

Processing Text with NNs

- Map each word to a vector of 0s and 1s (“1-hot repr.”):

$$\text{cat} \mapsto (0, 0, \dots, 0, 1, 0, \dots, 0)$$

- Sentence is then a matrix:

		the	cat	is	on	the	mat
	↑	a	0	0	0	0	0
		about	0	0	0	0	0
	
		cat	0	1	0	0	0
Vocabulary size:	
1.3M English		is	0	0	1	0	0
2.2M Czech	
		the	1	0	0	1	0
	
	↓	zebra	0	0	0	0	0