Language Models and the Discomfort of Not Knowing

Some pop-philosophy, nuclear physics, and dependency parsing.

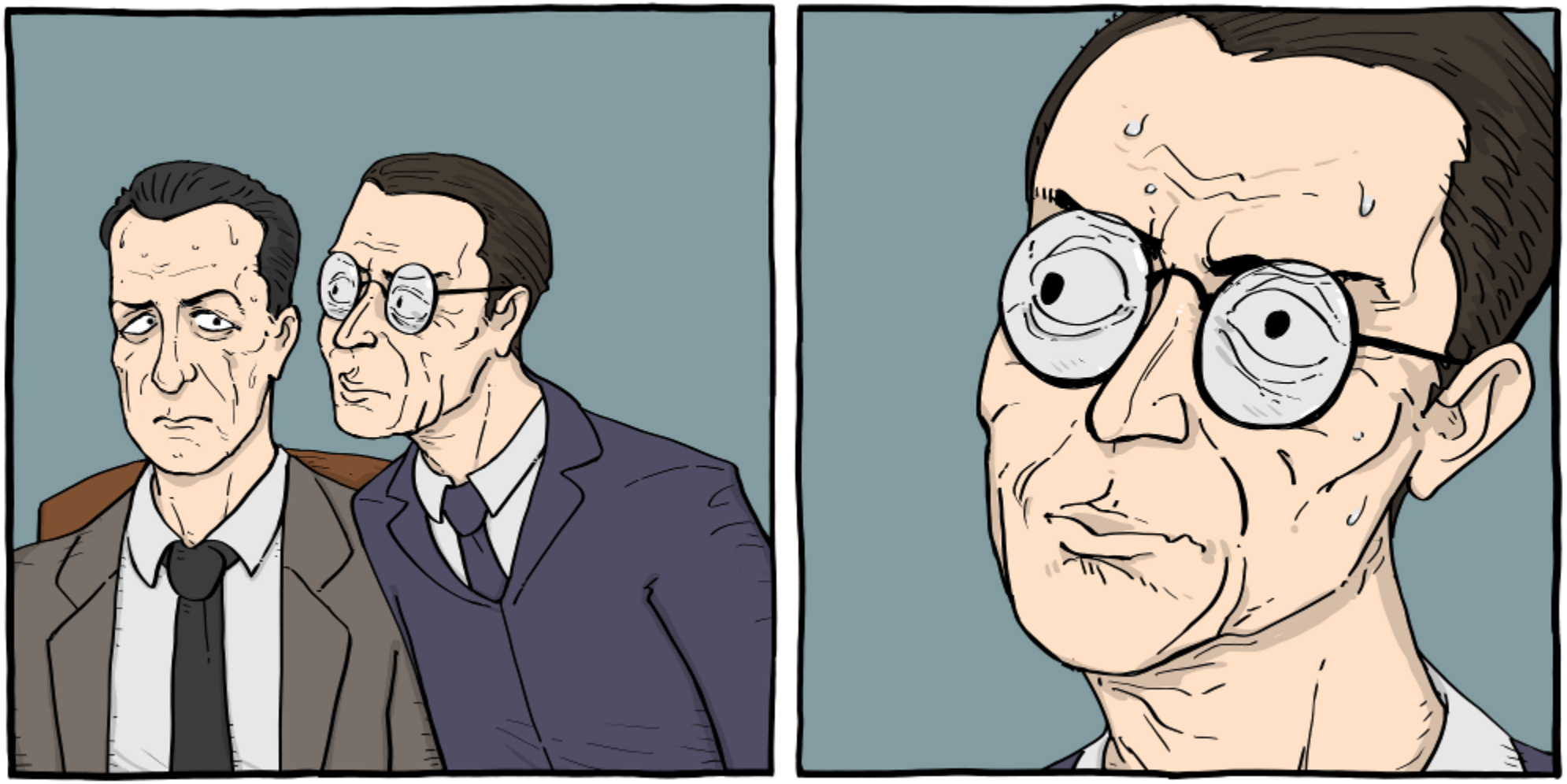
Prelude

Sarte and the "Look" — The discomfort of being the object

Sartre and the "Look"

