(N) A Token of Your Attention (1/3) [15 Points]

When computers process human language, the fundamental unit of language they use is usually the **token**, which can be anywhere from a single letter to a whole word. **In this problem, we will assume that tokens correspond to *morphemes*, or the smallest units of meaning in each word.** For example, we can represent the word *chasing* with a token for the verb *chase* and another for its suffix *ing*.

Modern approaches for processing language with computers use a mechanism called **attention**[1] to decide which words matter most when processing a sequence. NACLOLabs has trained a multilingual system that uses attention, and your task is to figure out how its attention works!

The cat eats the rat on the mat.

	the	*cat*	*eat*	*s*	*the*	*rat*	*on*	*the*	*mat*
the	.60	.05	.05	.05	.05	.05	.05	.05	.05
cat	.05	.35	.30	.05	.05	.05	.05	.05	.05
eat	.05	.15	.25	.20	.05	**.10**	.05	.05	.10
s	.05	.05	.40	.25	.05	.05	.05	.05	.05
the	.05	.05	.05	.05	.60	.05	.05	.05	.05
rat	.05	.05	**.20**	.05	.05	.45	.05	.05	.05
on	.05	.05	.05	.05	.05	.05	.60	.05	.05
the	.05	.05	.05	.05	.05	.05	.05	.60	.05
mat	.05	.05	.20	.05	.05	.05	.05	.05	.45

In this problem, attention is visualized as a matrix of decimals[2]. Each cell shows how much the word at the left of that row pays attention to the word at the top of that column. Attention has the following constraint placed on it: the values in each row must add up to 1. For example, in the sentence *The cat eats the rat on the mat*, shown in the matrix on the right, the word *eat* gives .10 (i.e. 10%) of its attention to *rat*, while *rat* gives .20 of its attention to *eat*. Because each row must sum to 1, attention has to be selective: it cannot give full weight to every word, so it must split its attention across the sequence.

Below is another attention matrix for the sentence *The dog was chasing the fast cat around the yard*. However, only the first five rows are visible!

The dog was chasing the fast cat around the yard.

	the	*dog*	*was*	*chase*	*ing*	*the*	*fast*	*cat*	*around*	*the*	*yard*
the	.60	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
dog	.04	.40	.04	.24	.04	.04	.04	.04	.04	.04	.04
was	.04	.04	.32	.32	.04	.04	.04	.04	.04	.04	.04
chase	.04	.12	.16	.20	.16	.04	.04	.08	.04	.04	.08
ing	.04	.04	.04	.32	.32	.04	.04	.04	.04	.04	.04

N1. Write the attention values (as decimals) for the word pairs **(a)**–**(d)** from the sentence *The dog was chasing the fast cat around the yard*. Here, A → B stands for the amount of attention A gives to B; for example, *chase* → *dog* = .12 .

(a) *fast* → *fast* ☐ **(b)** *fast* → *dog* ☐ **(c)** *cat* → *cat* ☐ **(d)** *yard* → *chase* ☐

[1]The mechanism in this problem is more specifically referred to as **encoder self-attention**, where a sequence is compared to itself rather than to another sequence.

[2]This problem uses a simplified but consistent procedure for calculating attention values. All decimal values are rounded down.

(N) A Token of Your Attention (2/3)

N2. Fill in the gaps **(e)**–**(i)** in the attention matrix on the right. You don't have to fill in the shaded cells.

(e) ☐ **(f)** ☐ **(g)** ☐

(h) ☐ **(i)** ☐

The cat meowed.

	the	*cat*	*meow*	*ed*
the				
cat		.40	**(e)**	
meow		**(f)**	.40	**(g)**
ed			**(h)**	**(i)**

Santali is an Austroasiatic language spoken by approximately 7.6 million people around Bengal, with a range that extends toward the foothills of the Himalayas. Santali is primarily written in Ol Chiki, a writing system developed by Raghunath Murmu in 1925.

NACLOLabs' attention mechanism works broadly similarly across Santali and English, placing attention on roughly analogous parts of the sequence. Below are six Santali sentences and their attention matrices, in arbitrary order.

However, two of the sentences have the exact same attention values, so they correspond to the same matrix! Fortunately, you are told that Santali sentence 4 has the same meaning as one of the prior three English sentences with an attention matrix.

1. ᱜᱩᱰᱩᱥᱮᱱ ᱢᱮᱱᱮᱠᱟᱵᱫ
2. ᱯᱟᱚᱩ ᱢᱟᱵᱫ ᱱᱟᱡᱵᱮᱯᱮᱟᱭ
3. ᱡᱟᱲᱟ ᱥᱮᱛᱟ ᱨᱟᱚᱮᱠᱟᱮ
4. ᱯᱩᱥᱤ ᱜᱩᱰᱩ ᱯᱟᱛᱤᱨᱮ ᱱᱟᱡᱮᱟᱮ
5. ᱥᱮᱛᱟ ᱯᱩᱥᱤᱥᱟᱶ ᱢᱮᱱᱫᱜᱟᱮ
6. ᱥᱮᱛᱟᱥᱟᱶ ᱨᱟᱚᱮᱠᱟᱭ

A.

						(j)	
	.352	.058	.058	.058	.352	.058	.058
	.058	.470	.058	.058	.235	.058	.058
	.058	.058	.470	.058	.235	.058	.058
	.058	.058	.058	.647	.058	.058	.058
	.176	.117	.117	.058	.235	.117	.176
(j)	.058	.058	.058	.058	.235	.470	.058
	.058	.058	.058	.058	.352	.058	.352

B.

						(k)
	.444	.055	.055	.333	.055	.055
	.055	.555	.055	.222	.055	.055
	.055	.055	.722	.055	.055	.055
	.166	.111	.055	.277	.222	.166
	.055	.055	.055	.444	.333	.055
(k)	.055	.055	.055	.333	.055	.444

C.

		(l)			
	.533	.066	.266	.066	.066
(l)	.066	.733	.066	.066	.066
	.133	.066	.333	.266	.200
	.066	.066	.533	.266	.066
	.066	.066	.400	.066	.400

D.

			(m)			
	.705	.058	.058	.058	.058	.058
	.058	.529	.235	.058	.058	.058
(m)	.058	.117	.294	.235	.117	.176
	.058	.058	.470	.294	.058	.058
	.058	.058	.235	.058	.529	.058
	.058	.058	.352	.058	.058	.411

E.

				(n)	
	.666	.083	.083	.083	.083
	.083	.417	.333	.083	.083
	.083	.166	.333	.250	.166
(n)	.083	.083	.500	.250	.083
	.083	.083	.333	.083	.417

N3. Match each Santali sentence 1–6 with its corresponding attention matrix A–E. Remember that two sentences correspond to the same attention matrix!

1. ☐ 2. ☐ 3. ☐ 4. ☐ 5. ☐ 6. ☐

N4. Identify the English sentence that has the same meaning as sentence 4.

○ *The cat eats the rat on the mat.* ○ *The dog was chasing the fast cat around the yard.* ○ *The cat meowed.*

(N) A Token of Your Attention (3/3)

N5. Fill in gaps **(j)** through **(n)** in the attention matrices with the correct Ol Chiki tokens in the corresponding Santali sentences. For the matrix that corresponds to two sentences, provide both tokens. You don't have to fill in the shaded cells. **Please write clearly!**

(j)

(k)

(l)

(m)

(n)

N6. Explain your observations about how NACLOLabs calculates attention weights between tokens in a sequence, regardless of which language: