A Supplementary Material

A.1 Precision@10 for the 80 test words

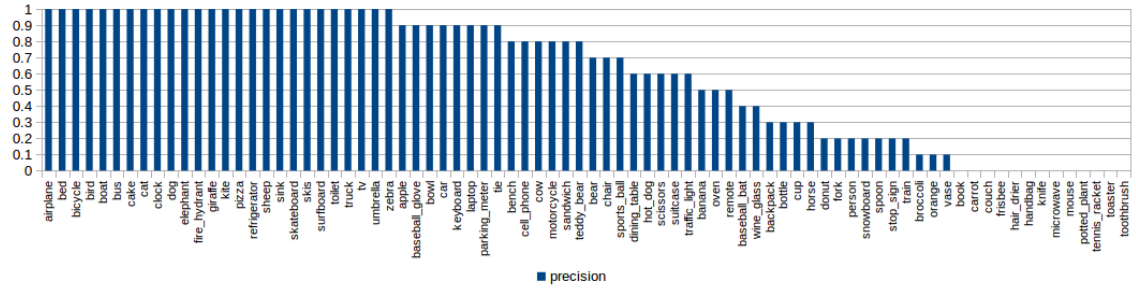
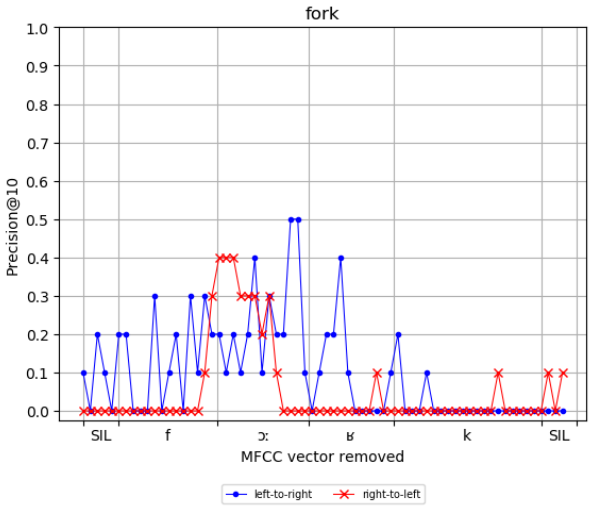
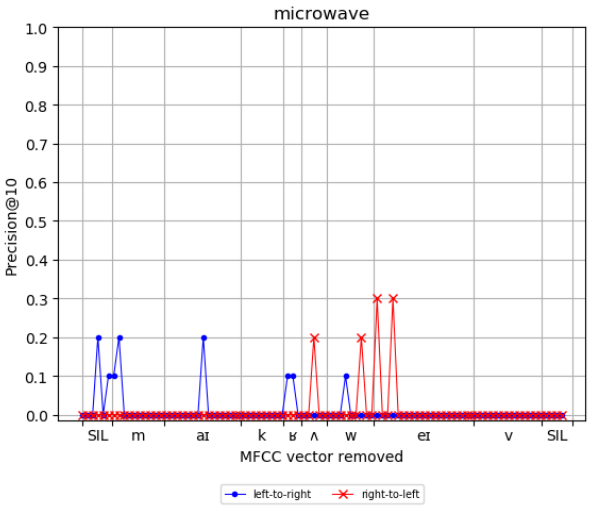


Figure: Precision@10 for each of the 80 test words

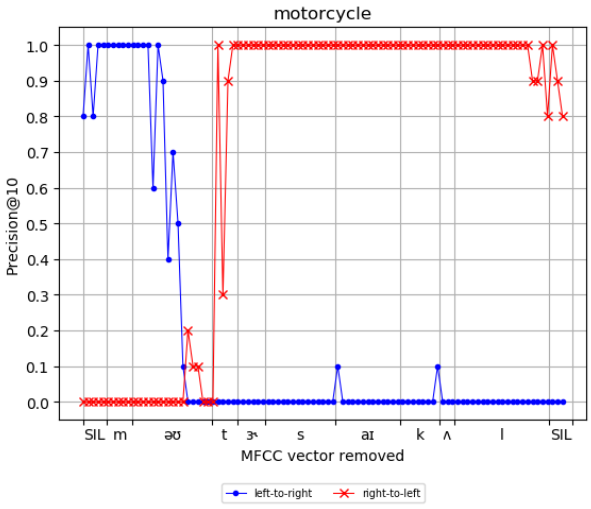
A.2 Gating Samples



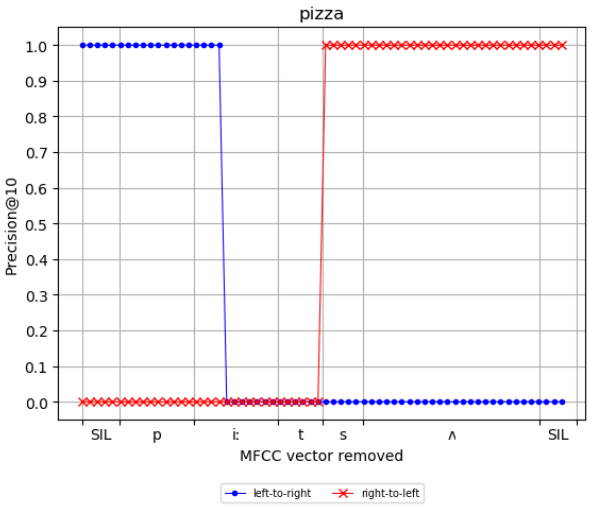
(a)



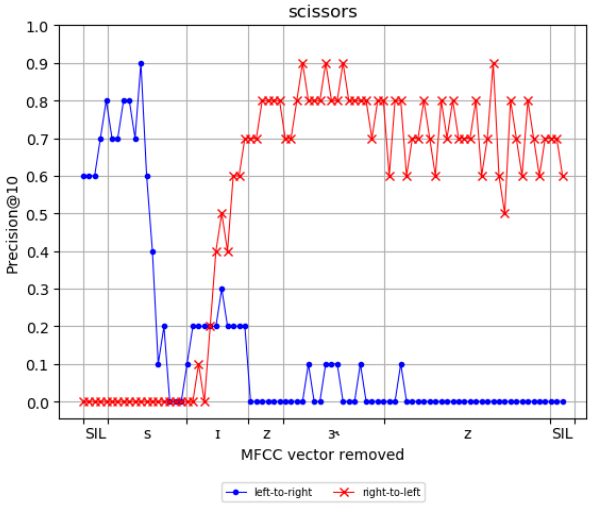
(b)



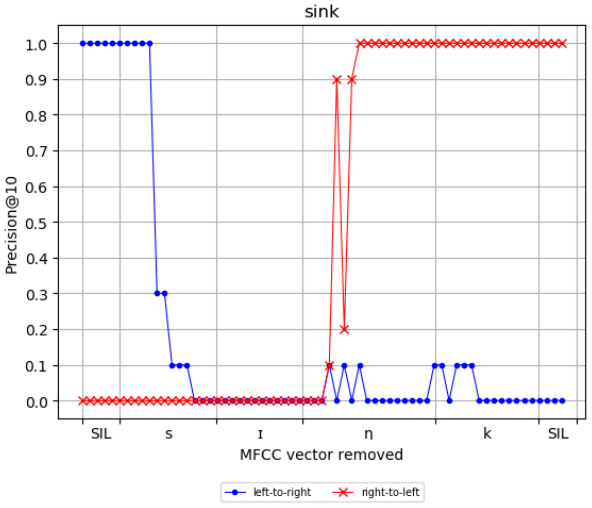
(c)



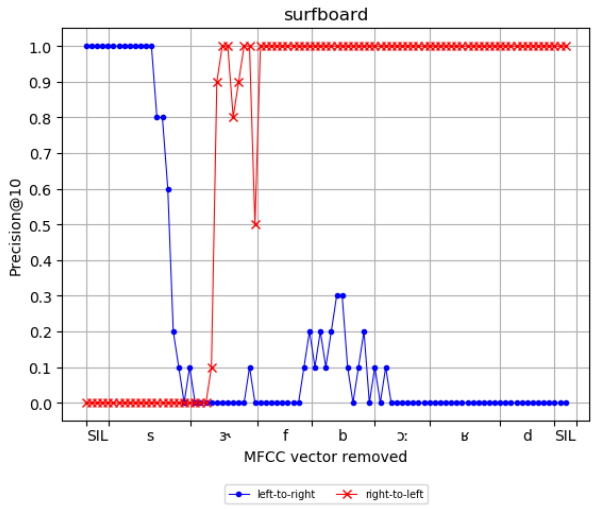
(d)



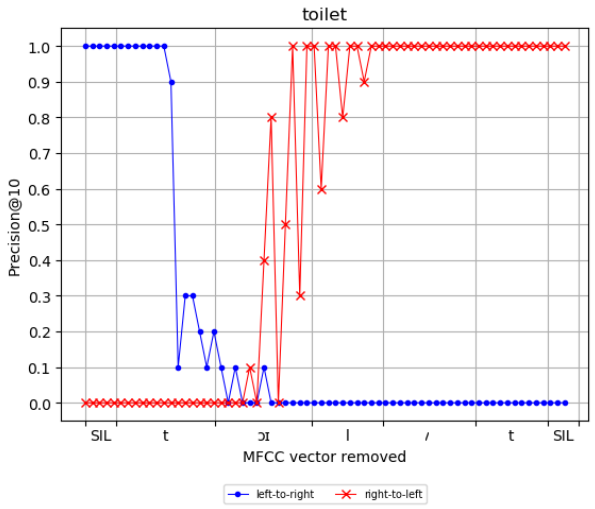
(e)



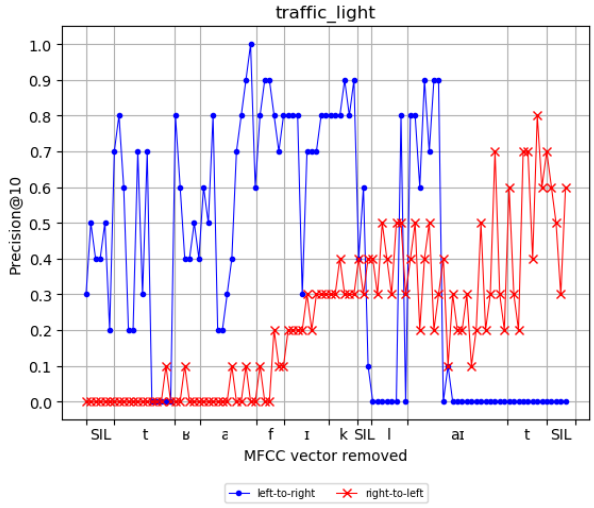
(f)