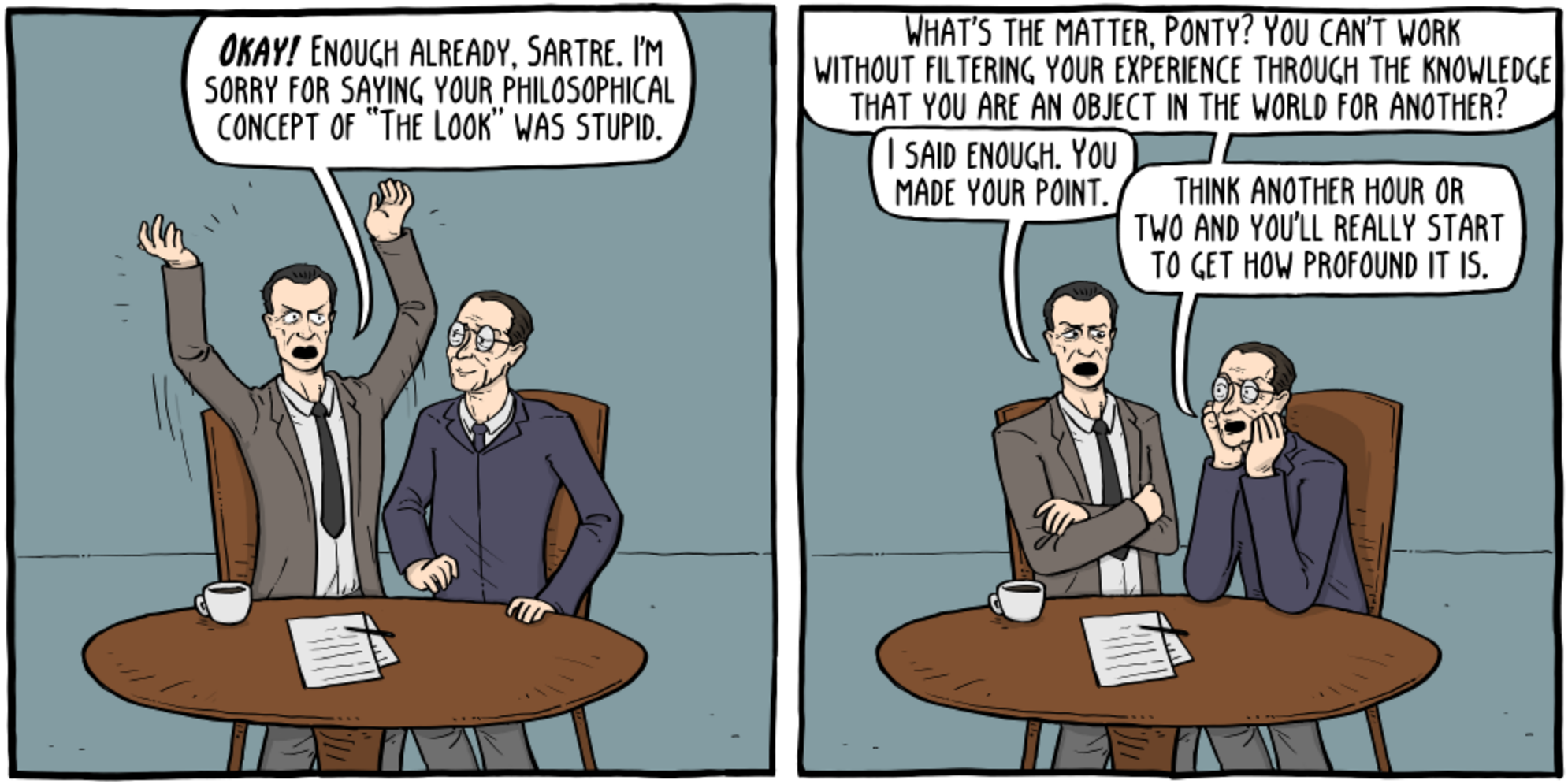
https://existentialcomics.com/comic/235

Sartre and the "Look"



https://existentialcomics.com/comic/235

Sartre and the "Look"



https://existentialcomics.com/comic/235

The look

- On my own, I am the only subject.
- When someone looks at me, my sense of identity collapses.
- I feel shame at being an object (self-conscious)
- Not all bad as I can experience others as subjects (i.e. humans)

Knowing

- It gives me info about myself (e.g. no laughter = humorless)
- But I can't truly know if I am an object or not
- I also can't know how I am experienced as an object, I can just observe the exterior

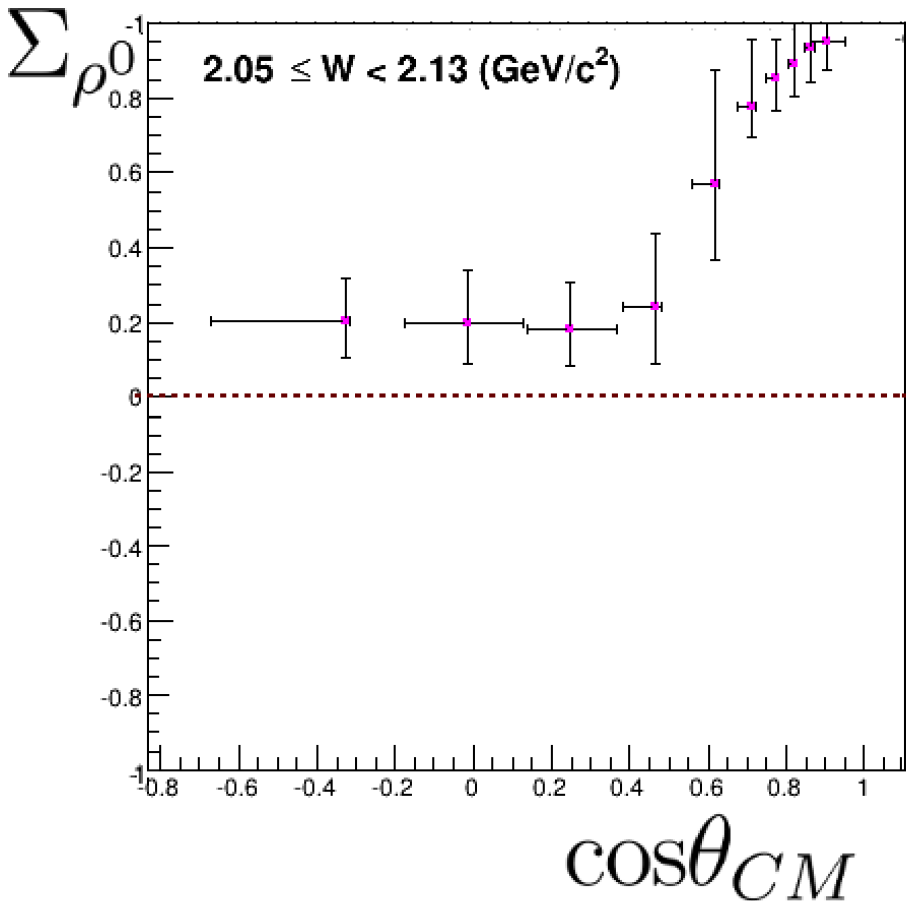
%[//]: <> (you could be looking at me but my performance, my voice could be nothing more than a background noise like the sound of birds or the rumbling of a nearby trikk)

