

Traitement Automatique des Langues Naturelles

Cours 4: Sémantique

Chloé Braud, Philippe Muller

Master IAFA 2024-2025

Les niveaux d'analyse

- phonologie: les sons
- morphologie: les mots et leur forme
- syntaxe: l'organisation des mots en phrase
- **sémantique: le sens dans la phrase**
- pragmatique: le sens en contexte

Sémantique

- la **sémantique** est l'étude du **sens** d'expressions de langage
- deux sous-champs :
 - sémantique lexicale: le sens des mots individuellement
 - sémantique compositionnelle: le sens d'unités plus larges à partir des unités de base
- la sémantique est reliée à la pragmatique
 - la sémantique traite la signification du langage *en isolation*
 - la pragmatique traite la signification du langage *en usage*

Sémantique compositionnelle

- décomposition du sens sur les parties d'une phrase
- composition des éléments pour faire des représentations au niveau des groupes de mots
- nécessite de définir la représentation cible
- pour quoi ? inférences/comparaison

Il fait beau et chaud → *Il fait beau*

Tous les hommes sont mortels. → *Socrate est mortel.*

Sémantique compositionnelle

à partir des règles de composition syntaxique par exemple

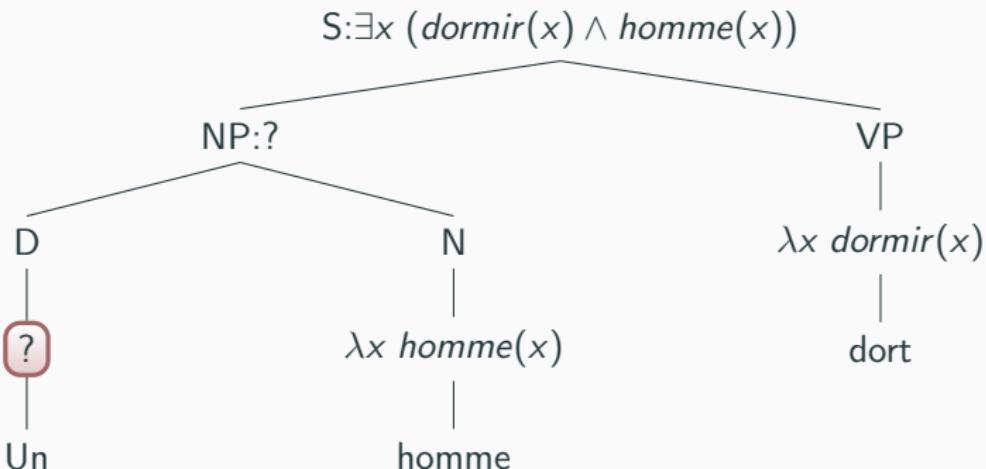
$$\begin{aligned} S(\text{sem}:f(\text{semNP}, \text{semVP})) &\rightarrow \\ \text{NP}(\text{sem}: \text{semNP}) \\ \text{VP}(\text{sem}: \text{semVP}) \end{aligned}$$

- représentation exprimable sous formes diverses
 - forme logique
 - forme vectorielle
- les formalismes varient sur la nature de la fonction de composition du sens :
 - composition logique par la syntaxe
 - composition de vecteurs
(par exemple dans un réseau de neurones récurrents → plus tard)

Approche logique

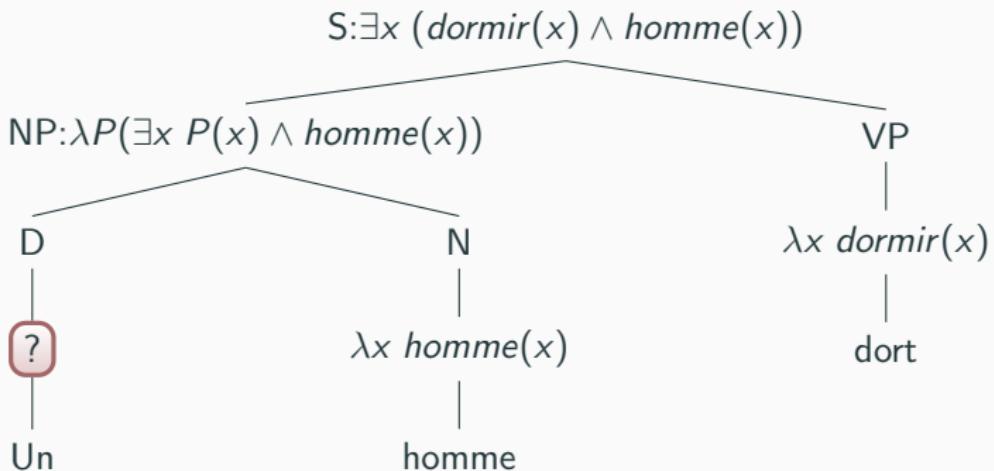
- cible des représentations : formules logiques (premier ordre)
- éléments de base (lexique) représentés par des fragments/des expressions de lambda calcul
- s'attache surtout aux combinaisons avec quantificateurs (existential, universel, ...)
- suit la structure donnée par l'analyse syntaxique
- modèle essentiellement théorique

Exemple: déterminant indéfini



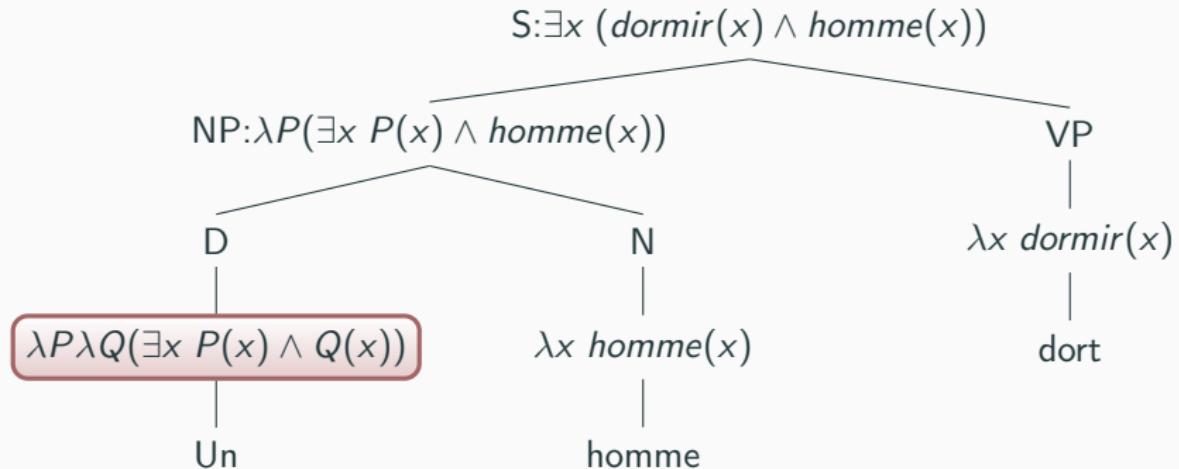
Un homme dort.

Exemple: déterminant indéfini



Un homme dort.

