(D) Pay Attention (1/2) [10 Points]

The meaning of a word depends on its context. For example, in the sentence "The farmer *seeded* the field with corn," the word *seeded* means "added seeds to." However, in the sentence "The chef seeded the tomato," the word *seeded* means "took seeds away from." Similarly, *bat* means different things in "The baseball player swung the *bat*" vs. "The *bat* flew through the air."

If you were building a model of language, how would you get it to recognize the way that a word's meaning depends on context? One popular technique for achieving this goal is a mechanism called **attention**. In the way that attention is implemented in current state-of-the-art models of language, the model has a large number of **attention heads**, each of which is denoted by a pair of numbers (for examples, 8-10). When the model processes a sentence, for every pair of words in the sentence, each head calculates the "relatedness" of the two words.

The one wrinkle is that we do not know what exactly "relatedness" should mean, so instead of telling the model how to define "relatedness," we let the model learn its own definition of relatedness. Recently, computer scientists have started to analyze what these attention heads have learned, and this analysis shows that they often reflect linguistic information! For example, here's the output of one attention head (head 8-10) when we feed the following sentence into BERT[1], which is one of the most popular models that uses attention heads:

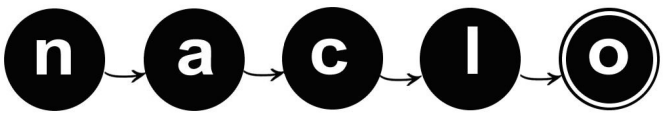
Example: I_1 see_2 my_3 $sister_4$, but_5 she_6 $can't_7$ see_8 me_9 $because_{10}$ she_{11} is_{12} $reading_{13}$ a_{14} $linguistics_{15}$ $book_{16}$.

Output: 4 → 2, 9 → 8, 16 → 13

This output signifies that head 8-10 connects word #4 (*sister*) to word #2 (*see*), as well as word #9 (*me*) to word #8 (*see*) and word #16 (*book*) to word #13 (*reading*). If you consider what all of those pairs of words have in common, you'll see that each one is a verb and its direct object: *sister* is the direct object of the first instance of *see*, *me* is the direct object of the second instance of *see*, and *book* is the direct object of *reading*. It appears that head 8-10 has learn to connect verbs to their objects! (Note that these activations are directional; for example, word #2 is not connected to word #4.)

Why is this information useful? If we go back to the example with the verb *seed*, this sort of information can help the model figure out which version of *seed* is being used: If its direct object is something like *field* or *lawn*, then it probably means "to add seeds to;" if its direct object is something like *tomato* or *watermelon*, then it probably means "to take seeds away from." Of course, one sentence isn't enough to draw strong conclusions. Instead, computer scientists tend to use a *corpus*, or a database, of example sentences to find patterns in the data. On the next page is a small corpus, the NacloWeb Corpus[1], which has 7 sentences.

[1] Some sentences derived from data in the English Web Treebank.



(D) Pay Attention (2/2)

NacloWeb Corpus

1. *My$_1$ experience$_2$ with$_3$ Gelda$_4$'s$_5$ House$_6$ of$_7$ Gelbelgarg$_8$ has$_9$ been$_{10}$ extremely$_{11}$ wonderful$_{12}$*
2. *We$_1$ use$_2$ Google$_3$'s$_4$ models$_5$ to$_6$ delve$_7$ into$_8$ the$_9$ inner$_{10}$ workings$_{11}$ of$_{12}$ language$_{13}$*
3. *At$_1$ this$_2$ corporation$_3$'s$_4$ meeting$_5$ people$_6$ are$_7$ concerned$_8$ about$_9$ the$_{10}$ company$_{11}$'s$_{12}$ plans$_{13}$*
4. *In$_1$ July$_2$ we$_3$ will$_4$ interview$_5$ the$_6$ candidate$_7$ and$_8$ review$_9$ her$_{10}$ resumé$_{11}$ again$_{12}$*
5. *The$_1$ platypus$_2$ is$_3$ a$_4$ strange$_5$ animal,$_6$ with$_7$ its$_8$ eggs$_9$ and$_{10}$ its$_{11}$ webbed$_{12}$ feet$_{13}$*
6. *I$_1$ think$_2$ that$_3$ although$_4$ my$_5$ NACLO$_6$ exam$_7$ was$_8$ difficult,$_9$ it$_{10}$ was$_{11}$ a$_{12}$ lot$_{13}$ of$_{14}$ fun$_{15}$*
7. *Linguistics$_1$ is$_2$ a$_3$ beautiful$_4$ science$_5$ that$_6$ provides$_7$ interdisciplinary$_8$ insight$_9$ into$_{10}$ the$_{11}$ human$_{12}$ experience$_{13}$*

Note that the NacloWeb corpus treats the possessive element *'s* as a separate word. (So in Sentence #1, word #5 is *'s* and word #6 is *House*.)

In our experiment on the NacloWeb Corpus, we ran each of the corpus' sentences through BERT and recorded the outputs of four attention heads (8-11, 7-6, 9-6, and 5-4). Unfortunately, due to some extremely sloppy experimental procedure, we don't remember in which order we ran them through the model; in addition, we forgot to record some data. Your job is to fill in the blanks! Note that some blanks may have more than one connection, and some may have none at all.

Sentence	**8-11**	**7-6**	**9-6**	**5-4**
Sentence A	12 → 13	5 → 7	14 → 15	10 → 7, 5 → 1
Sentence B	**(a)**	8 → 9, 11 → 13	7 → 9, 7 → 13	8 → 2, 11 → 2
Sentence C	2 → 3, 10 → 11	4 → 5, 12 → 13	**(b)**	11 → 3
Sentence D	**(c)**	4 → 5	12 → 13, 8 → 11	None
Sentence E	3 → 5, 11 → 13	**(d)**	10 → 13	5 → 1
Sentence F	**(e)**	1 → 2, 5 → 6	7 → 8, 3 → 6	None
Sentence G	**(f)**	**(g)**	**(h)**	**(i)**

D1. Identify sentences A-G. Record your answers in the Answer Sheets.

D2. Fill in the missing data in the table in the Answer Sheets.