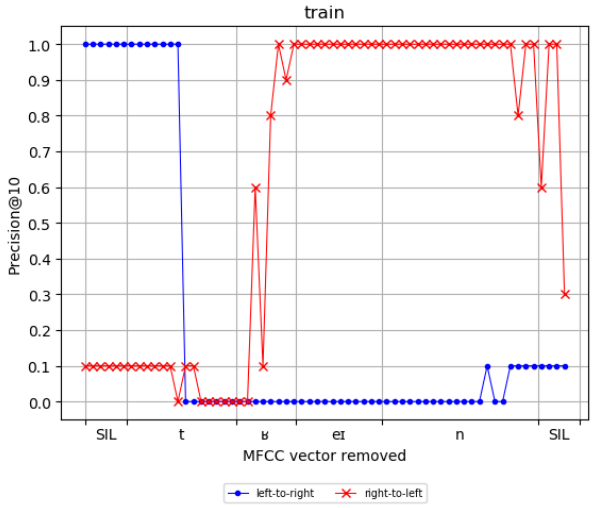
(g)



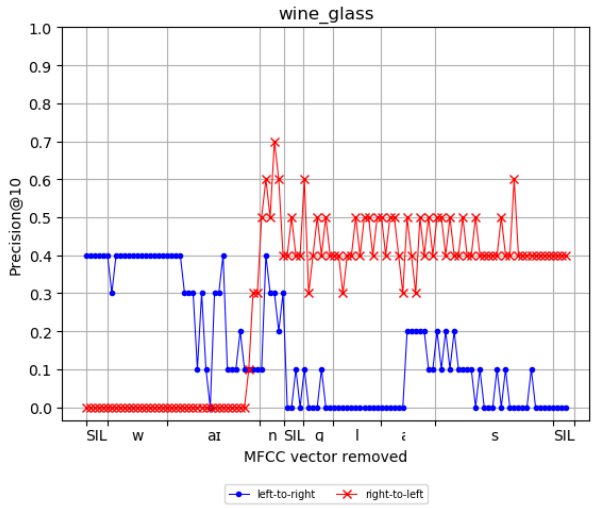
(h)



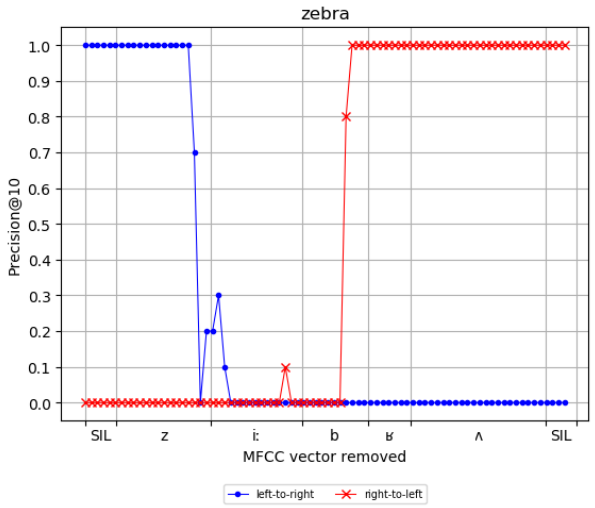
(i)



(j)



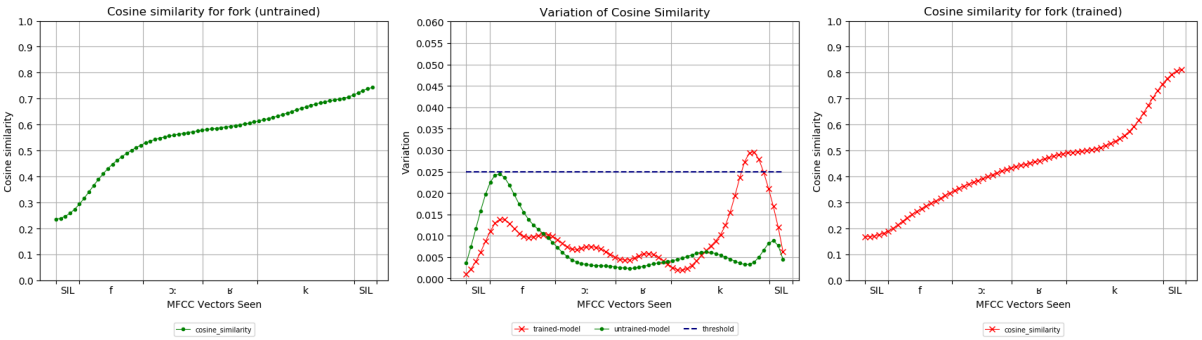
(k)



(l)