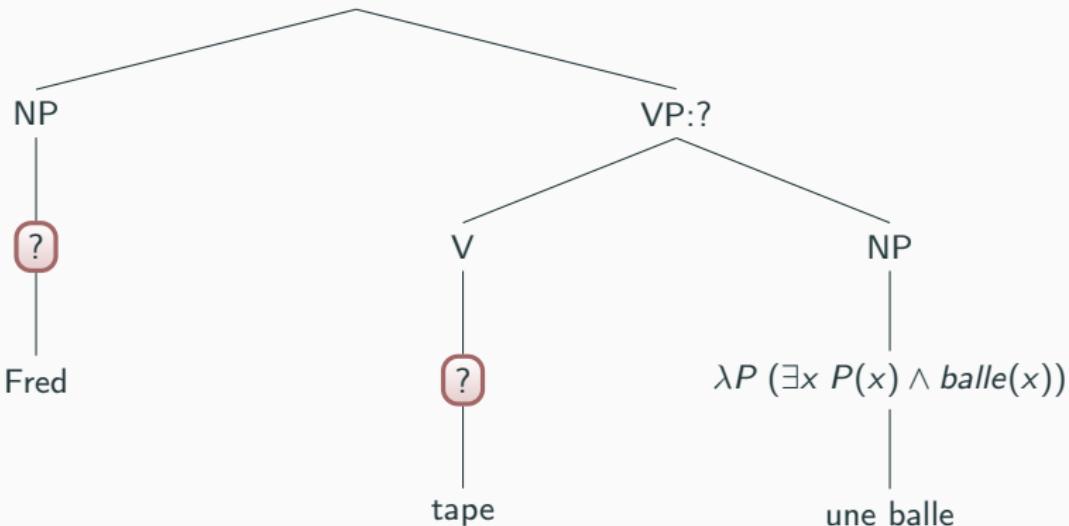
Exemple: déterminant indéfini



Un homme dort.

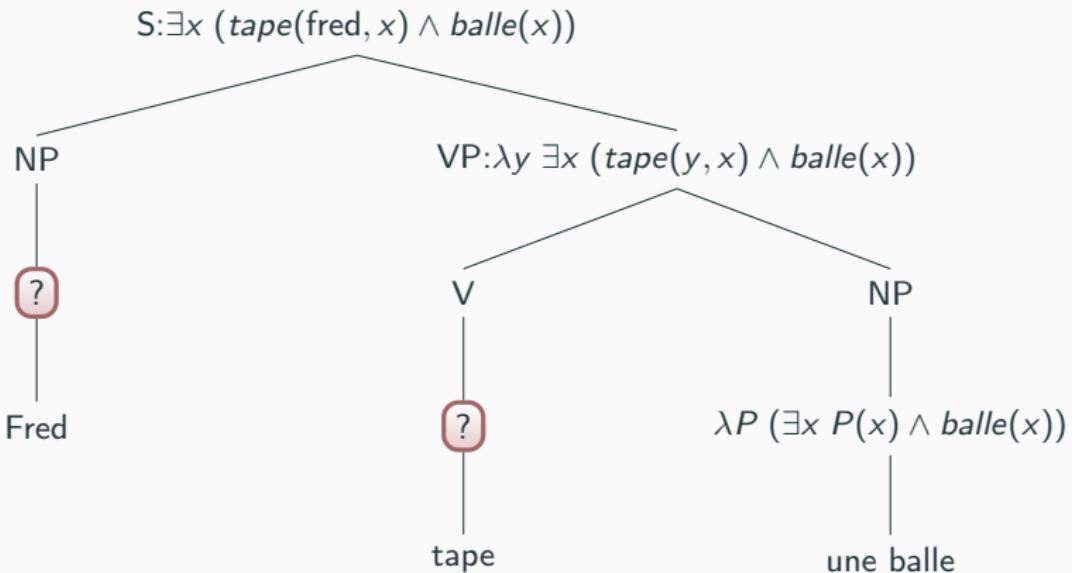
Exemple: nom propre, verbe transitif

S: $\exists x \ (tape(fred, x) \wedge balle(x))$



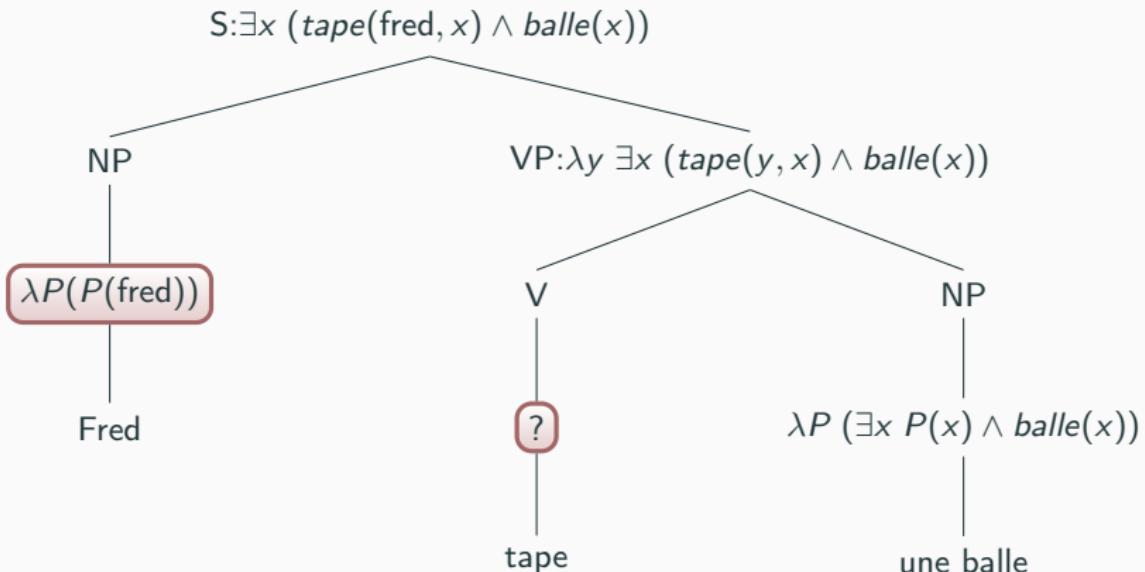
Fred tape une balle.