%[//]: <> (or you could be looking at me and observe a metaphorical goose, not golden, or something more dissonant still with how I perceive myself, naturally as the physical manifestation of truth)

Chapter 1

Measuring the impact of unobservable particles – The discomfort of abstraction

Discomfort in abstraction



https://theses.gla.ac.uk/6957/

- Σ measures sensitivity to beam orientation (ish)
- ρ^0 has mean lifetime $4 \times 10^{-24} s$
- $\rho^0 p \rightarrow \pi^- \pi^+ p$ - decay path.
- Measure angular distribution of π^- and π^+ to get Σ.
- Different decay paths ($\Delta^{++}\pi^-$ or $\Delta^0\pi^+$).
- Other mesonic states (f^0)

Discomfort in abstraction

- Σ is a combination of spin density matrix elements SDMEs.
- We measure Σ because we can't
- SDMEs give probabilities of spin states of mesonic resonances (ρ^0) (ish)
- The spin states of mesonic resonances give us info as to what other hadronic resonances may have occurred before the mesonic resonance.
- The existence (or not) of other hadronic states gives us info about the underlying structure of nucleons (protons and neutrons).
- We learn something about the structure of nucleons.
- Nucleonic mass accounts for about 98% of mass.

Discomfort in abstraction

Allows us to update models and likelihood of certain intermediate states *indirectly*.

Does **not** allow us to say anything directly about nucleonic structure!

Chapter 2

Evaluating the syntactic "knowledge" of models – The discomfort of reification

Discomfort of reification

Originally intended to help highlight students lagging behind a curriculum.

IQ scores — an attempt to distill a complex, multifaceted abstract idea into a single number.

This then results in treating intelligence as a thing, an intentity that exists.

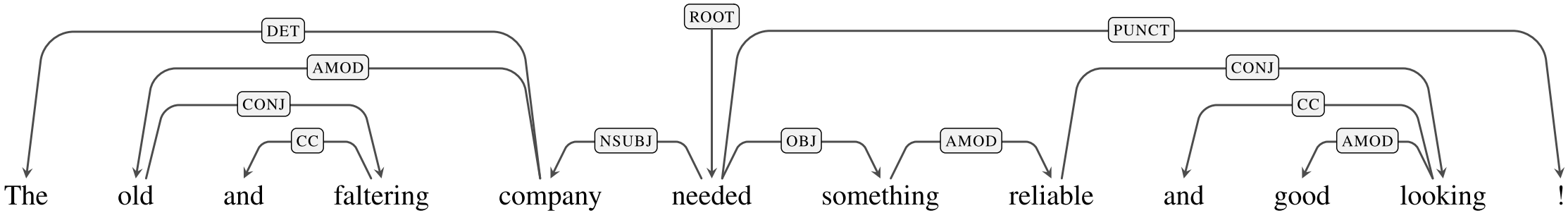
And typically is considered fixed.

Leads to a lot of nonsense.

