# Intelligent Minds and Machines

PSYC 3043

Week 5: Language

# today's agenda

+ language!

# an experiment

+https://dbvmtbylbs.cognition.run

# activity: arrange on the board

+jhool
+kha
+ladka
+tendua
+masal
+bhediya
+dekh
+seengh
+masakkali
+ber
+gaddi
+shatir
+soongh
+hoga
+komal
+tez
+qalabaz
+kela
+bhaag
+shikar
+pakad
+bandar
+hil
+nariyal
+gabbar
+shandaar
+gulabo

# activity: arrange on the board

+jhool
+kha
+ladka
+tendua
+masal
+bhediya
+dekh

+seengh
+masakkali
+ber
+gaddi
+shatir
+soongh
+hoga

+komal
+tez
+qalabaz
+kela
+bhaag
+shikar
+pakad

+bandar
+hil
+nariyal
+gabbar
+shandaar
+gulabo

# english to hindi translations!

- tarzan = gulabo
- jane = masakkali
- boy = ladka
- cheetah = tendua
- chimp = bandar
- rhino = seengh
- bigfoot = gabbar
- junglebeast = bhediya
- coconut = nariyal
- banana = kela
- berries = ber
- jeep = gaddi

- fierce = shatir
- yummy = shandaar
- soft = komal
- quick = tez
- acrobatic = qalabaz

- will = hoga
- flee = bhaag
- hunt = shikaar
- chase = pakad
- squish = masal
- move = hil
- eat = kha
- see = dekh
- smell = soongh
- swing = jhool

# language model demo

+code notebook

# concept check: Lake and Murphy (2021)

+ word representations / embeddings
+ dimensionality
+ semantic similarity / cosine similarity
+ testing models

Brendan M. Lake

Gregory L. Murphy

# dimensions and similarity

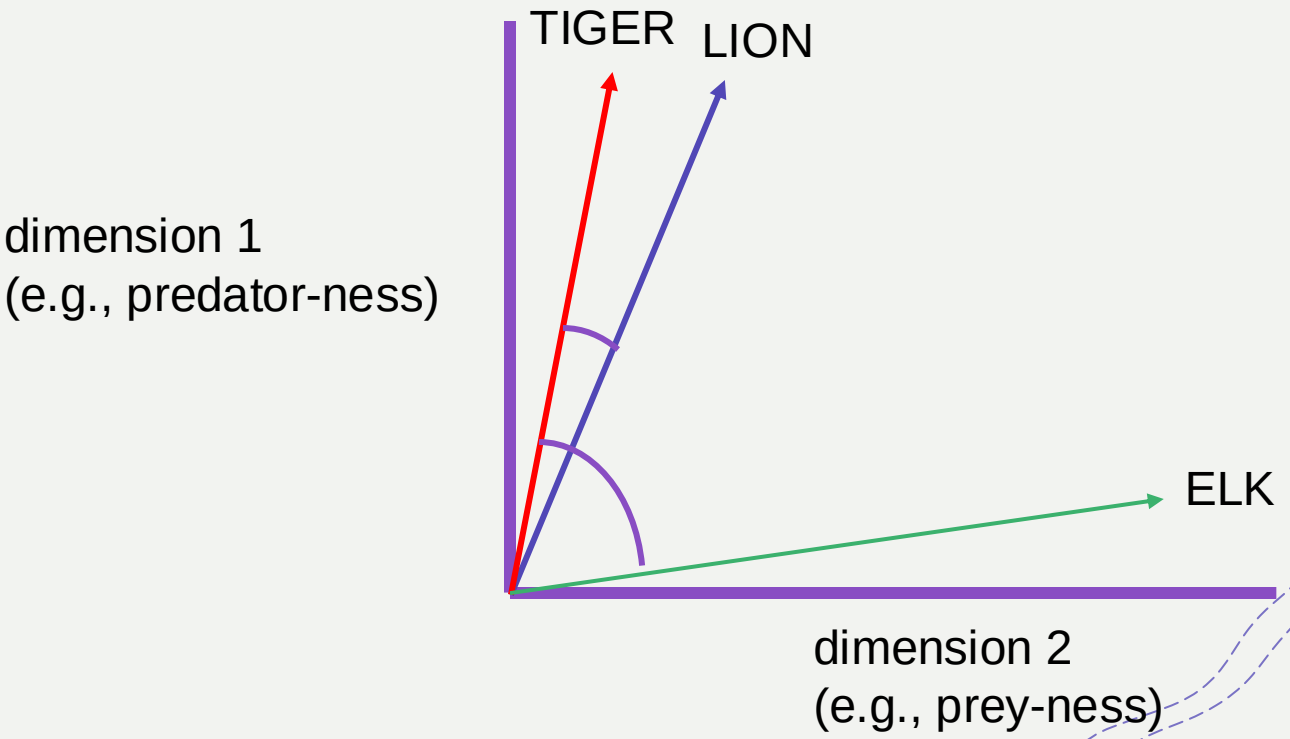
**lions** are **tigers** are carnivorous predators

