

Hypothèse distributionnelle

- If A and B have almost identical environments we say that they are synonyms. *(Harris, 1954)*
- You shall know a word by the company it keeps. *(Firth, 1957)*
- The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear. *(Lenci, 2008)*

Matrice documents-termes

	QuatreVT 119 Kw	Voyage Bal 82 kw	Bête Hum. 128 kw	Bovary 117 kw
bataille	35	4	6	2
clair	105	26	96	52
facile	12	19	6	10
politique	11	0	9	5
voyage	17	196	94	44
idiot	2	1	2	6
amour	19	0	47	94

Table – Matrice documents-termes pour quelques mots et 4 romans (Quatrevingt-treize (Hugo); Le voyage en ballon (Verne); La bête humaine (Zola); Mme Bovary (Flaubert)).

Des romans dans l'espace

QuatreVT	(17,19)
Voyage Bal	(196,0)
Bête Hum.	(94,47)
Mme Bovary	(44,94)

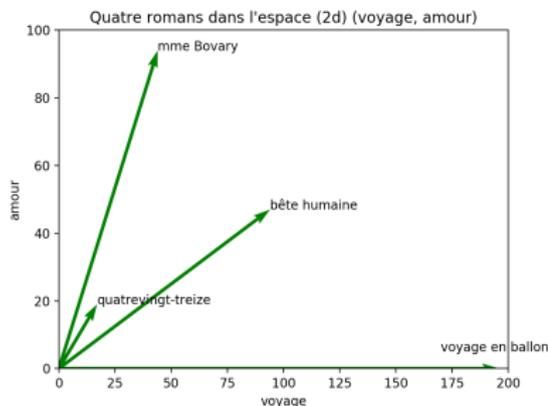


Figure – Représentation graphique de quelques romans dans le plan (voyage, amour)

Plongement dans un espace

- ▶ Système dense (il y a toujours un point entre deux points)
- ▶ Notion de déplacement (vecteur)
- ▶ Notion de distance (euclidienne)
- ▶ Notion de proximité (similarité cosinus)

Recherche d'information

Moteur de recherche (type google) :

- ▶ Tous les documents indexés sont des points dans un espace
- ▶ La requête (après traitement) peut être plongée dans le même espace
- ▶ Les documents pertinents sont les documents les plus « proches » de la requête

Inversion de la matrice

	QuatreVT 119 Kw	Voyage Bal 82 kw	Bête Hum. 128 kw	Bovary 117 kw
bataille	35	4	6	2
clair	105	26	96	52
facile	12	19	6	10
politique	11	0	9	5
voyage	17	196	94	44
idiot	2	1	2	6
amour	19	0	47	94

amour	(19,94)
bataille	(35,2)
facile	(12,10)
politique	(11,5)
voyage	(17,44)

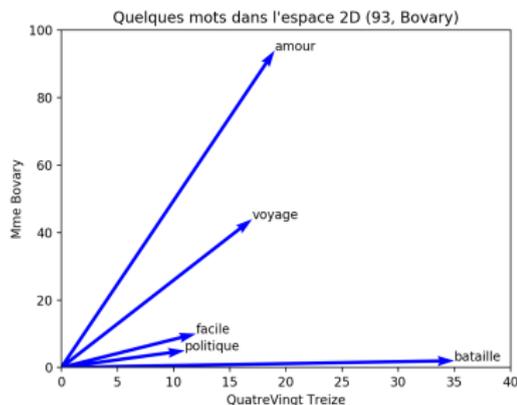


Figure – Représentation graphique de quelques mots dans le plan (93, Bovary)

Voisinages comme dimensions

je connaissais le	voyagiste	qui les amène
voyages : Différent du	voyagiste	ou tour-opérateur . En
compagnie à un	voyagiste	. Animaux domestiques : Lorsqu'
que disent certains	voyagistes	, légalement , aucune n'
Transport / Hébergement	voyagiste	FRAM propose les
Le plus vite votre	voyagiste	. Le Comité scientifique
le monde ... 6ème	voyagiste	européen , Kuoni est
autres , les grands	voyagistes	essayent de prendre
fév 2008 Ces	voyagistes	, qui opèrent principalement
parmis les meilleurs	voyagistes	en ligne . Service
clients chez un	voyagiste	- Développement d' un
les 15 principaux	voyagistes	en ligne en
ses amies des	voyagistes	locaux qui fêtent
foyers ...) et des	voyagistes	. La commercialisation Le
voyage à cheval .	Voyagiste	spécialiste du voyage
hôtel , à votre	voyagiste	, à votre ambassade