Syntax — dependency parsing

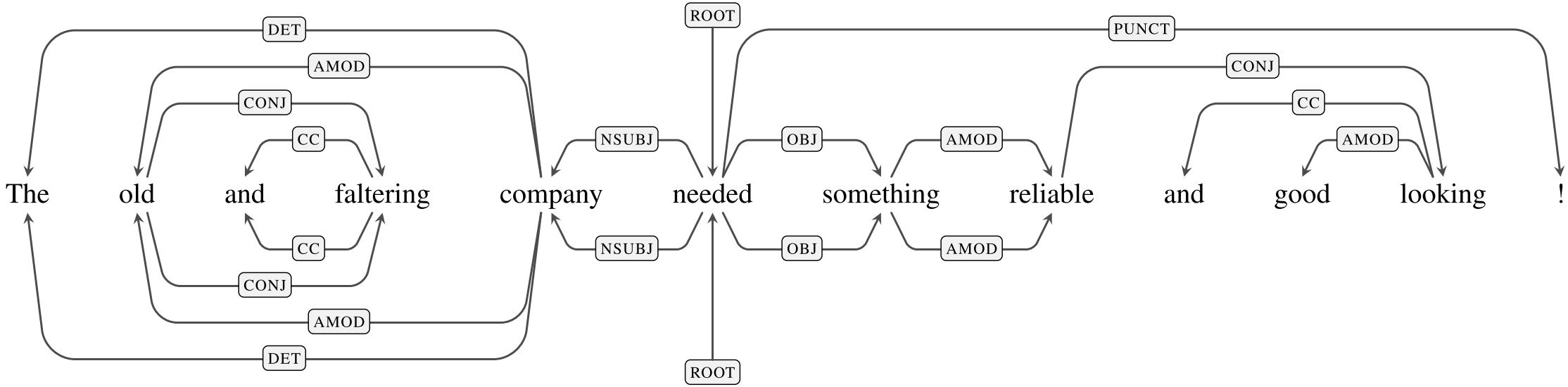
Attachment scores reifying syntax

Syntactic errors



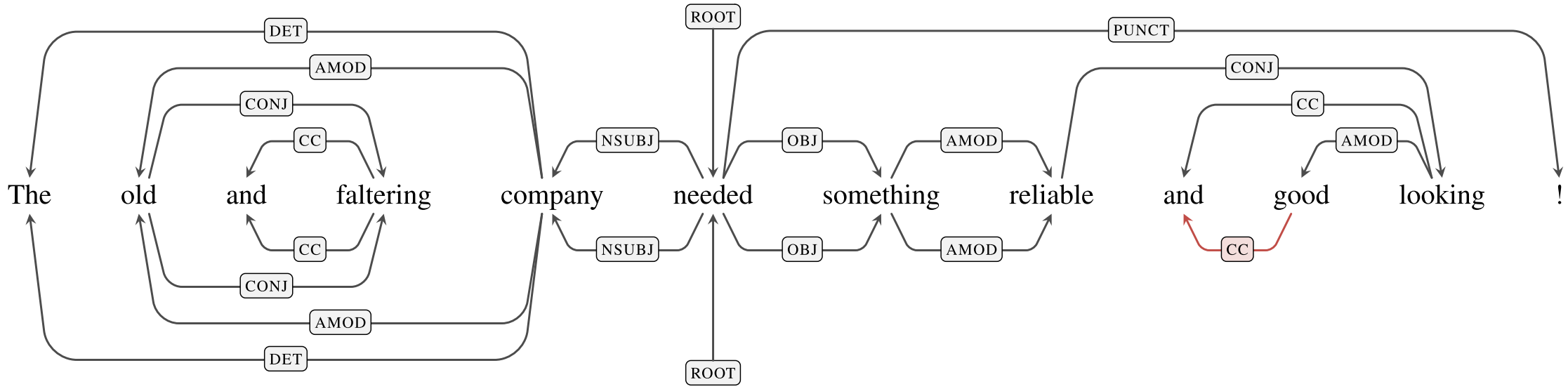
- UDPipe 2.10 (https://lindat.mff.cuni.cz/services/udpipe/)
- English-EWT
- UAS: 92.46, LAS: 90.08

Syntactic errors



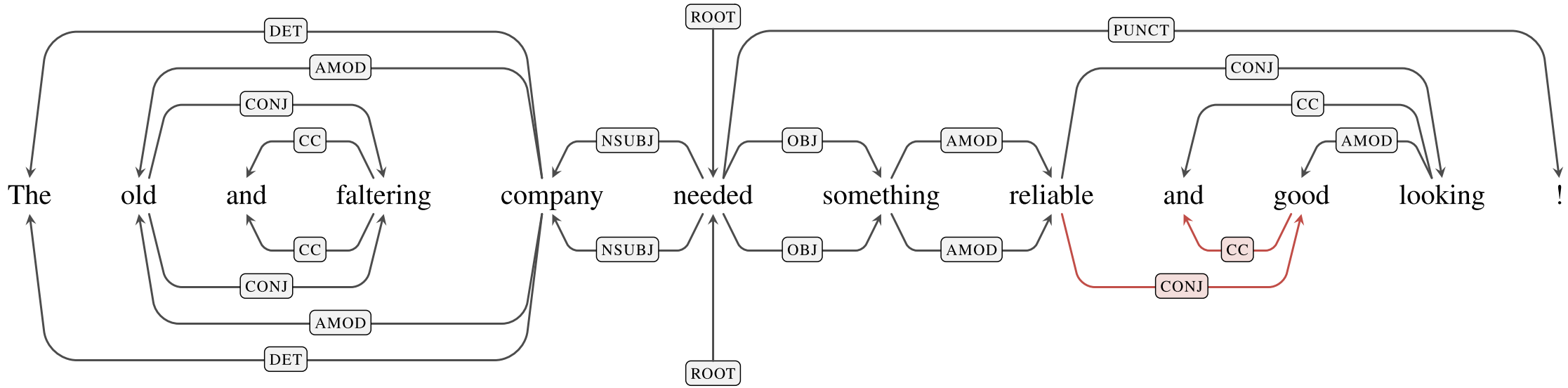
- UDPipe 2.10 (https://lindat.mff.cuni.cz/services/udpipe/)
- English-EWT
- UAS: 92.46, LAS: 90.08