Exemple: nom propre, verbe transitif



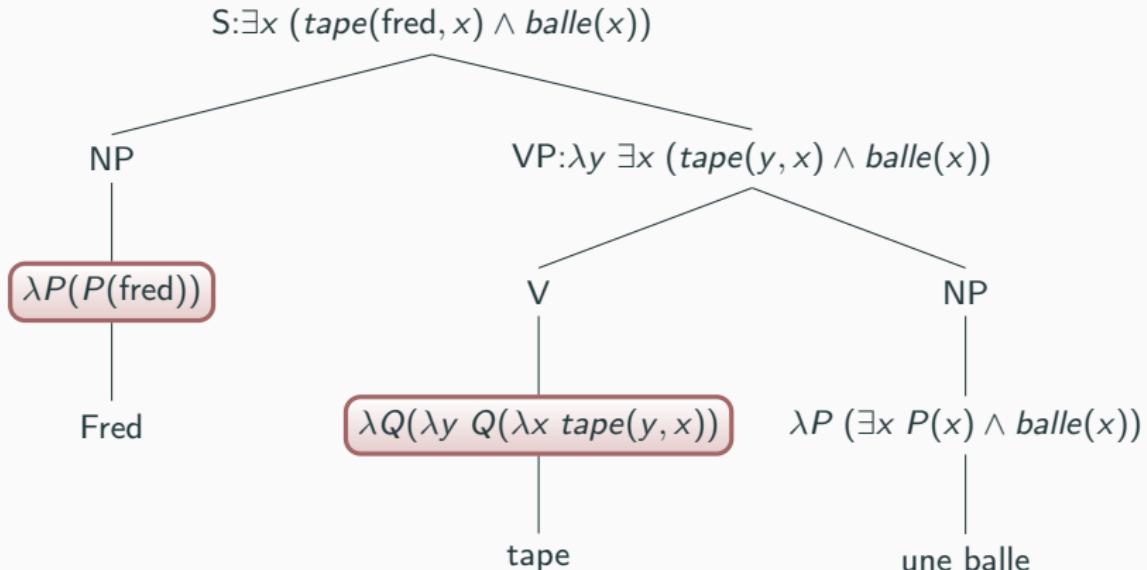
Fred tape une balle.

Exemple: nom propre, verbe transitif



Fred tape une balle.

Exemple: nom propre, verbe transitif



Fred tape une balle.

Problème de cette approche

- demande une couverture parfaite du lexique pour être applicable:
peu robuste
- déterministe sur l'interprétation, alors qu'il peut y avoir des ambiguités sémantiques sur une même structure syntaxique:
Marie aime bien Jean mais elle n'aime pas son frère.
- ne dit rien sur le sens des mots, seulement la combinaison des sens:
Jean est célibataire $celibataire(j)$
Jean n'est pas marié $\neg marie(j)$

Suite du cours

- → importance des relations entre mots et leurs sens : **sémantique lexicale**
 - Sens des mots
 - Principales ressources sémantiques existantes
 - Désambiguïsation lexicale (WSD)
- → composition des sens des mots au sein de la phrase :
sémantique de la phrase
 - Rôles sémantiques
 - Principales ressources existantes
 - Etiquetage des rôles sémantiques (SRL)

Suite du cours

- → importance des relations entre mots et leurs sens : **sémantique lexicale**
 - Sens des mots
 - Principales ressources sémantiques existantes
 - Désambiguïsation lexicale (WSD)

Sémantique lexicale

Petit rappel

- **lemme**: unité constituante du lexique, forme canonique du mot ('entrée de dictionnaire')
- **forme de mot**: mot fléchi tel qu'il apparaît dans un texte

forme de mot	lemme
tables	table
grandes	grand
buvions	boire

- Dans la sémantique lexicale, on traite les lemmes de mots
- Dans un contexte sémantique, on parle de **lexème**
- Tous les lexèmes constituent le **lexique** d'une langue

Le(s) sens d'un mot

- Un mot (lexème) peut avoir différents sens
- e.g. *avocat*

Le(s) sens d'un mot

- Un mot (lexème) peut avoir différents sens
- e.g. *avocat*



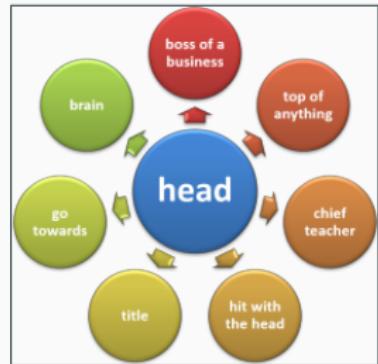
Le(s) sens d'un mot

- Un mot (lexème) peut avoir différents sens
- e.g. *avocat*



Le(s) sens d'un mot

- Un mot (lexème) peut avoir différents sens
- e.g. *avocat* ; *head*



Le(s) sens d'un mot

- Un mot (lexème) peut avoir différents sens
- e.g. *avocat* ; *head*



- les **sens**: représentations discrètes de différents aspects de la signification d'un mot
- le lexème *avocat* a deux sens dans cet exemple, mais les mots peuvent être très polysémiques

Synonymie

- Des mots qui ont le même sens sont des **synonymes**
 - auto / voiture
 - eau / H₂O
 - colossal / énorme
 - patate / pomme de terre
- deux lexèmes sont synonymes si l'un peut être substitué par l'autre dans tous les contextes
- ils ont alors la même signification propositionnelle (dénotation)

Synonymie

- Mais: il y a très peu de synonymes parfaits
- Beaucoup d'aspects de la signification peuvent être le même
- Bien que d'autres aspects ne sont pas préservés (par rapport à la politesse, argot, registre, genre, ...)
 - mec / homme
 - clope / cigarette
 - eau / H_2O

Synonymie

- La synonymie est une relation entre sens plutôt qu'entre mots
- Exemple: grand / large
- Ils sont synonymes ?
 - une *grande* étude a été menée
 - une *large* étude a été menée
- Mais:
 - elle est comme une *grande* soeur pour moi
 - ? elle est comme une *large* soeur pour moi
- Pourquoi ?
 - *grand* a le sens 'plus vieux, adulte'
 - *large* n'a pas ce sens

Les relations entre mots

Antonymie

- Des sens qui sont opposés par rapport à un aspect de leur signification
- Autrement, ils sont très similaires
 - court / long
 - chaud / froid
 - haut / bas
 - accord / désaccord
 - begin / end

Hyponymie et hypéronymie

- un sens est un **hyponyme** d'un autre si le premier sens est plus spécifique (un sous-classe) de l'autre
 - *voiture* est un hyponyme de *véhicule*
 - *pomme* est un hyponyme de *fruit*
 - *vin* est un hyponyme de *boisson*
- l'inverse est un **hypéronyme**
 - *véhicule* est un hypéronyme de *voiture*
 - *fruit* est un hypéronyme de *pomme*
 - *boisson* est un hypéronyme de *vin*
- On appelle la relation d'hyponymie également **IS-A**
 - *vin* IS-A *boisson*

Hyponymie et hypéronymie

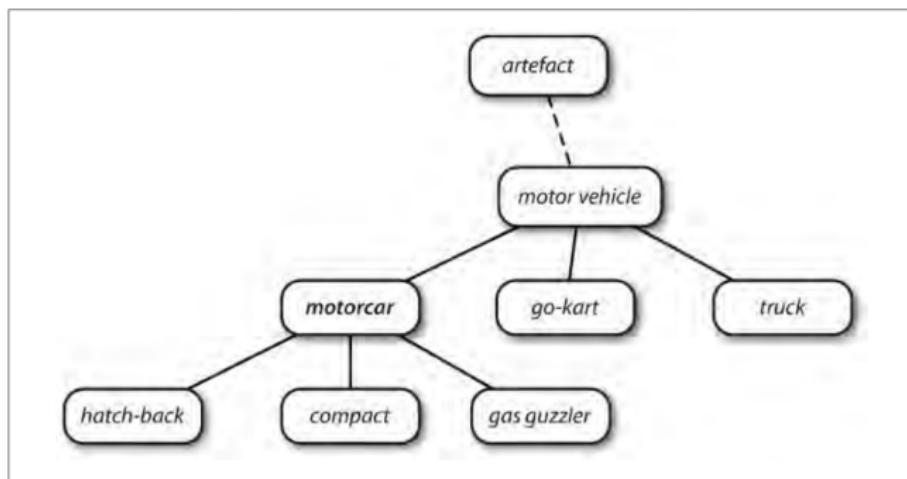


Figure 2-8. Fragment of WordNet concept hierarchy: Nodes correspond to synsets; edges indicate the hypernym/hyponym relation, i.e., the relation between superordinate and subordinate concepts.

