.
29. <sup>c</sup> Pointless as you think it will be, this post is actually worth spending the two minutes that it will take to read, because these are so obvious that they're most often overlooked.
30. <sup>c</sup> Crazy as some might think that is, the customer service was incredible and the product even more so.
31. <sup>c</sup> Difficult as they say it will be, Inshā' Allāh it will be better than every Ramaḍān before.
32. <sup>c</sup> Unightly as some say it is, wind farming is more sustainable compared to the generating energy using fossil fuel.
33. <sup>c</sup> prejudiced as the test seems to say they are.
34. <sup>c</sup> Smart though we might think we are, we do not have access to all information, nor are we experts in every field of human endeavor, so we have to rely on others.
35. <sup>c</sup> Clean as you think your hands are, they usually have dirt and bacteria from items you've touched.
36. <sup>c</sup> Happy as I want you to be and hope you will be, you must yet understand that marriage is God's design and His purposes must be pursued in order for you to be truly happy.
37. <sup>c</sup> Neither as clever nor as interesting as it appears to think it is, The Words maroons its talented stars in an overly complex, dramatically inert literary thriller that's ultimately a poor substitute for a good book.
38. <sup>c</sup> Because as terrible a thought as she knows it is, Irene is just so damn tempting.
39. <sup>c</sup> For as small and insignificant as some other luxury and green car brands might want us to think Tesla is, the name keeps popping up all over the news, and even when it's about Tesla's losses in the courtroom, most of the American public seems to side with Tesla.

40. <sup>◦</sup> For as independent and fierce as Miley Cyrus wants everyone to think she is, she's actually pretty scared of being alone – and as a result, she's starting to blame herself for Liam Hemsworth's alleged wandering eye.
41. <sup>◦</sup> So as safe as the CSE promoters want you to believe the process is, there is a significant potential for both loss and inconvenience— enough so that managed municipal, corporate, and government CEFs, REITs, preferred stocks, etc.
42. <sup>◦</sup> Clever as I thought my nerd humour was, it would appear that this was a bit of a recurring joke among nerd tweeps and just proved once again that there is no original thought.
43. <sup>◦</sup> As entertaining as we may think we are everyone in the room wants to be in their own area getting ready for kids.
44. <sup>◦</sup> Ardent, as pretentious and intelligent as you want to keep trying to tell people u are, u should know what AI is and how far Siri is from AI.
45. <sup>◦</sup> As exceptional as your client wants to think you are, you must come in second to him.
46. <sup>◦</sup> Because as terrible a thought as she knows it is, Irene is just so damn tempting.
47. <sup>◦</sup> As odd as I know it sounds I think that adding a few extra bits is part of the experience of this set.
48. <sup>◦</sup> As smart as I think I am now I know that it is only because I had to walk some long and hard steps to get here.
49. <sup>◦</sup> It is my understanding (as limited as you seem to think it is) is that trustees (perhaps not the names you've mentioned) were the ones who decided to create an investment office and the parameters of the office's responsibilities.
50. <sup>◦</sup> As crazy as I thought it sounded I told him it was fine with me.
51. <sup>◦</sup> As anonymous as I thought I wanted to be, there was something about being there and singled out for anything other than panhandling that felt like a casting call.
52. <sup>◦</sup> As practical as I can imagine being sighted must be I sometimes think sight can be a very big obstacle.
53. <sup>◦</sup> However, as secure as many people want to believe blockchain is, it is not without its vulnerabilities.
54. <sup>◦</sup> As busy as I know Kelly will want to be on her first day at Gonzaga, . . . Hoopfest must come first.
55. <sup>◦</sup> As profound and self-important as Arrival appears to think it is, your humble reviewer found it a yawn-inducing snoozefest that borrowed from every science fiction film from The Day the Earth Stood Still to Close Encounters of the Third Kind to ET to Contact.
56. <sup>◦</sup> Well-intentioned though many people may have imagined that the CIA probably thought they were, their foreign-policy operations were confused, duplicitous failures.
57. <sup>◦</sup> As for planning, as sinister as I think this student thinks our meetings may be, they are really not!