Syntactic errors



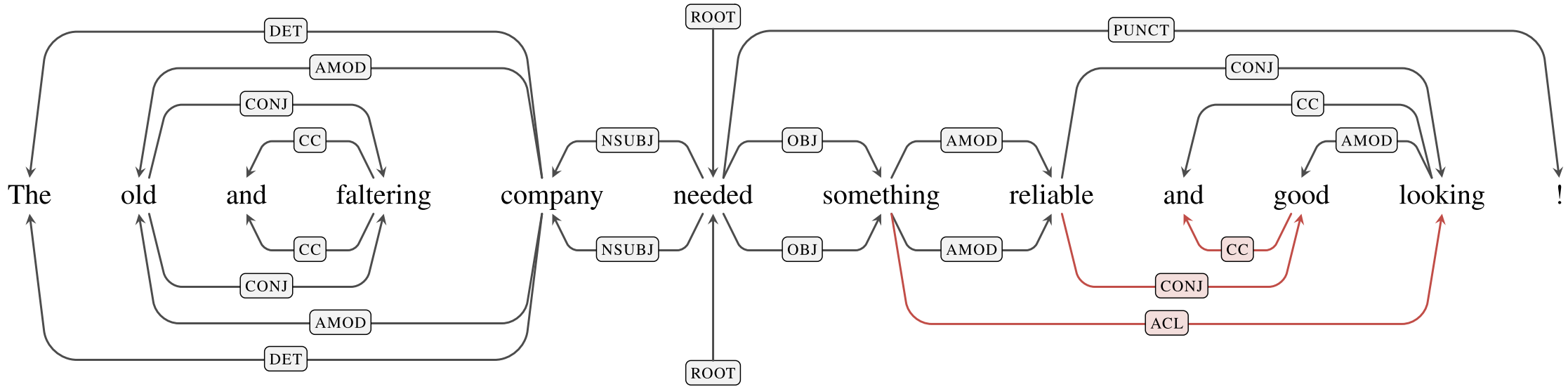
- Conjunction has wrong head.

Syntactic errors



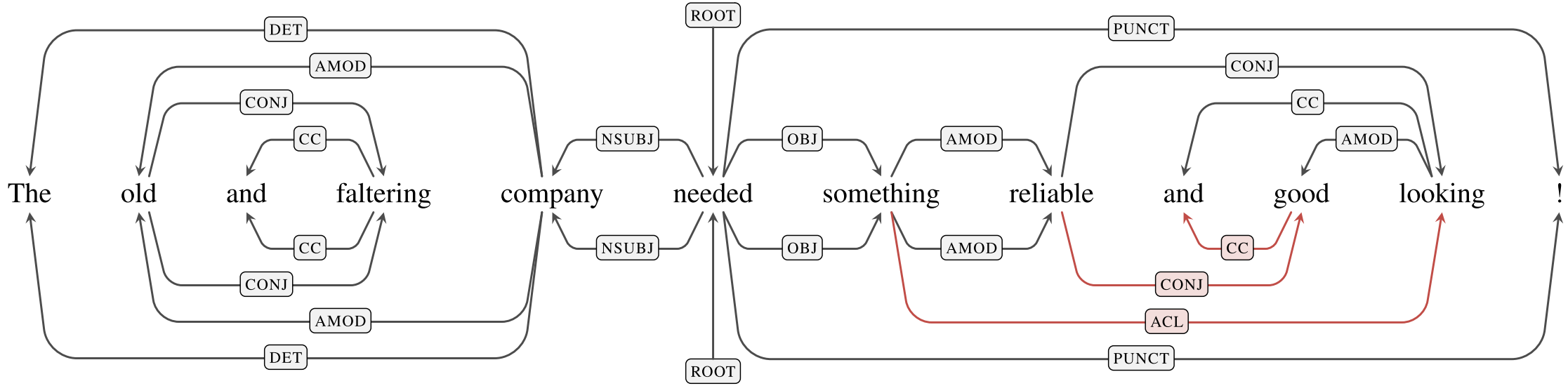
- Conjunction has wrong head.
- Because "good" is predicted as a conjunct to "reliable"

Syntactic errors



- Conjunction has wrong head.
- Because "good" is predicted as a conjunct to "reliable"
- ACL: adnominal clause/clausal modifier of noun, i.e. "something" is doing some "looking"

Syntactic errors



- Tree UAS: 75.00
- Treebank UAS: 92.46
- *something reliable and of good quality that is looking at something unspecified ≠ something reliable and aesthetically pleasing*

Evaluation

	UAS
Norsk-Bokmål	0.944
Cymraeg	0.879
Gallego	0.873
Αρχαία ελληνική	0.803

- UAS$= \frac{correct}{gold+predicted}$
- correct — number of tokens with correctly predicted head
- gold — number of tokens in treebank
- predicted — number of predicted tokens (gold $\neq$ predicted if tokenization stage)

Evaluation

	UAS	$^{(\text{tokens-edit})}/_{\text{tokens}}$	$^{\text{edit}}/_{\text{tokens}}$
Norsk-Bokmål	0.944	0.944	0.056
Cymraeg	0.879	0.879	0.121
Gallego	0.873	0.873	0.127
Αρχαία ελληνική	0.803	0.803	0.197