Homonymie

Homonymie

- L'homonymie pose des problèmes pour le TAL
- Recherche d'information: “prix d'un avocat”
- Traduction automatique: FR *avocat* → EN *lawyer or avocado*
- synthèse vocale: *couvent* (nom) vs. *couvent* (verbe)

Les relations entre mots

Polysémie

1. La construction de l'université a été interrompue par la première Guerre Mondiale.
 2. Il a commencé à l'université à l'âge de 17 ans.
- Est-ce que c'est le même sens?
 - Sens 2: centre d'enseignement supérieur
 - Sens 1: les bâtiments d'un centre d'enseignement supérieur
 - Un mot **polysémique** a des sens qui sont reliés
 - La plupart des mots sont polysémiques

Polysémie

- Beaucoup de types de polysémie sont systématiques
 - *école, université, banque, hôpital*
 - une relation systématique
 - 'institution' ↔ 'bâtiment'
- Différents exemples:
 - auteur ↔ livre de l'auteur ('J'aime Baudelaire')
 - contenant ↔ contenu ('Je bois un verre')
- aussi connu comme **métonymie**

Ressources lexicales : Wordnet

- La ressource de sémantique lexicale la plus utilisée en anglais
- Elle contient trois bases de données (Wordnet v3.0):
 - verbes (11.488 entrées)
 - noms (117.097 entrées)
 - adjectives (22.141) et adverbes (4.601)
- Chaque lemma contient les champs suivants:
 - gloss:** définition (comme dans un dictionnaire classique)
 - exemples:** exemples d'utilisation (pas pour tous les lemmas)
 - synset:** **synonyme set** un ensemble de presque synonymes.
On pourrait dire qu'un synset représente un **concept**.
 - relations:** relations lexicales entre les synsets.

nouns

hypernym: from concepts to superordinates (breakfast → meal)

hyponym: from concepts to subtypes (meal → lunch)

member meronym: from groups to their members (faculty → professor)

has-instance: from concepts to instances of the concept (composer → Bach)

instance: from instances to their concepts (Austen → author)

member holonym: from members to their groups (copilot → crew)

part meronym: from wholes to parts (table → leg)

part holonym: from parts to wholes (course → meal)

antonym: opposites (leader → follower)

verbs

hypernym: from events to superordinate events (fly → travel)

troponym: from a verb (event) to a specific manner elaboration of that event
(walk → stroll / flâner)

entails: from verbs (events) to the verbs (events) they entail (snore → sleep)

antonym: opposites (increase → decrease)

adjectives and adverbs

antonym: opposite adjectives (heavy → light)

antonym: opposite adverbs (quickly → slowly)

Sémantique lexicale : exercices

- Allez sur le site WordNet:
<http://wordnetweb.princeton.edu/perl/webwn>. Tapez un mot et explorez les relations qui le relient aux autres mots.
- Allez sur le notebook TD-WordNet disponible sur Moodle et explorer les différentes relations sémantiques ainsi que le calcul de similarité entre les concepts
- Wordnet dans d'autres langues. Explorez les langues actuellement couvertes (<https://github.com/goodmami/wn>). Qu'observez vous?
- D'après vous, comment cette ressource peut-elle être utilisée dans des applications TAL/RI?

Sémantique lexicale : autres ressources (lexicales) existantes

BabelNet (<https://babelnet.org/>) combine un ensemble de ressources multilingues

- dictionnaire encyclopédique (wikipedia)
- réseau sémantique (concepts, termes, entités nommées) dans +200 langues; 13 millions d'entrées
- analyseur de termes / détection d'entités

Et aussi:

- SentiWordNet (<https://github.com/aesuli/SentiWordNet>)
- ConceptNet (<https://conceptnet.io/>)
-

Sémantique lexicale : Word sense disambiguation

On a un inventaire fixe de sens de mots, e.g. WordNet: sens/synset

- une définition (gloss) + des exemples (souvent)

→ désambiguïser des instances spécifiques des mots dans un contexte

- Exemple: pièce

- Ce **concert** réunit des **airs** pour **soprano** et des **pièces** pour **violon**.
→ *composition musicale*
- Elle découvrit une **pièce immense**, au **parquet** brillant.
→ *salle, chambre*

- Approches :

- Heuristique - Lesk : comparer l'ensemble des mots de la définition vs le contexte + Variante: Augmenté avec des mots similaires (e.g. méthodes distributionnelles...)
- Apprentissage :
 - supervisé = nécessite des corpus annotés, tâche difficile selon la granularité (même pour un humain)
 - non-supervisé = discrimination, sémantique distributionnelle et clustering

Composition au sein de la phrase

- → composition des sens des mots au sein de la phrase :
sémantique de la phrase
 - Rôles sémantiques
 - Principales ressources existantes
 - Etiquetage des rôles sémantiques (SRL)

Rôles thématiques: intuition

jusqu'ici on s'est concentré sur les *prédictats*

les arguments sont importants aussi

- (Je) veux (le livre)
(NP) **vouloir** (NP)
- (Le gros chien noir) a voulu (manger un petit os)
(NP) **vouloir** (VP)
- Il y a une structure prédicative (un prédictat et un certain nombre d'arguments) pour un verbe
- Cette structure prédicative ne se limite pas aux verbes, mais également aux autres parties du discours comme les noms et les adjectifs. Ex:
 - **L'acquisition** (de l'entreprise A) (par l'entreprise B).
 - **Fier** (de son score) (Nicolas) (souriait constamment).

Rôles thématiques: intuition, encore

- *Nicholas a cassé la fenêtre.*
- *Marie ouvre la porte.*

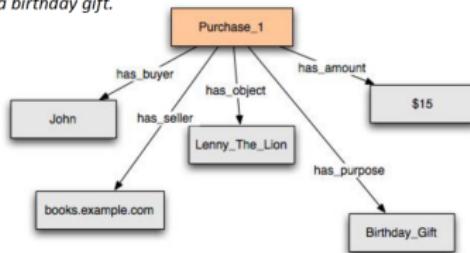
On observe que:

- il y a une action (casser)
- il y a un agent de cette action (le casseur)
- il y a un patient de cette action (l'objet cassé)

On appelle les arguments d'un prédicat **rôles thématiques**
→ catégories d'arguments généraux

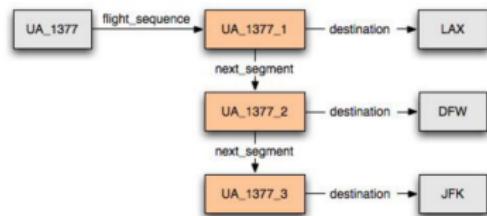
Rôles thématiques: intuition

John buys a "Lenny the Lion" book from books.example.com for \$15 as a birthday gift.