58. <sup>c</sup> As much of a downer as I think we both agree the pistols are, for us, do you not find the only thing worse than using one yourself is when someone else in the lobby absolutely dominates with them, when running them akimbo?

## F. ADDITIONAL LARGE LANGUAGE MODEL RESULTS

Figure 5 shows long-distance dependency results for Pythia 70M, the smallest model in the Pythia series. Figure 6 shows the results for Pythia 410M, the smallest model to show the full set of effects reported in Figure 2 in the main text. Results for the full set of Pythia models are included in the code repository for this paper.

## G. TESTING PREPOSITIONAL-HEAD EFFECTS WITH AUTOREGRESSIVE LMS

In Section 5.2, I moved from autoregressive language models to masked language models in order to probe for prepositional-head effects. My primary reason for doing this is that the prepositional head occurs too early to fully identify the construction as a PiPP. Masked language models use the entire bidirectional context and so do not impose this limitation.

An anonymous reviewer suggested an alternative design that allows us to continue using autoregressive language models to test for prepositional-head effects. The key idea is to use paradigms like the following:

- (43) a. The food smelled fresh though it was. (though/+gap)  
 b. \*The food smelled fresh although it was. (although/+gap)
- (44) a. The food smelled fresh though it was **very** old. (though/-gap)  
 b. The food smelled fresh although it was **very** old. (although/-gap)

The main comparison is in (43), which is a +gap condition. Example (43a) can be parsed as a PiPP, whereas (43b) cannot. We can test for this by comparing the surprisals for the period at the end. This should be very high for (43b) and comparatively low for (43a). The examples in (44) serve as controls here; both are fully grammatical (with non-PiPP parses), so we expect any differences between them, as measured by the surprisal for *very*, to trace to baseline differences in specific examples.

I created ten example paradigms of the sort seen in (43)–(44). These are included in the code repository for this paper (in the file `materials-autoregressive-prepeffects.txt`). Figure 7 summarizes the wh-effects analysis (using the same methods as in Section 5.1) for the smallest and largest Pythia models. The results clearly support the claim that these models are sensitive to the prepositional-head restrictions for PiPPs.

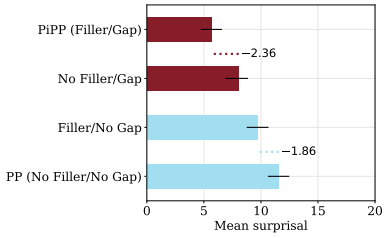
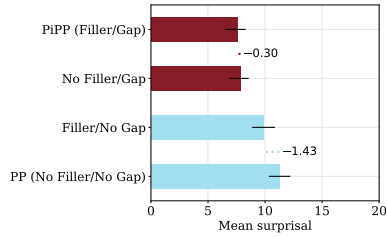
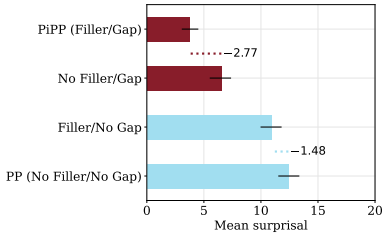
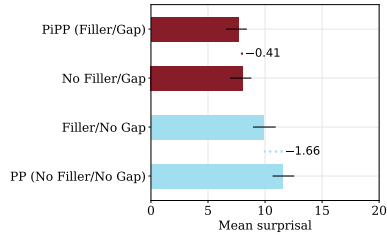
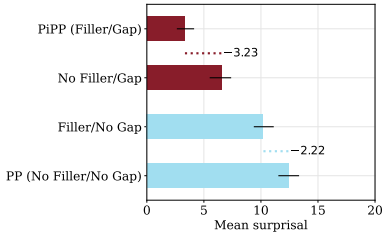
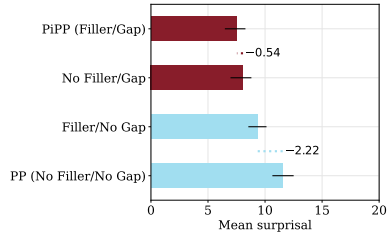
(a) Single clause, *though*-headed.(b) Multi-clause, *though*-headed.(c) Single clause, *as*-headed.(d) Multi-clause, *as*-headed.(e) Single clause, *as...as*-headed.(f) Multi-clause, *as...as*-headed.