Edit distance

$$\mathrm{ED}_{\text{tree}} = \sum_{g \in G, p \in P} \mathrm{ED}(g, p)$$

$$\mathrm{ED}(g, p) = \begin{cases} 1 & \text{if } g_{\text{head}} \neq p_{\text{head}} \\ 0 & \text{otherwise.} \end{cases}$$

Evaluation

	UAS	$(tokens-edit)/tokens$	$edit/tokens$	$resistance/tokens$
Norsk-Bokmål	0.944	0.944	0.056	0.180
Cymraeg	0.879	0.879	0.121	0.488
Gallego	0.873	0.873	0.127	0.563
Αρχαία ελληνική	0.803	0.803	0.197	0.365

Resistance (perturbation) distance

- Less locally focused than edit distance
- Useful to detect perturbations on strongly connected subgraphs
- Treats edges as resistors and compares "resistance" of graphs.

Evaluation

	UAS	$(tokens-edit)/tokens$	$edit/tokens$	$resistance/tokens$	$exact\_trees/trees$
Norsk-Bokmål	0.944	0.944	0.056	0.180	0.655
Cymraeg	0.879	0.879	0.121	0.488	0.336
Gallego	0.873	0.873	0.127	0.563	0.103
Αρχαία ελληνική	0.803	0.803	0.197	0.365	0.245

Exact tree score

- Simply the proportion of trees that have no incorrectly predicted heads.

Rankings

Edit distance	Resistance distance	Exact trees
Norsk-Bokmål	Norsk-Bokmål	Norsk-Bokmål
Cymraeg	Αρχαία ελληνική	Cymraeg
Gallego	Cymraeg	Αρχαία ελληνική
Αρχαία ελληνική	Gallego	Gallego