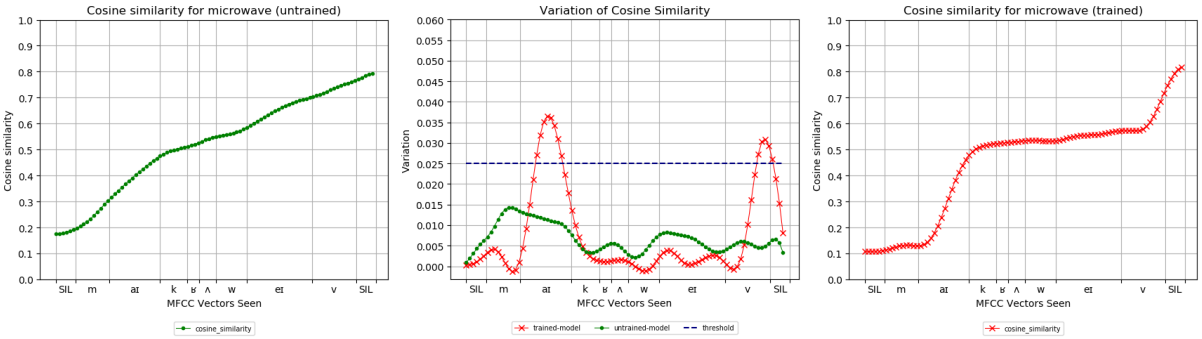
Figure: Evolution of Precision@10 for each ablation step for the words (a) “fork”, (b) “microwave”, (c) “motorcycle”, (d) “pizza”, (e) “scissors”, (f) “sink”, (g) “surfboard, (h) “toilet”, (i) “traffic light”, (j) “train”, (k) “wine glass”, (l) “zebra”. It should be noted that “fork” and “microwave” do not display the same behaviour as the other words, this could be explained by the fact these two words both have a very low Precision@10.

A.3 Activation Sample



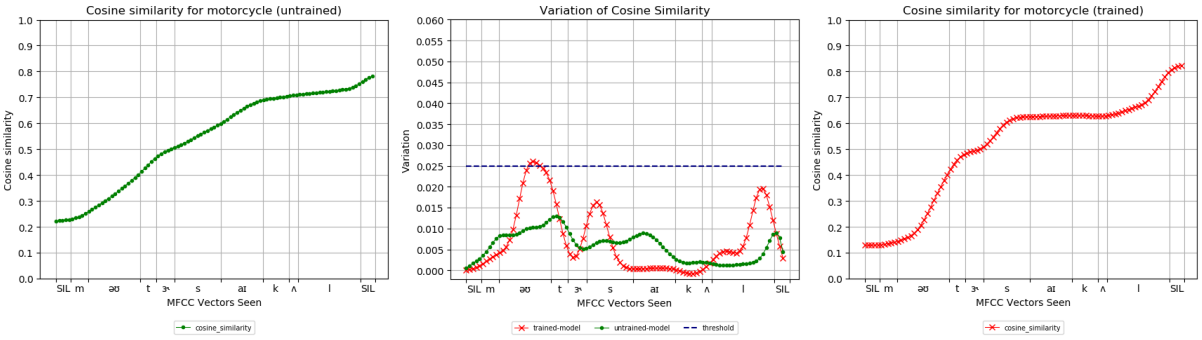
(a) (b) (c)

Figure 1: Evolution of cosine similarity for the word “fork”. Figure 1b shows peaks indicating the inflection points of curve 1a (untrained model, green) and 1c (trained model, red)



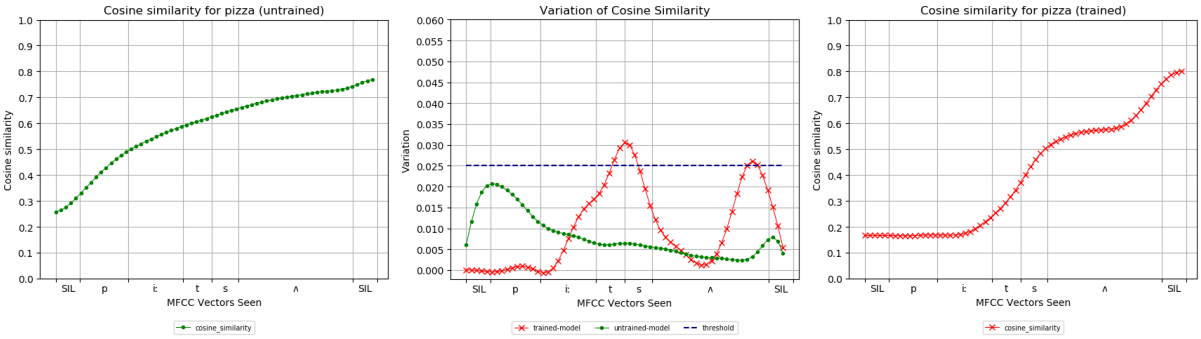
(a) (b) (c)

Figure 2: Evolution of cosine similarity for the word “microwave”. Figure 2b shows peaks indicating the inflection points of curve 2a (untrained model, green) and 2c (trained model, red)



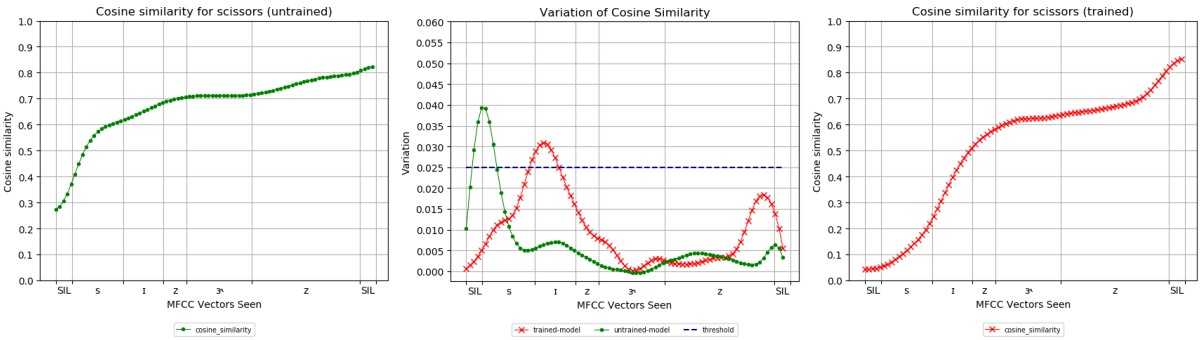
(a) (b) (c)