Figure 5: Testing wh-effects for Pythia 70M, the smallest model in the original Pythia series. The +gap effects (red bars) are reasonably clear, especially for the single-clause cases, but the -gap effects (blue bars) are not in the expected direction.

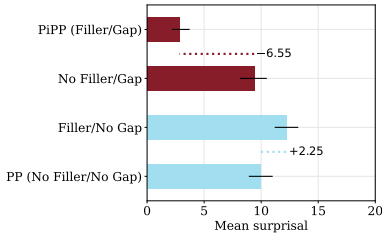
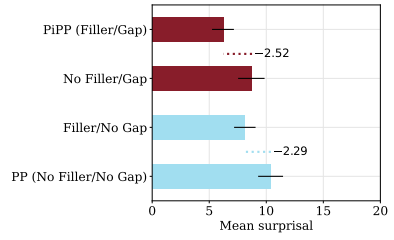
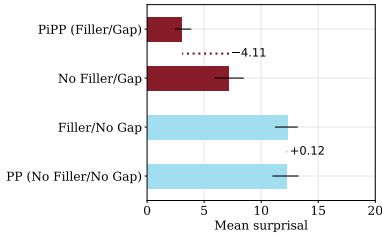
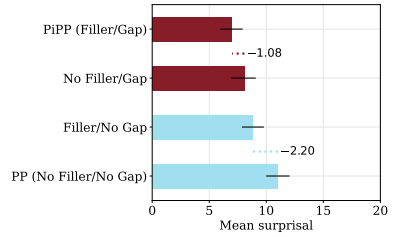
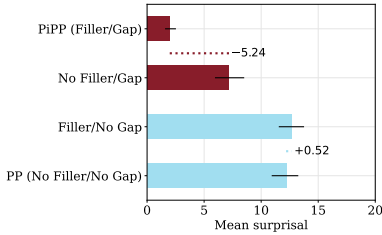
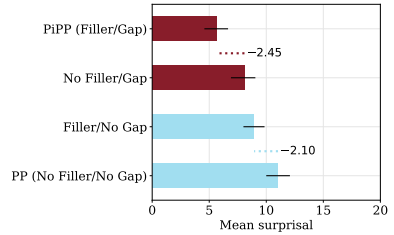
(a) Single clause, *though*-headed.(b) Multi-clause, *though*-headed.(c) Single clause, *as*-headed.(d) Multi-clause, *as*-headed.(e) Single clause, *as...as*-headed.(f) Multi-clause, *as...as*-headed.

Figure 6: Testing *wh*-effects for Pythia 410M. This is the smallest model in the Pythia series to show the same qualitative patterns that we see for the 12B model (Figure 2).



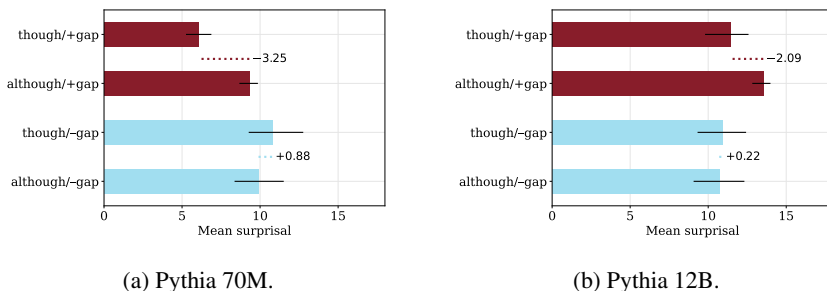


Figure 7: Prepositional-head effects in autoregressive language models. The models show high surprisal for PiPP gaps where the preposition is *although*, compared with such gaps where the preceding preposition is *though* (red bars). The *-gap* cases are included as controls; these examples cannot easily be parsed as involving PiPPs, but rather only as involving concessive adjunct clauses that can be headed by both *though* and *although*.