Epilogue

Two Counterpoints

Literature on detailed syntactic experiments and LMs

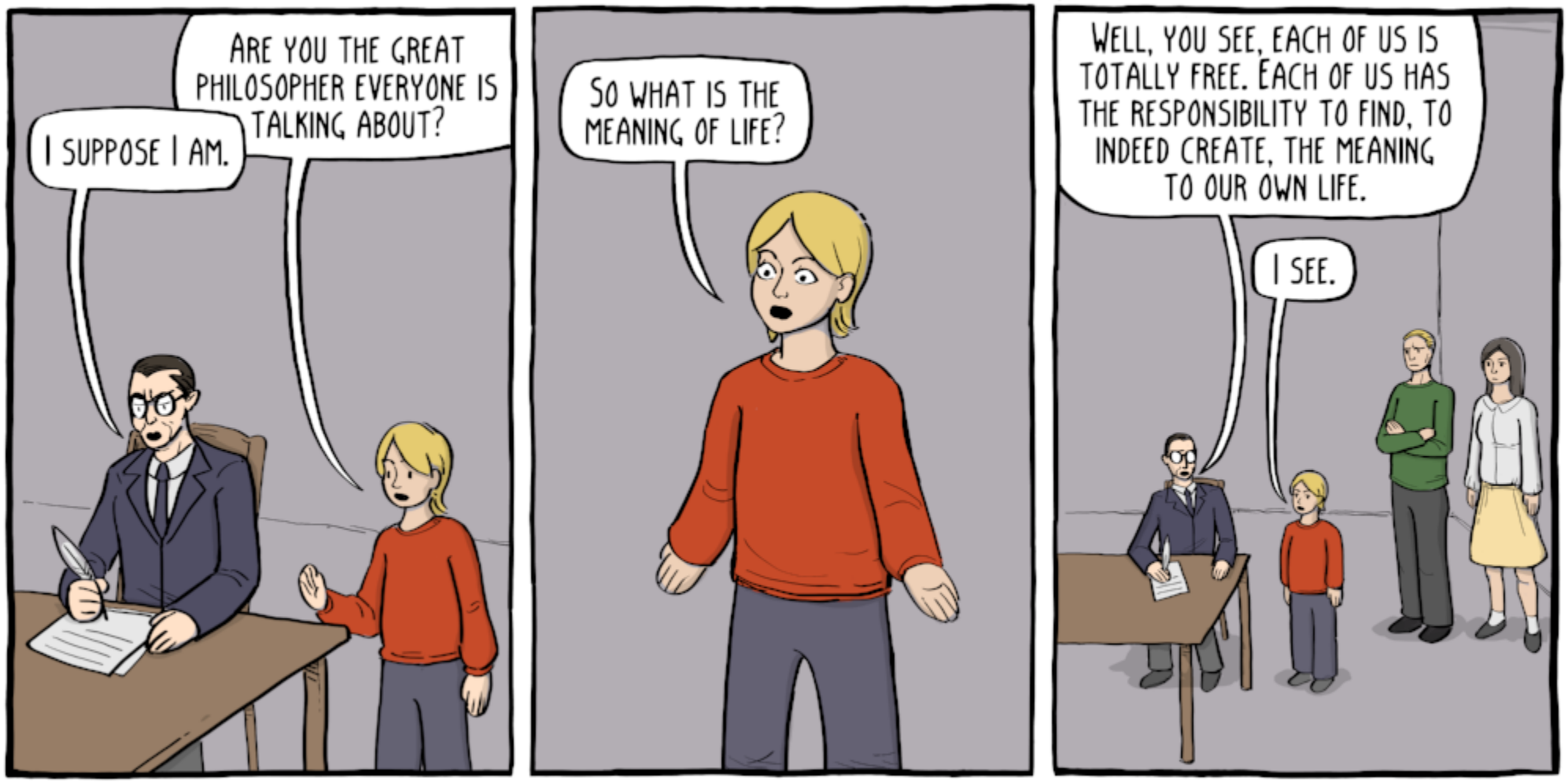
Evidence for LMs encoding syntax

- Deeper, hierarchical syntactic structures but RNNs (Gulordava et al. 2018)
- Handle subset of filler-gap dependency restrictions, i.e. empty syntactic positions - again RNNs (Wilcox et al. 2017).
- LMs (RNNs/LSTMs) encode basic syntactic state but not more complex, fine-grained details (Futrell et al. 2019).
- BERT/ElMo represent "classical NLP pipeline" (Tenney et al. 2019/Peters et al. 2018).
- LMs encode something regarding licensing contexts and corresponding negative polarity items (Jumelet and Hupkes, 2018).
- BERT does well on a number of simple syntactic tests (Goldberg, 2019).
- Some evidence that attention mechanism encodes some syntactic information (Clark et al. 2019).

Evidence against

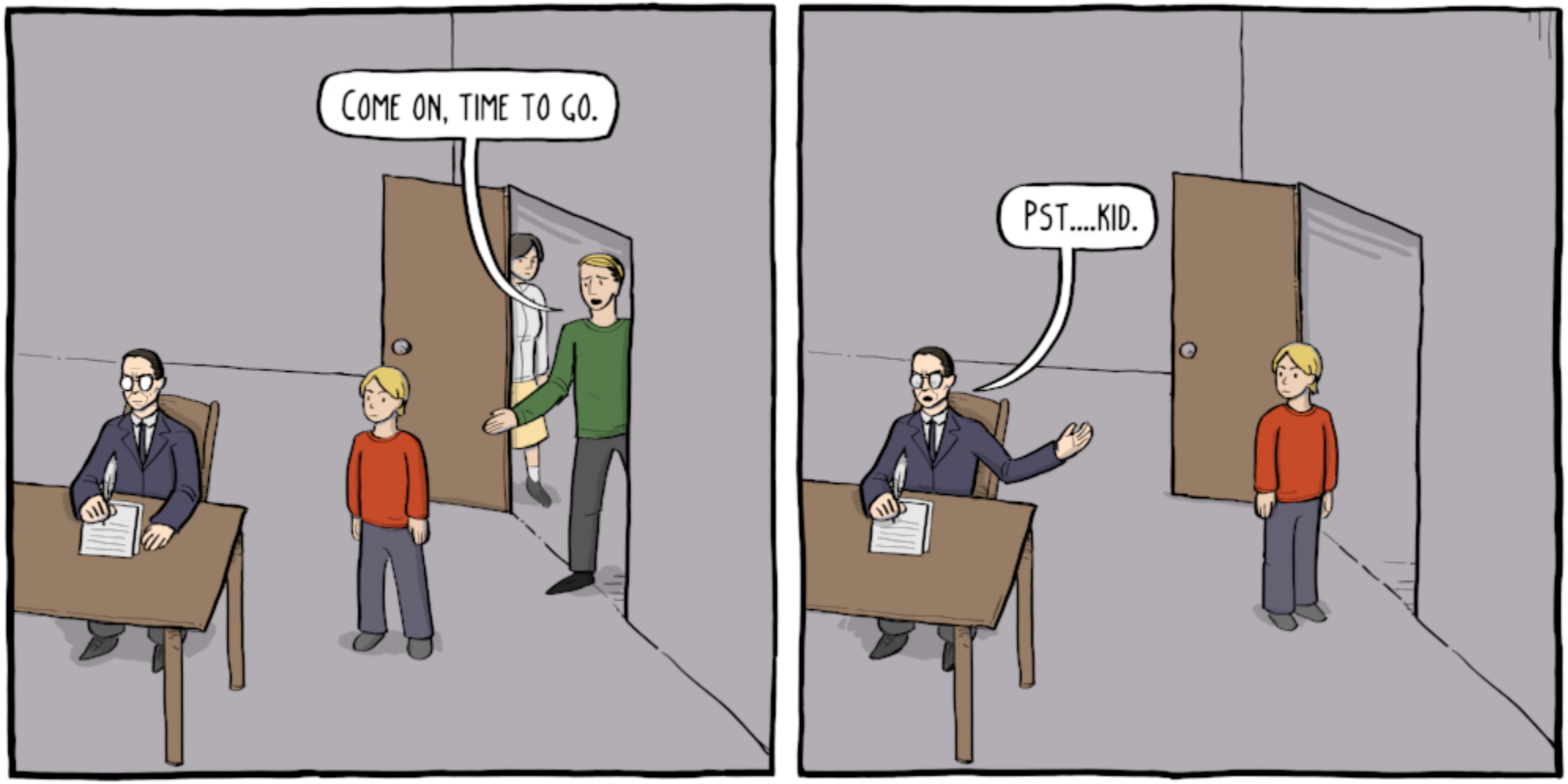
- LMs w/o explicit signal higher error rate for verb subject number agreement (Linzen et al. 2016)
- licensing reflexive pronouns and negative polarity but RNNs (Futrell et al. 2018)
- LMs with and without explicit signals struggled to differentiate between grammatical and ungrammatical sentences (Linzen et al. 2018)
- Architecture impacts LM's ability wrt syntactic generalisation and perpelexity of LMs not strongly associated with this ability (Hu et al. 2020)
- LMs rely on "fallible syntactic heuristics" (McCoy et al. 2019).
- LMs substantial worse than humans on predicting grammaticality of sentences (Warstadt et al. 2019).
- Doubt cast on probing metrics for measuring syntactic encoding in LMs by using pseudowords to generate syntactically valid sentences without meaningful lexical units (Maudslay and Cotterell, 2021).

