TP semaine 8: Analogie avec des embeddings distributionels

L'objectif de ce TP est d'observer les possibilités de raisonnement analogique dans un espace obtenu par plongements lexicaux statiques.

Dans l'espace obtenu par plongement lexical du TP de la semaine 6, sélectionner 3 candidats pour tester le raisonnement analogique, récupérer leurs embeddings dans la matrice termeterme (on prendra une valeur moyenne pour k, par exemple k=4). Faire les additions et différences vectorielles pour obtenir un vecteur dont on va rechercher les 10 voisins les plus proches. Choisir des relations sémantiques (par exemple femme-fille+homme), des relations morphologiques (par exemple cheval-le+les), des relations factuelles (par exemple femme-fille+les).

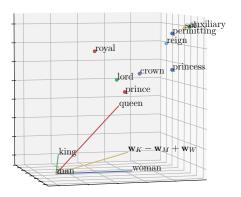


Figure 1. The relative locations of word embeddings for the analogy "man is to king as woman is to ..?". The closest embedding to the linear combination $\mathbf{w}_K - \mathbf{w}_M + \mathbf{w}_W$ is that of queen. We explain why this occurs and interpret the difference between them.

FIGURE M.1 – Graphique sur la fameuse analogie homme/femme|roi/x, version de Allen et Hospedales (2019)

woman – queen		queen – woman		king - man		man - king	
+ man		+ king		+ queen		+ woman	
$_{ m king}$	0.749	man	0.733	woman	0.745	son	0.728
prince	0.708	who	0.712	beautiful	0.726	queen	0.716
kingdom	0.694	men	0.711	girl	0.672	elizabeth	0.710
victoria	0.644	whom	0.708	my	0.655	brother	0.706
scotland	0.643	killed	0.685	lady	0.646	emperor	0.699
wales	0.640	person	0.676	she	0.628	wife	0.693
lord	0.638	young	0.673	thing	0.613	henry	0.691
great	0.627	himself	0.639	good	0.606	younger	0.686
elizabeth	0.622	said	0.637	her	0.595	daughter	0.681
throne	0.614	father	0.636	naked	0.592	prince	0.665

TABLE M.1 – Quelques calculs analogiques réalisés avec la démo de GloVe (https://github.com/stanfordnlp/GloVe) avec les 10 mots les plus proches par distance cosinus du point obtenu par l'opération additive sur les vecteurs.