Figure 3: Evolution of cosine similarity for the word “motorcycle”. Figure 3b shows peaks indicating the inflection points of curve 3a (untrained model, green) and 3c (trained model, red)



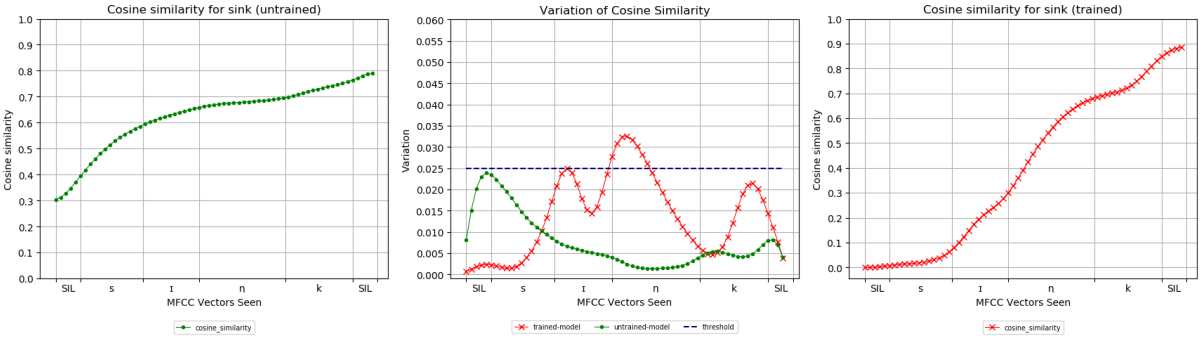
(a) (b) (c)

Figure 4: Evolution of cosine similarity for the word “pizza”. Figure 4b shows peaks indicating the inflection points of curve 4a (untrained model, green) and 4c (trained model, red)



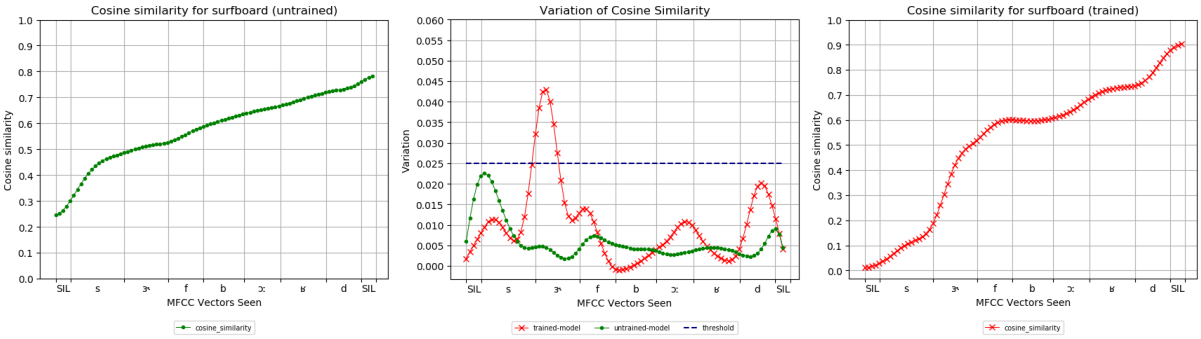
(a) (b) (c)

Figure 5: Evolution of cosine similarity for the word “scissors”. Figure 5b shows peaks indicating the inflection points of curve 5a (untrained model, green) and 5c (trained model, red)



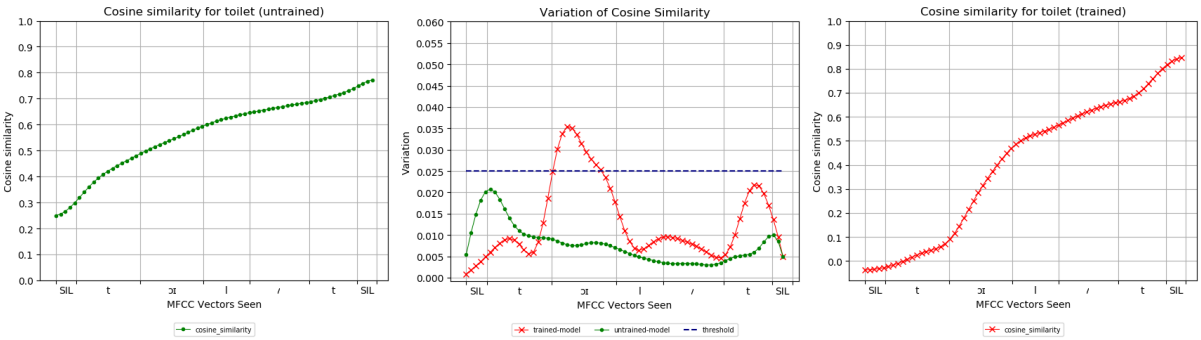
(a) (b) (c)

Figure 6: Evolution of cosine similarity for the word “sink”. Figure 6b shows peaks indicating the inflection points of curve 6a (untrained model, green) and 6c (trained model, red)