(c)

United Airlines flight 3177 visits the following airports: LAX, DFW, and JFK



(d)

Pourquoi détecter la structure argumentale d'une phrase?

Ensemble des rôles thématiques

- Il n'y a pas un ensemble de rôles thématiques accepté par tout le monde
- Néanmoins, il y a un petit ensemble de rôles thématiques plus ou moins accepté par la majorité des chercheurs.

Rôles thématiques: exemples i

agent: The volitional causer of an event

The waiter spilled the soup

experiencer: The experiencer of an event

John has a headache

patient/theme The entity undergoing the effect of some action, often undergoing some change of state / The participant most directly affected by an event

Nicolas a coupé le gazon

Fred threw the rock.

percept/stimulus: the entity which is perceived or experienced.

Mary fears thunder.

instrument: An instrument used in an event

Fred opened the lock with a paper clip.

Rôles thématiques: exemples ii

beneficiary: The beneficiary of an event

*Mary made a hotel reservation **for her boss.***

source: The origin of the object of a transfer event

*I flew **from Paris.***

goal: The destination of an object of a transfer event

*I drove **to Marseille.***

Rôles thématiques : Principales ressources existantes

- PropBank (<https://propbank.github.io/>): The Proposition Bank (Kingsbury and Palmer, 2002)
 - Penn TreeBank (40 000 phrases)
 - rôles thématiques (ARG0, ARG1,..., ARG5)
 - conçu pour les prédicts verbaux (cf NomBank (Meyers et al., 2004))
- FrameNet (Fillmore, 1976)
<https://framenet.icsi.berkeley.edu/>
 - rôles sémantiques (vs thématiques) associés à un cadre (frame) (vs un verbe spécifique)
 - des rôles nécessaires (core) vs non nécessaires (peripheral) + contextuels (extra-thematic)

Permet de s'abstraire de la réalisation syntaxique et de se focaliser uniquement sur le sens de la phrase

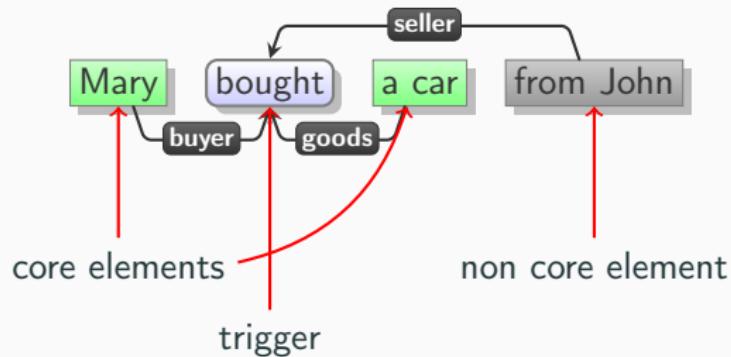
[_{Arg0} Big Fruit Co.] increased [_{Arg1} the price of bananas].

[_{Arg1} The price of bananas] was increased again [_{Arg0} by Big Fruit Co.]

[_{Arg1} The price of bananas] increased [_{Arg2} 5%]

Framenet: Exemple

Commerce_buy

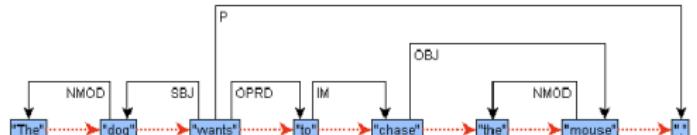


Equivalents à :

- John sold a car to Mary
- A car was sold to Mary by John
- The selling of a car by John to Mary (failed).
- ...

Comparaisons des ressources

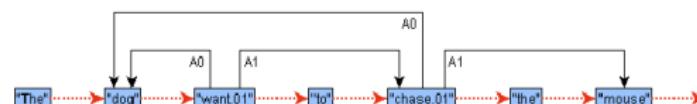
surface-syntax



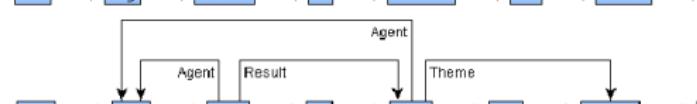
deep-syntax



PropBank/NomBank



VerbNet



FrameNet

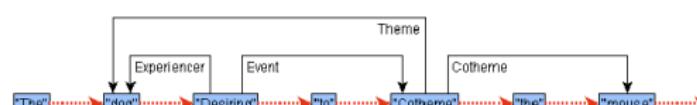


Figure extraite du tutoriel (http://taln.upf.edu/pages/eswc2016-tutorial/ESWC_tuto_RelExtr_v1.0.pdf)

Semantic Role Labeling (SRL)

- Identifier automatiquement les prédicats de la phrase
- Étiquetage des rôles thématiques
- Evaluation des systèmes : extrinsèque vs. intrinsèque (discussion)

Démo :

https://cogcomp.seas.upenn.edu/page/demo_view/SRLEnglish

Modèle BIO :