## H. LLMs AS LINGUISTS

Can LLMs reliably transform PPs into PiPPs? This is an unusual task, and so it should not be considered a prerequisite for mastering PiPPs, but a positive result here seems like it would be informative. To begin such an assessment, I ran some small pilots with the GPT-3 models `text-davinci-001` and `text-davinci-003`. It should be emphasized that, in these experiments, the models are frozen objects. To the extent that they “learn”, it is entirely the result of the prompt placing them in a particular temporary state. This is what the NLP literature refers to as *few-shot in-context learning* (Brown et al. 2020).

The primary materials for my pilot are the 33 base sentences from Section 5.1.3. As before, we can automatically create variants of these basic sentences with different prepositions and different levels of embedding.

The prompt to the LLM includes some high-level instructions, and then it offers some number of demonstrations of the intended behavior: translating PPs into PiPPs. The demonstrations are drawn from the experimental materials, always disjoint from the target item. After the demonstrations, I include the instruction `Now apply the transformation to this input:`, a novel input, and the string `Output:`. Here is a toy example of the prompt, using two invented short examples as demonstrations, and a toy target case:

---

You are an expert grammarian. Your task is to convert the Input example to a new Output by applying a transformation. Here are some examples of these transformations:

Input: `Though they were happy, they said no.`  
 Output: `Happy though they were, they said no.`

Input: I am, though it seems odd, friends with a robot.  
 Output: I am, odd though it seems, friends with a robot.

Now apply the transformation to this input:

Input: Though they felt sad, they smiled.  
 Output:

The model’s entire continuation (with peripheral whitespace removed) is taken to be its prediction, and we say the model is correct if and only if its prediction is an exact string match (EM) with the gold PiPP. For the above, the correct output would be *Sad though they felt, they smiled*.

Table 2 summarizes the results. The `text-davinci-001` engine struggles to perform the task, but `text-davinci-003` is outstanding at it. For that model, the assessment is perhaps unfairly strict, as the three cases that were marked as incorrect are the following, in which the model created a well-formed PiPP and happened also to change the position of the entire PiPP in the string:

- (45) PP We liked the end of the movie, as they said that we knew that it was tragic.  
 Gold We liked the end of the movie, as tragic as they said that we knew that it was.  
 Pred As tragic as they said that we knew that it was, we liked the end of the movie.
- (46) PP The proposal is still being assessed, as they said that we knew that it seemed inspired.  
 Gold The proposal is still being assessed, as inspired as they said that we knew that it seemed.  
 Pred As inspired as they said that we knew that it seemed, the proposal is still being assessed.
- (47) PP They skipped the movie, as they said that we knew that it seemed exciting.  
 Gold They skipped the movie, as exciting as they said that we knew that it seemed.  
 Pred As exciting as they said that we knew that it seemed, they skipped the movie.

In these materials, the fronted material is always a single adjective. To assess whether models could perform the PiPP transformation on a wider range of constituents, I created nine additional “stress test” cases. These are included in the code repository for the paper, as `materials-stress-test.csv`. I repeated the above experiments using these items. With demonstrations drawn from the stress-test examples (always disjoint from the target), `text-davinci-003` gets only 3/9 correct. Essentially the same result obtains (2/9) when the demonstrations are drawn from the basic materials (randomly sampling from different preposition types and different embeddings). This suggests that, in the

Engine	Preposition	Embedding	Accuracy (EM)
text-davinci-001	as	None	0.70
	as	they said that we knew that	0.64
	though	None	0.48
	though	they said that we knew that	0.55
	as...as	None	0.70
	as...as	they said that we knew that	0.88
text-davinci-003	as	None	0.67
	as	they said that we knew that	0.64
	though	None	0.70
	though	they said that we knew that	0.70
	as...as	None	0.88
	as...as	they said that we knew that	0.91

Table 2: Assessment of model abilities to transform PPs into PiPPs using only few-shot, in-context learning.

general case, applying this transformation is challenging for these models – as it would be for many people.