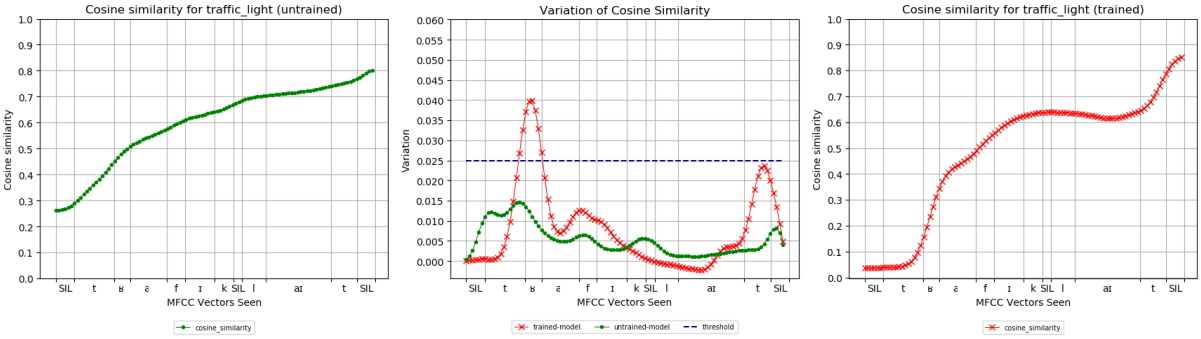
(a) (b) (c)

Figure 7: Evolution of cosine similarity for the word “surfboard”. Figure 7b shows peaks indicating the inflection points of curve 7a (untrained model, green) and 7c (trained model, red)



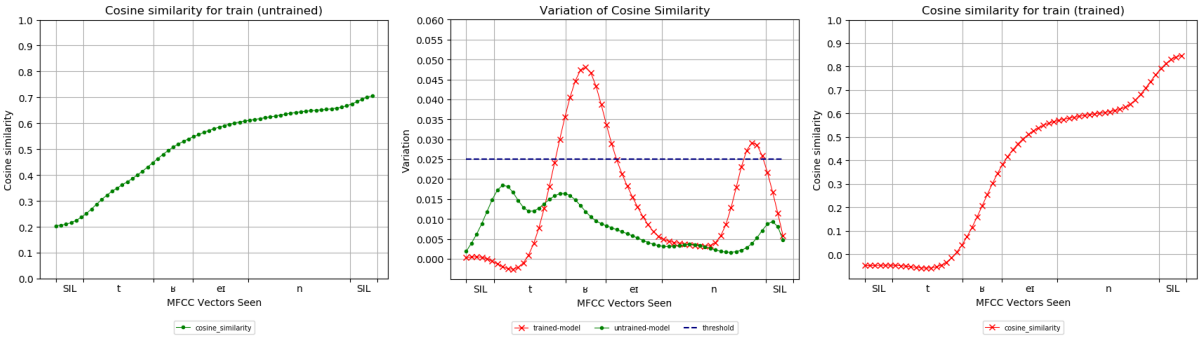
(a) (b) (c)

Figure 8: Evolution of cosine similarity for the word “toilet”. Figure 8b shows peaks indicating the inflection points of curve 8a (untrained model, green) and 8c (trained model, red)



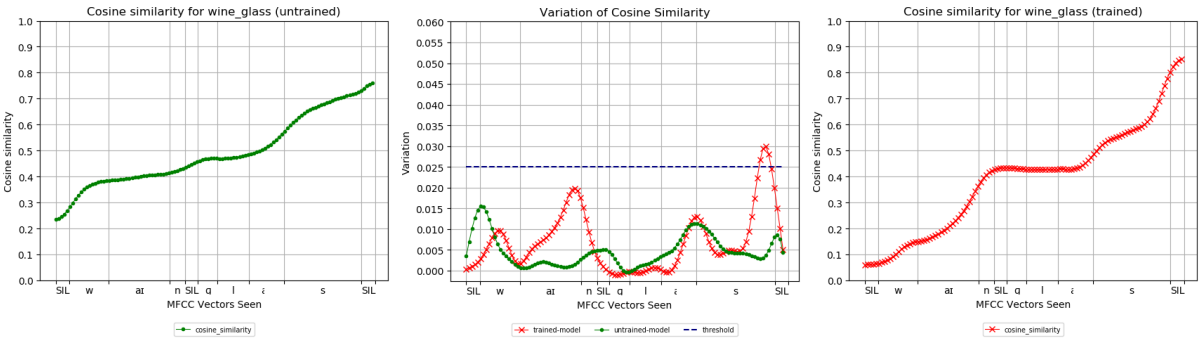
(a) (b) (c)

Figure 9: Evolution of cosine similarity for the word “traffic light”. Figure 9b shows peaks indicating the inflection points of curve 9a (untrained model, green) and 9c (trained model, red)



