Readme RSM of stimuli

1. RSA

Let’s grab 5 random sentences from the 200 most prototypical, and only between 8 and 14 words (See prototypical\_sentences\_8\_14\_randomN.csv)

I have been always doing the RSMs using **ward** as a linkage method, and **euclidean** as a distance metric. Regard linkage method, **ward** does better when clustering categories, but **average** does better when clustering sentences:

Distance metric:

because of the activation function ‘elu’ so values are between like -1 and 19 (you can see sample of feature vectors here: prototypical\_random10\_layerX.csv where you have 10 sentences per category, each sentence represented by last\_layer (dense\_final) of cnn20\_multitask, size 128 per sentence.

It seems cosine is better than euclidean at clustering sentences.

But this method of clustering isn’t the best. It’s better to do our method (correlate sentence from two categories and insert in RSM).

Dendrogram

Distance metric = ‘correlation’

Linkage = ‘ward’