<https://paperswithcode.com/task/semantic-role-labeling>

Frame-SRL : Performances

	CoNLL05												CoNLL12											
	Dev			WSJ			Brown			Dev			Test											
	F ₁	P	R	F ₁	P	R	F ₁	F ₁	P	R	F ₁	F ₁	P	R	F ₁	F ₁	P	R	F ₁	F ₁	P	R	F ₁	
He et al. (2017)	80.3	80.2	82.3	81.2	67.6	69.6	68.5	75.5	78.6	75.1	76.8	-	-	-	-	-	-	-	-	-	-	-	-	-
He et al. (2018a)	81.6	81.2	83.9	82.5	69.7	71.9	70.8	79.4	79.4	80.1	79.8	-	-	-	-	-	-	-	-	-	-	-	-	-
Li et al. (2019)	-	-	-	83.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhou et al. (2020a)	82.27	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF	83.70	83.18	85.38	84.27	70.40	72.97	71.66	81.03	79.47	82.80	81.10	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF2O	83.91	83.26	86.20	84.71	70.70	74.16	72.39	81.16	79.27	83.24	81.21	-	-	-	-	-	-	-	-	-	-	-	-	-
Li et al. (2019) _{ELMo}	-	85.2	87.5	86.3	74.7	78.1	76.4	-	84.9	81.4	83.1	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhou et al. (2022) _{BERT}	86.79	87.15	88.44	87.79	79.44	80.85	80.14	84.74	83.91	85.61	84.75	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF _{BERT}	86.82	86.98	88.28	87.63	79.19	80.92	80.05	85.35	84.47	86.24	85.35	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF2O _{BERT}	87.03	87.00	88.76	87.87	79.08	81.50	80.27	85.53	84.53	86.41	85.45	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF _{ROBERTA}	87.31	87.20	88.67	87.93	79.29	81.48	80.38	86.08	84.98	86.86	85.91	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF2O _{ROBERTA}	87.46	87.35	89.34	88.33	79.95	82.32	81.12	86.34	85.30	87.02	86.15	-	-	-	-	-	-	-	-	-	-	-	-	-
<i>w/ gold predicates</i>																								
He et al. (2017)	81.6	83.1	83.0	83.1	72.9	71.4	72.1	81.5	81.7	81.6	81.7	-	-	-	-	-	-	-	-	-	-	-	-	-
Ouchi et al. (2018)	82.5	84.7	82.3	83.5	76.0	70.4	73.1	82.9	84.4	81.7	83.0	-	-	-	-	-	-	-	-	-	-	-	-	-
Tan et al. (2018)	83.1	84.5	85.2	84.8	73.5	74.6	74.1	82.9	81.9	83.6	82.7	-	-	-	-	-	-	-	-	-	-	-	-	-
Strubell et al. (2018)	-	84.7	84.24	84.47	73.89	72.39	73.13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhou et al. (2020a)	83.16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhang et al. (2021b)	84.45	85.30	85.17	85.23	74.98	73.85	74.41	82.83	83.09	83.71	83.40	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF	84.42	85.38	85.56	85.47	75.05	74.05	74.55	83.22	83.21	83.85	83.53	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF2O	84.65	85.47	86.40	85.93	74.92	75.00	74.96	83.39	83.02	84.31	83.66	-	-	-	-	-	-	-	-	-	-	-	-	-
Strubell et al. (2018) _{ELMo}	85.26	86.21	85.98	86.09	77.1	75.61	76.35	83.23	84.39	82.21	83.28	-	-	-	-	-	-	-	-	-	-	-	-	-
Shi and Lin (2019) _{BERT}	-	88.6	89.0	88.8	81.9	82.1	82.0	-	85.9	87.0	86.5	-	-	-	-	-	-	-	-	-	-	-	-	-
Jindal et al. (2020) _{BERT}	-	88.7	88.0	87.9	80.3	80.1	80.2	-	86.3	86.8	86.6	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhang et al. (2021b) _{BERT}	87.38	87.70	88.15	87.93	81.52	81.36	81.44	86.27	86.00	86.84	86.42	-	-	-	-	-	-	-	-	-	-	-	-	-
Zhou et al. (2022) _{BERT}	87.54	89.03	88.53	88.78	83.22	81.81	82.51	86.97	87.26	87.05	87.15	-	-	-	-	-	-	-	-	-	-	-	-	-
Conia and Navigli (2020) _{BERT}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Biloshmi et al. (2021) _{BART}	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	87.8	86.8	87.3	-	-	-	
CRF _{BART}	87.76	88.93	88.58	88.76	82.87	81.67	82.27	87.33	87.45	87.56	87.51	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF2O _{BART}	88.05	89.00	89.03	89.02	82.81	83.35	82.58	87.52	87.52	87.79	87.66	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF _{ROBERTA}	88.21	89.29	88.99	89.15	83.22	82.42	82.82	87.97	87.99	88.22	88.11	-	-	-	-	-	-	-	-	-	-	-	-	-
CRF2O _{ROBERTA}	88.49	89.45	89.63	89.54	83.89	83.39	83.64	88.29	88.11	88.53	88.32	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 2: Results on CoNLL05 and CoNLL12 data. All results are averaged over 4 runs with different random seeds.

- **SRL scores (2022):** <https://aclanthology.org/2022.coling-1.370/> ;
- **NMT basé sur représentations sémantiques (2018)** <https://aclanthology.org/N18-2078/>
- **Sentiment analysis (2022)** <https://aclanthology.org/2022.acl-long.314/>

avantages

- On peut faire des inférences
- on peut **représenter**—*qui a fait quoi à qui, où, quand, comment.*
- utile: questions/réponses, résumés, détection d'opinions, etc

problèmes

- Pas de consensus sur un ensemble commun des rôles thématiques
- Difficile de produire des définitions formelles → annotation difficile et manque d'homogénéité = apprentissage automatique difficile
- Manque de précision: e.g. différentes sortes d'agent selon le prédicat (agents vivant, agents avec volition, conscient, etc).

Introduction

- Synonymie: relation binaire
 - Deux mots sont synonymes, ou ils ne le sont pas
- Similarité (ou distance): métrique moins stricte
 - Deux mots sont similaires s'ils partagent plus de traits
- Théoriquement, la similarité est une relation entre **sens**
 - avocat¹ est similaire à juriste¹
 - avocat² est similaire à mangue¹
- En pratique, on calcule la similarité entre mots

Problèmes avec la similarité de mots à base d'un thésaurus

- On a besoin d'un thésaurus
 - ça n'existe pas pour chaque langue
- s'il existe, il y a souvent des problèmes de couverture ("rappel")
 - il manque des mots
 - il manque des liens entre sens
 - certains domaines sont mieux représentés que d'autres
- la méthode est moins appropriée pour les verbes et les adjectifs
 - il est plus difficile de structurer les verbes et les adjectifs de manière hiérarchique
- les ressources construites de manière semi-automatique contiennent beaucoup d'erreurs (problème de "précision")

Similarité distributionnelle

- L'induction de sens à partir de texte est basés sur l' hypothèse distributionnelle
- Prenons un mot et ses contextes :
 - *sooluceps* délicieux
 - *sooluceps* sucré
 - *sooluceps* rassis
 - *sooluceps* fraîchement sorti du four
- En regardant les contextes d'un mot, on peut en induire le sens
 - Zellig Harris (1954): "*If A and B have almost identical environments we say that they are synonyms*".
 - Firth (1957): "*You shall know a word by the company it keeps!*"

Similarité distributionnelle

- L'induction de sens à partir de texte est basés sur l' hypothèse distributionnelle
- Prenons un mot et ses contextes :
 - *soulceps* délicieux
 - *soulceps* sucré
 - *soulceps* rassis
 - *soulceps* fraîchement sorti du four

⇒ **nourriture**
- En regardant les contextes d'un mot, on peut en induire le sens
 - Zellig Harris (1954): "*If A and B have almost identical environments we say that they are synonyms*".
 - Firth (1957): "*You shall know a word by the company it keeps!*"

Similarité distributionnelle

- L'induction de sens à partir de texte est basés sur l' hypothèse distributionnelle
- Prenons un mot et ses contextes :
 - *sooluceps* délicieux
 - *sooluceps* sucré
 - *sooluceps* rassis
 - *sooluceps* fraîchement sorti du four
- En regardant les contextes d'un mot, on peut en induire le sens
 - Zellig Harris (1954): "*If A and B have almost identical environments we say that they are synonyms*".
 - Firth (1957): "*You shall know a word by the company it keeps!*"



Distributional similarity

The functional interplay of philosophy and
...and among works of dystopian
The rapid advance in
...calculus, which are more popular in
But because
...the value of opinions formed in
...if
...certainty which every new discovery in
...if the new technology of computer
...frightened by the powers of destruction
...but there is also specialization in

should, as a minimum, guarantee..
fiction...
today suggests...
-oriented schools.
is based on mathematics...
as well as in the religions...
can discover the laws of human nature
either replaces or reshapes.
is to grow significantly
has given
and technology...