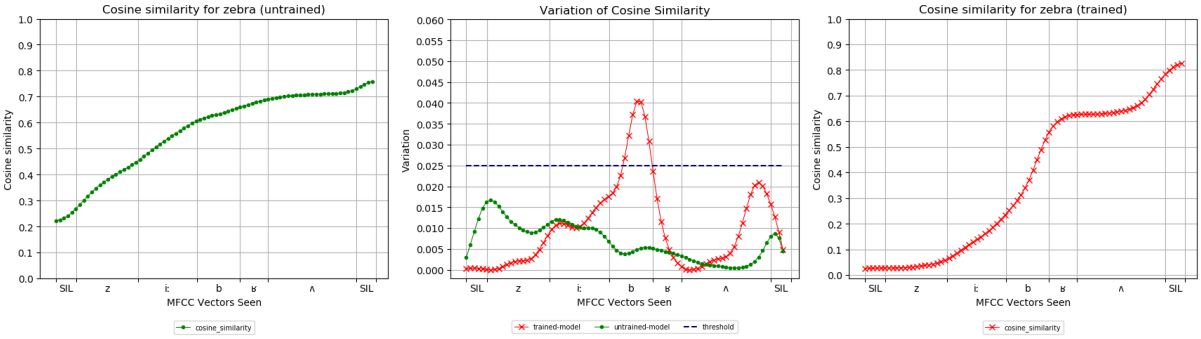
(a) (b) (c)

Figure 10: Evolution of cosine similarity for the word “train”. Figure 10b shows peaks indicating the inflection points of curve 10a (untrained model, green) and 10c (trained model, red)



(a) (b) (c)

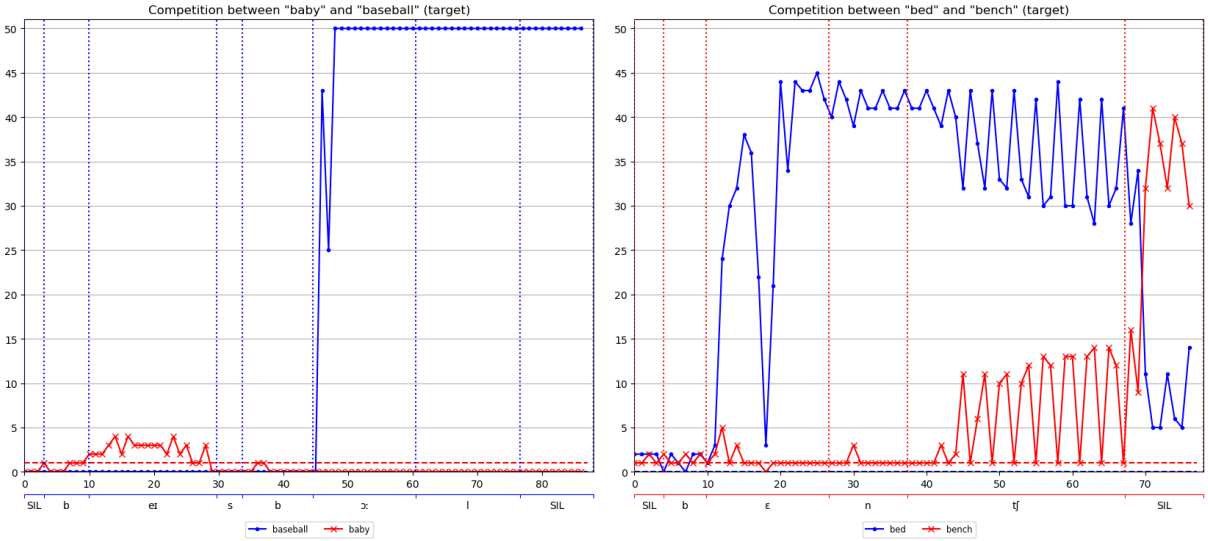
Figure 11: Evolution of cosine similarity for the word “wine glass”. Figure 11b shows peaks indicating the inflection points of curve 11a (untrained model, green) and 11c (trained model, red)



(a) (b) (c)

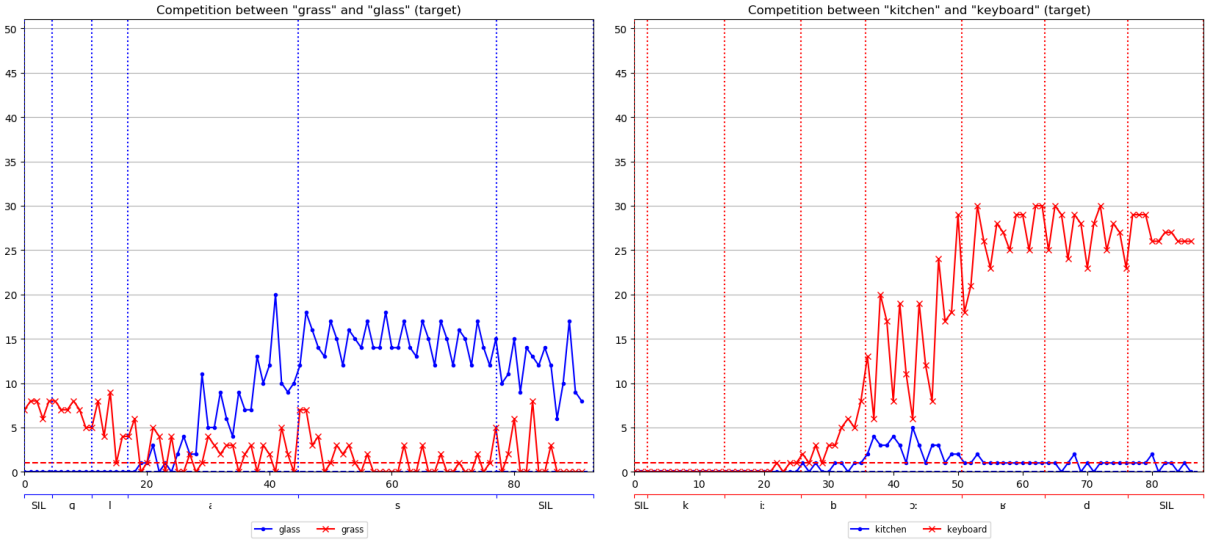
Figure 12: Evolution of cosine similarity for the word “zebra”. Figure 12b shows peaks indicating the inflection points of curve 12a (untrained model, green) and 12c (trained model, red)