Sartre and the meaning of life



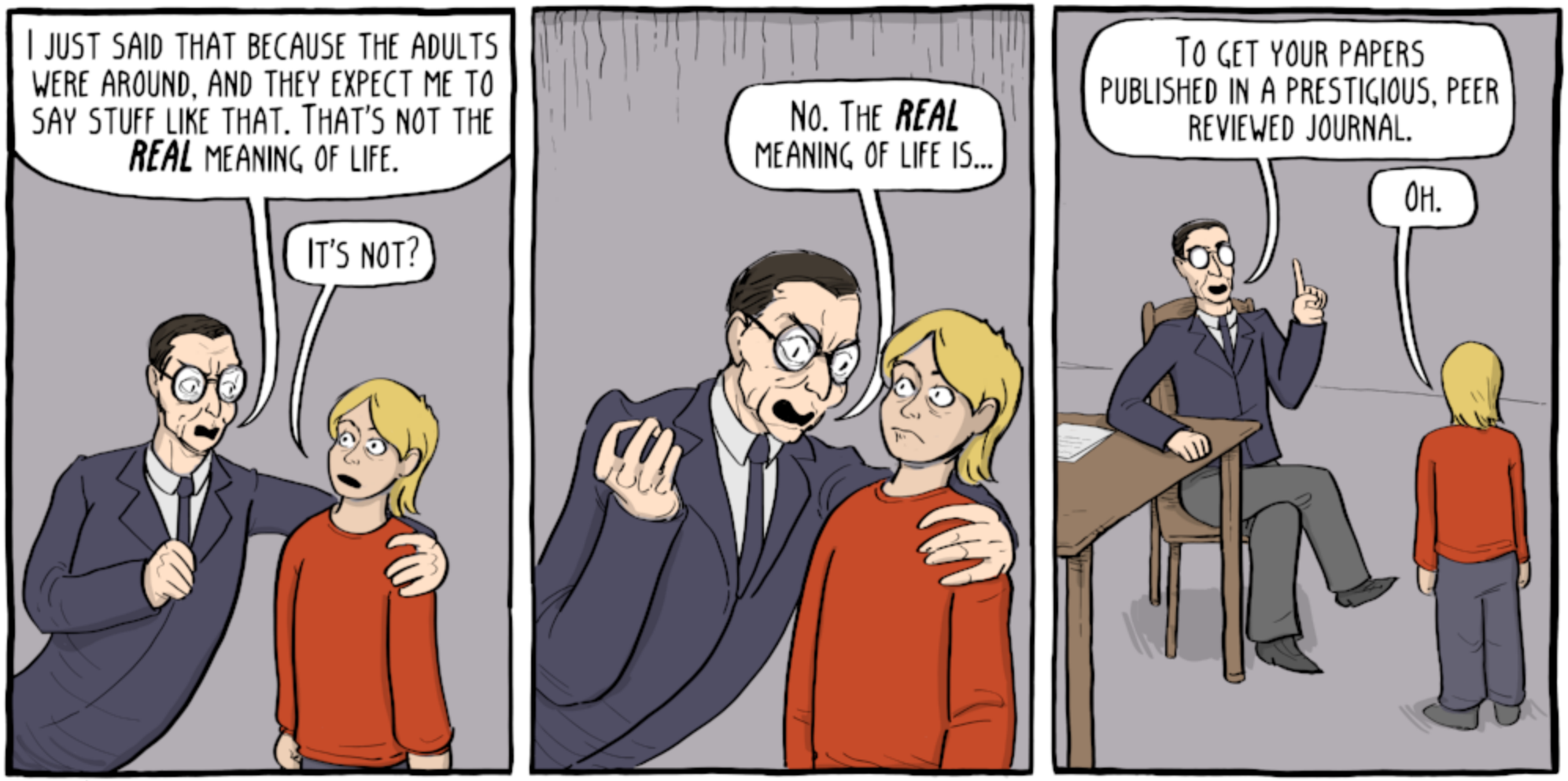
https://existentialcomics.com/comic/197

Sartre and the meaning of life



https://existentialcomics.com/comic/197

Sartre and the meaning of life



https://existentialcomics.com/comic/197