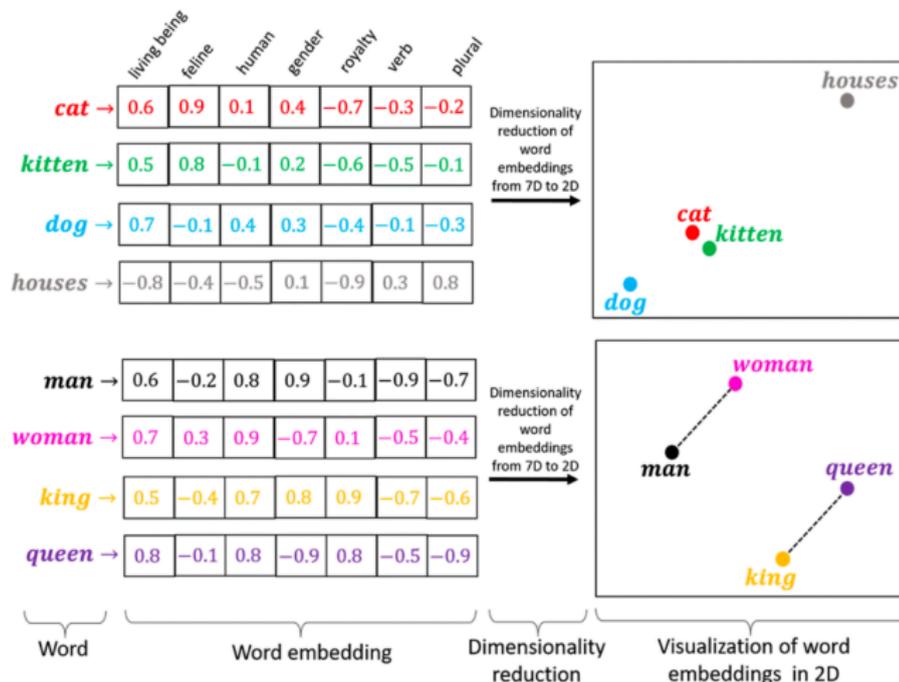
Extrait kwic de la concordance [-3,3] autour de voyagiste dans frWaC, consultation 2021-08-11

Matrice termes-termes

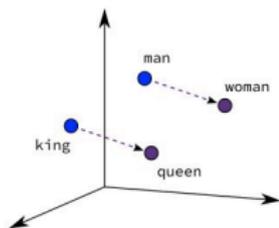
	bataille	voyage	homme	femme
arriver	246	470	1 819	890
tomber	100	83	1 205	384
habiller	2	4	339	384
mourir	180	116	339	1 088
	55 331	208 520	668 289	346 093

Table – Matrice Terme-Terme obtenue dans frWaC. Contextes en ligne.

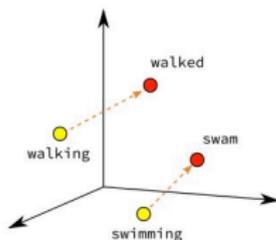
Similarité



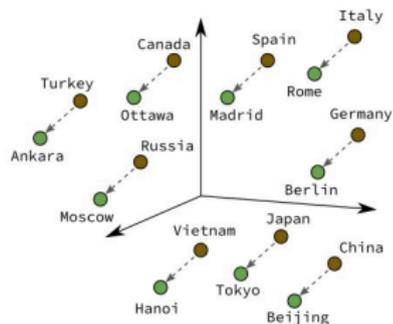
Analogie



Male-Female



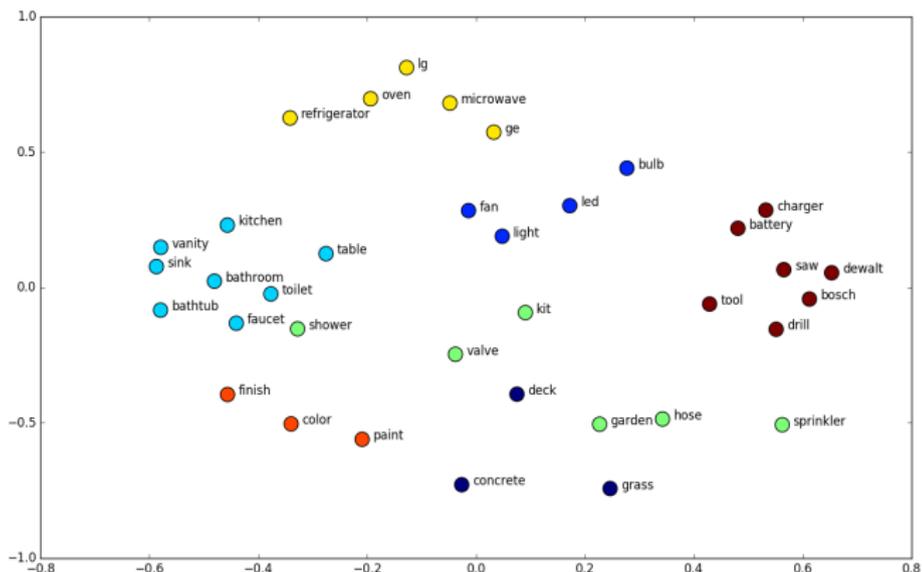
Verb Tense



Country-Capital

(Huang, 2025)

Visualisation



Synthèse

En représentant les mots comme des $\left\{ \begin{array}{l} \text{points} \\ \text{vecteurs} \\ \text{suites de nombres} \end{array} \right.$

on renouvelle la façon de faire de la sémantique lexicale,

et on permet de représenter les mots et les phrases de la meilleure façon possible pour l'entrée des systèmes neuronaux.