Distributional similarity

The functional interplay of philosophy and
...and among works of dystopian

The rapid advance in
...calculus, which are more popular in

But because
...the value of opinions formed in
...if

...certainty which every new discovery in
...if the new technology of computer
...frightened by the powers of destruction
...but there is also specialization in

science should, as a minimum, guarantee..
science fiction...
science today suggests...
science -oriented schools.
science is based on mathematics...
science as well as in the religions...
science can discover the laws of human nature
science either replaces or reshapes.
science is to grow significantly
science has given
science and technology...

Différentes sortes de similarité

- Souvent, on distingue des mots similaires et des mots associés
- mots similaires: (quasi-)synonymes ou co-hyponymes
 - poire, pomme, banane, cerise
 - lundi, mercredi, vendredi
- mots associés: relation quelconque, similarité de 'topique', thématique
 - football, joueur, dopage, médaille
 - voiture, roue, autoroute, rétroviseur

Différentes sortes de similarité

Exemple

- **médecin:** *docteur, médecin de famille, chirurgien, spécialiste, dermatologue, gynécologue*
- **médecin:** *malade, maladie, diagnostic, traitement, hôpital, stéthoscope*

Relation entre contexte et similarité

- Différentes contextes donnent lieu à différentes notions de similarité
- relations de dépendances ou fenêtre étroite ↔ fenêtre large, documents
- Les premiers modèles induisent une **similarité stricte, comme des synonymes**
- Les derniers modèles induisent une **similarité associative, thématique**

Similarité lexicale

- Méthode statique → A partir de vecteurs sparses (matrice de co-occurrences)
- Méthode statique → A partir de vecteurs denses (Singular Value Decomposition-SVD)
- Méthode prédictive → A partir de vecteurs denses (Word embeddings – plongement de mots)

Similarité lexicale

- Méthode statique → A partir de vecteurs sparses (matrice de co-occurrences)
- Méthode statique → A partir de vecteurs denses (Singular Value Decomposition-SVD)
- Méthode prédictive → A partir de vecteurs denses (Word embeddings – plongement de mots)

Etape 1: prétraitement linguistique

Cécile Duflot n'exclut pas la réquisition de logements vacants.

1. tokenisation

- Cécile_Duflot n' exclut pas la réquisition de logements vacants .

2. normalization (lemmatisation)

- Cécile_Duflot ne exclude pas la réquisition de logement vacant .

3. annotation (part of speech tagging, parsing)

- Cécile_Duflot/NPP ne/ADV exclude/V pas/ADV la/DET réquisition/NC de/P logement/NC vacant/ADJ .

Etape 2: Traitement mathématique

1. Construction d'une matrice de fréquence
2. Pondération des éléments de la matrice
3. Calcul de similarité

Etape 2: Traitement mathématique

1. Construction d'une matrice de fréquence

Construction d'une matrice de fréquence

Plusieurs types de matrices peuvent être construites

- Matrice termes-document → how often a word occurs in a document
- Matrice termes-termes → how often a word occurs with another
- Matrice termes-contexte → how often a word occurs with a given context

Approches statiques dites Bag of Words (Sac de mots)

Matrice terme-document

Each cell: count of word w in a document d :

- Each document is a **count vector** in \mathbb{N}^v : a column below

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Slides taken from D. Jurafsky's lexical semantics course

Two documents are similar if their vectors are similar

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Slides taken from D. Jurafsky's lexical semantics course

Matrice terme-document

Each word is a **count vector** in \mathbb{N}^D : a row below

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Slides taken from D. Jurafsky's lexical semantics course

Two words are similar if their vectors are similar

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	1	8	15
soldier	2	2	12	36
fool	37	58	1	5
clown	6	117	0	0

Construction d'une matrice de fréquence termes-termes

- enregistre les fréquences de co-occurrence de deux entités
- se focalise uniquement sur les associations syntagmatiques (nearby terms)

Matrice de co-occurrence

- enregistre les fréquences de co-occurrence de deux entités

	rouge	délicieux	rapide	d'occasion
framboise	2	1	0	0
fraise	2	2	0	0
voiture	1	0	1	2
camion	1	0	1	1

Matrice de co-occurrence

- enregistre les fréquences de co-occurrence de deux entités

	rouge	délicieux	rapide	d'occasion
framboise	7	9	0	0
fraise	12	6	0	0
voiture	7	0	8	4
camion	2	0	3	4

Matrice de co-occurrence

- enregistre les fréquences de co-occurrence de deux entités

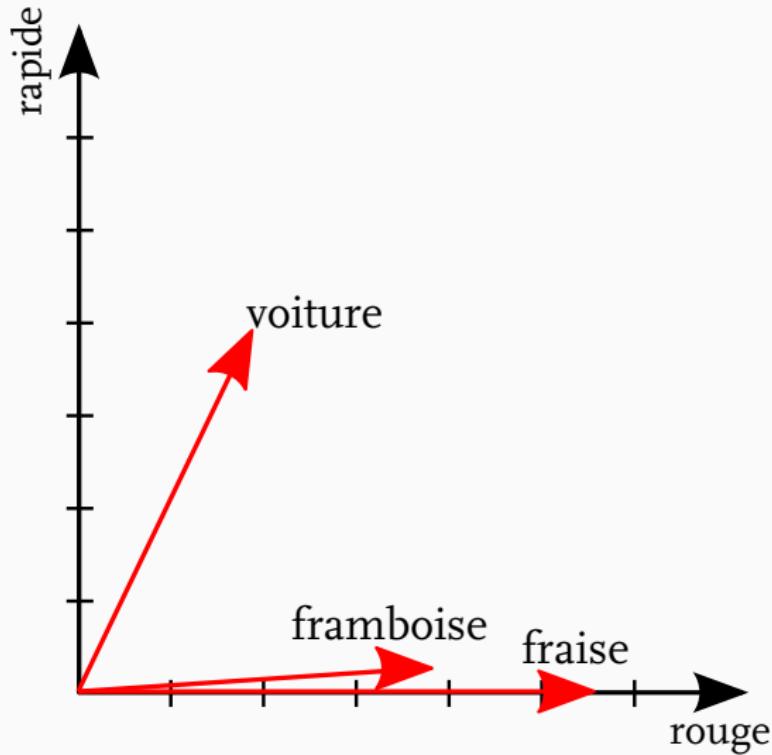
	rouge	délicieux	rapide	d'occasion
framboise	56	98	0	0
fraise	44	34	0	0
voiture	23	0	31	39
camion	4	0	18	29

Matrice de co-occurrence

- enregistre les fréquences de co-occurrence de deux entités

	rouge	délicieux	rapide	d'occasion
framboise	728	592	1	0
fraise	1035	437	0	2
voiture	392	0	487	370
camion	104	0	393	293

Modèle vectoriel



Matrice mot-contexte

	(obj,drove)	(mod,second-hand)	context3	context4
car				
word2				
word3				
word4				

- Différentes notions de contexte
 - fenêtre autour du mot
 - relations de dépendances (extrait d'un corpus parsé) —> **Relation paradigmatiques**

He drove his second-hand **car** a couple of miles down the road .