a group of **lions** is called a pride

**tigers** have stripes

**lions** and **tigers** hunt deer and **elk**

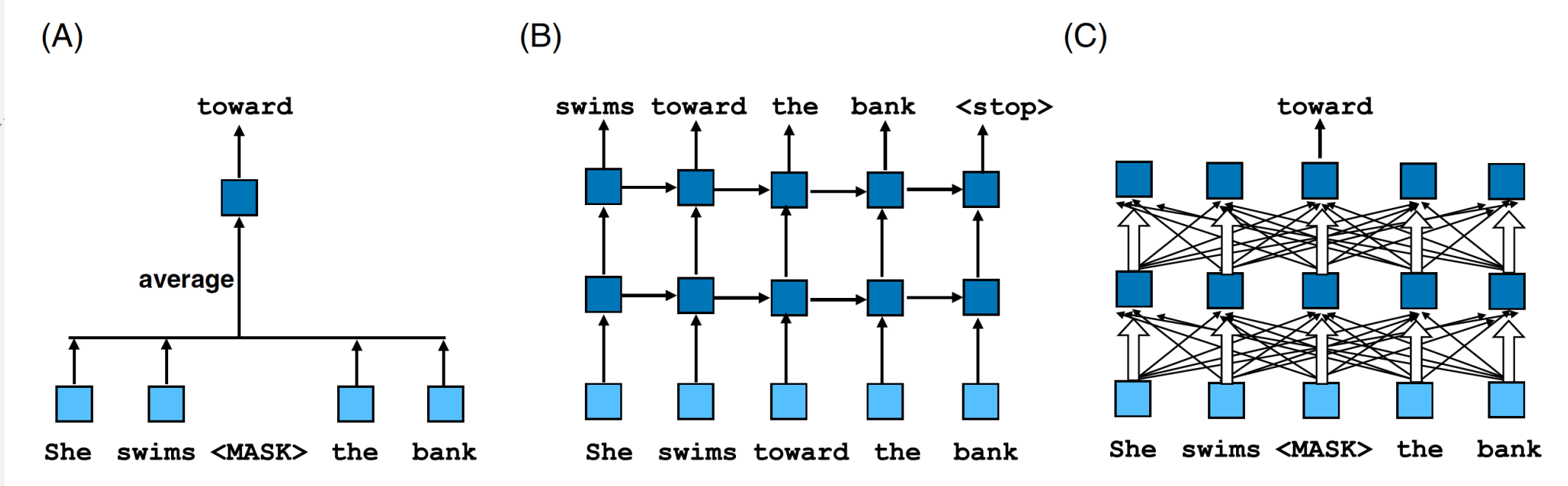
**elk** and deer are herbivores



# concept check: Lake and Murphy (2021)

- four types of NLP models
  - Latent Semantic Analysis (LSA): dimensionality reduction model : unsupervised
  - word2vec: predictive, semi-supervised (autoregressive)
  - RNN (Elman): *recurrent* neural network (semi-supervised)
  - BERT/GPT-2: predictive, autoregressive, attention/Transformer based

# concept check: models




word2vec
no concept of word order, averages all words' representations within a window

RNN
predicts each upcoming word using an indirect connection to previous words

BERT/GPT-2
all words contribute to different degrees in generating a target word's representation

# desiderata

**Table 1**

*Five Desiderata. Word Representations Should Support These Basic Functions of Language Use*

| Behavior to be explained | Examples |
|---|---|
| 1. Describing a perceptually present scenario, or understanding such a description. | That knife is in the wrong place.<br>The orangutan is using a makeshift umbrella. |
| 2. Choosing words on the basis of internal desires, goals, or plans. | I am looking for a knife to cut the butter.<br>I need a flight from New York to Miami. |
| 3. Responding to instructions and requests appropriately. | Pick up the knife carefully.<br>Find an object that is not the small ball. |
| 4. Producing and understanding novel conceptual combinations. | That's a real apartment dog.<br>The apple train left the orchard. |
| 5. Changing one's beliefs about the world based on linguistic input. | Sharks are fish but dolphins are mammals.<br>Umbrellas fail in winds over 18 knots. |

# annotations in themes