A.4 Competition Sample



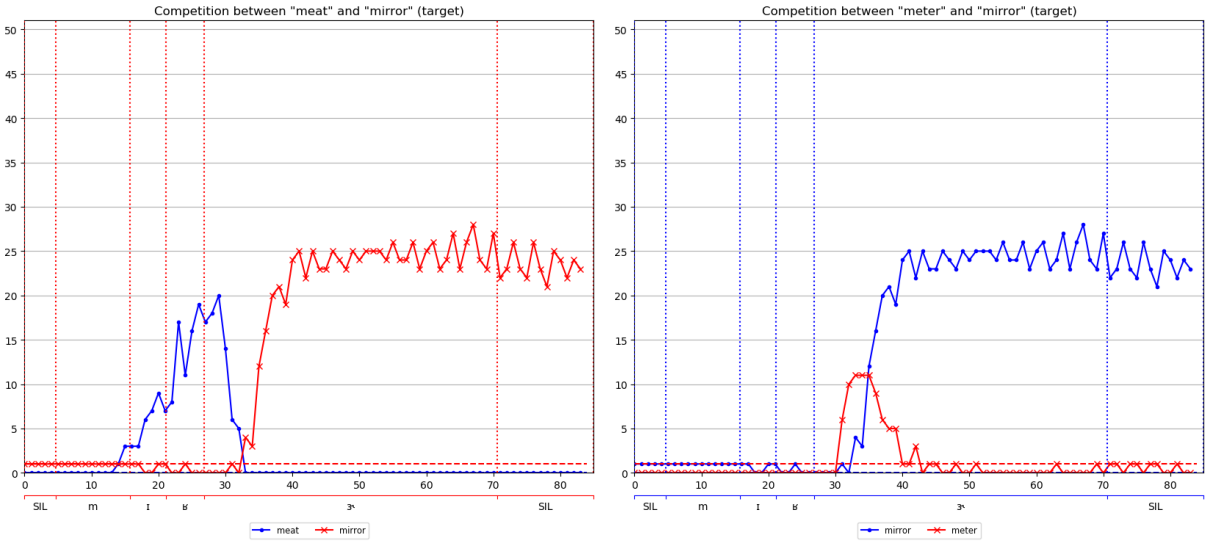
(a) (b)

Figure 13: Competition plots between 13a “baseball” and “baby” and 13b “bed” and “bench”.



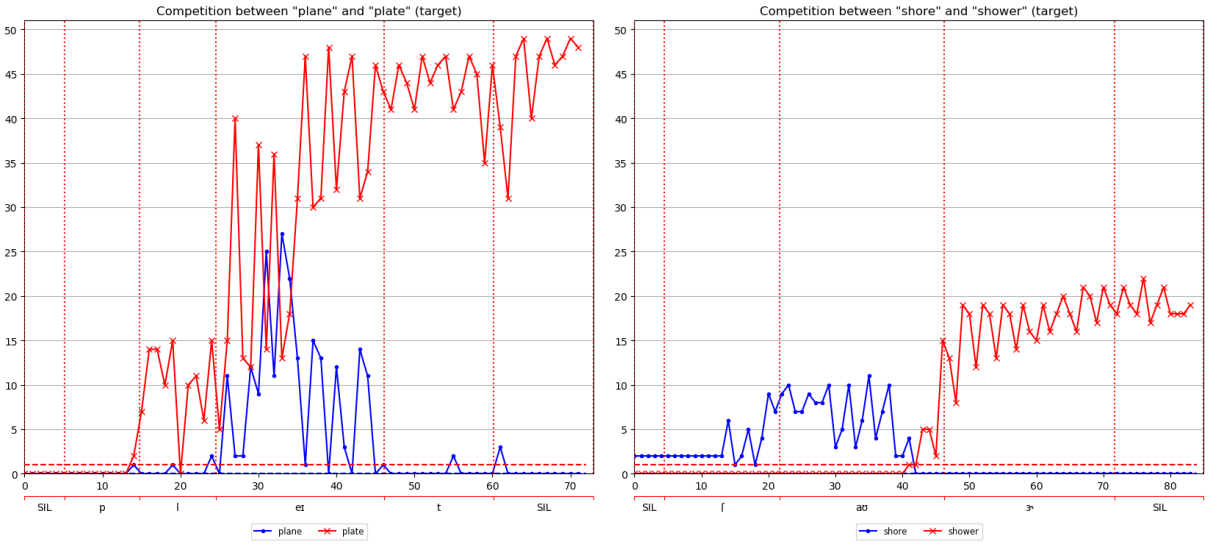
(a) (b)

Figure 14: Competition plots between 14a “glass” and “grass” and 14b “kitchen” and “keyboard”.



(a) (b)

Figure 15: Competition plots between 15a “meat” and “mirror” and 15b “mirror” and “meter”.



(a) (b)

Figure 16: Competition plots between 16a “plane” and “plate” and 16b “shore” and “shower”.