Matrice mot-contexte

	(obj,drove)	(mod,second-hand)	context3	context4
car				
word2				
word3				
word4				

- Différentes notions de contexte
 - fenêtre autour du mot (2 mots)
 - relations de dépendances (extrait d'un corpus parsé) —> **Relation paradigmatiques**

He drove [his second-hand car a couple] of miles down the road .

Matrice mot-contexte

	(obj,drove)	(mod,second-hand)	context3	context4
car				
word2				
word3				
word4				

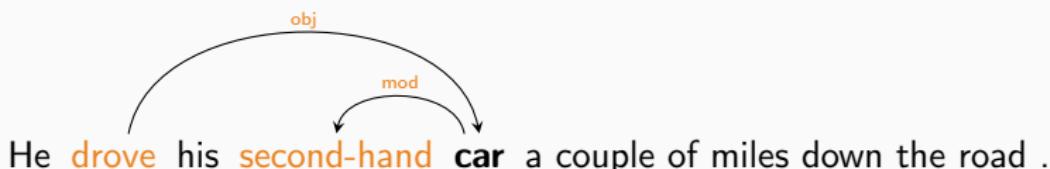
- Différentes notions de contexte
 - fenêtre autour du mot (phrase)
 - relations de dépendances (extrait d'un corpus parsé) —> **Relation paradigmatisques**

[He drove his second-hand car a couple of miles down the road .]

Matrice mot-contexte

	(obj,drove)	(mod,second-hand)	context3	context4
car	1	1		
word2				
word3				
word4				

- Différentes notions de contexte
 - fenêtre autour du mot
 - relations de dépendances (extrait d'un corpus parsé) → Relation paradigmatisques



Etape 2: Traitement mathématique

1. Construction d'une matrice de fréquence
2. Pondération des éléments de la matrice

Pondération des éléments de la matrice

- On veut donner plus de poids aux co-occurrences ‘surprenantes’, et moins aux co-occurrences attendues
- → (positive) pointwise mutual information (PPMI)
- Quantification du décalage entre la probabilité de la distribution jointe et la probabilité d’entités individuelles si on suppose qu’elles sont indépendantes
- $pmi(x, y) = \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$
- $ppmi(x, y) = \text{Max}\left(\log\left(\frac{p(x,y)}{p(x)p(y)}\right), 0\right)$
(généralement valeurs négatives problématiques)

Pondération des éléments de la matrice

- $pmi(x, y) = \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$

	rouge	délicieux	rapide	d'occasion	grand
framboise	728	592	0	0	823
fraise	1035	633	0	0	890
voiture	392	0	487	370	920
camion	104	0	393	293	846
banane	0	489	0	0	500

Pondération des éléments de la matrice

- $pmi(x, y) = \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$

	rouge	délicieux	rapide	d'occasion	grand
framboise	.08	.06	0	0	.09
fraise	.11	.07	0	0	.09
voiture	.04	0	.05	.04	.10
camion	.01	0	.04	.03	.09
banane	0	.05	0	0	.05

Pondération des éléments de la matrice

- $pmi(x, y) = \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$

	rouge	délicieux	rapide	d'occasion	grand	$p(X)$
framboise	.08	.06	0	0	.09	.23
fraise	.11	.07	0	0	.09	.27
voiture	.04	0	.05	.04	.10	.23
camion	.01	0	.04	.03	.09	.17
banane	0	.05	0	0	.05	.10
$p(Y)$.24	.18	.09	.07	.42	

Pondération des éléments de la matrice

- $pmi(x, y) = \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$

	rouge	délicieux	rapide	d'occasion	grand
framboise	.36	.42	0	0	-.09
fraise	.53	.31	0	0	-.19
voiture	-.27	0	.88	.89	.01
camion	-1.31	0	.95	.94	.21
banane	0	1.00	0	0	.19

Pondération des éléments de la matrice

- $pmi(x, y) = \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$

	rouge	délicieux	rapide	d'occasion	grand
framboise	.36	.42	0	0	0
fraise	.53	.31	0	0	0
voiture	0	0	.88	.89	.01
camion	0	0	.95	.94	.21
banane	0	1.00	0	0	.19

Attention : La PPMI est biaisée envers les occurrences peu fréquentes/rares (PMI très élevée) → lissage parfois appliqué

Etape 2: Traitement mathématique

1. Construction d'une matrice de fréquence
2. Pondération des éléments de la matrice
3. Calcul de similarité

Calcul de similarité ...

Cosinus

- $\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$
- $\cos(x) \in [-1, 1]$
- exemples:
 - $\cos(framboise, fraise) = .96$
 - $\cos(framboise, voiture) = .42$
- **voiture:** *véhicule, moto, avion, automobile, camion, train, bateau,*
...
- **fraise:** *framboise, groseille, mûre, myrtille, ...*

Calcul de similarité ...

Mais avec d'autres mesures possibles

$$\begin{aligned}\text{simJaccard}(\vec{v}, \vec{w}) &= \frac{\sum_{i=1}^N \min(v_i, w_i)}{\sum_{i=1}^N \max(v_i, w_i)} \\ \text{simDice}(\vec{v}, \vec{w}) &= \frac{2 \times \sum_{i=1}^N \min(v_i, w_i)}{\sum_{i=1}^N (v_i + w_i)} \\ \text{simJS}(\vec{v} || \vec{w}) &= D\left(\vec{v} \middle| \frac{\vec{v} + \vec{w}}{2}\right) + D\left(\vec{w} \middle| \frac{\vec{v} + \vec{w}}{2}\right)\end{aligned}$$

Similarité lexicale

- Méthode statique → A partir de vecteurs sparses (matrice de co-occurrences)
- **Méthode statique → A partir de vecteurs denses (Singular Value Decomposition-SVD)**
- Méthode prédictive → A partir de vecteurs denses (Word embeddings – plongement de mots)

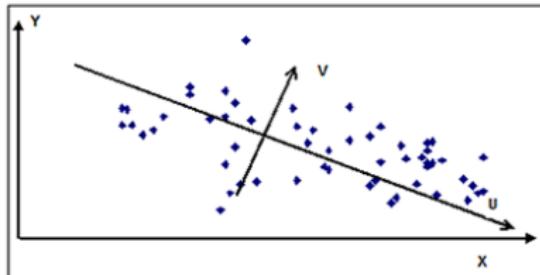
Lissage de la matrice

Deux raisons pour effectuer une réduction de dimensionalité:

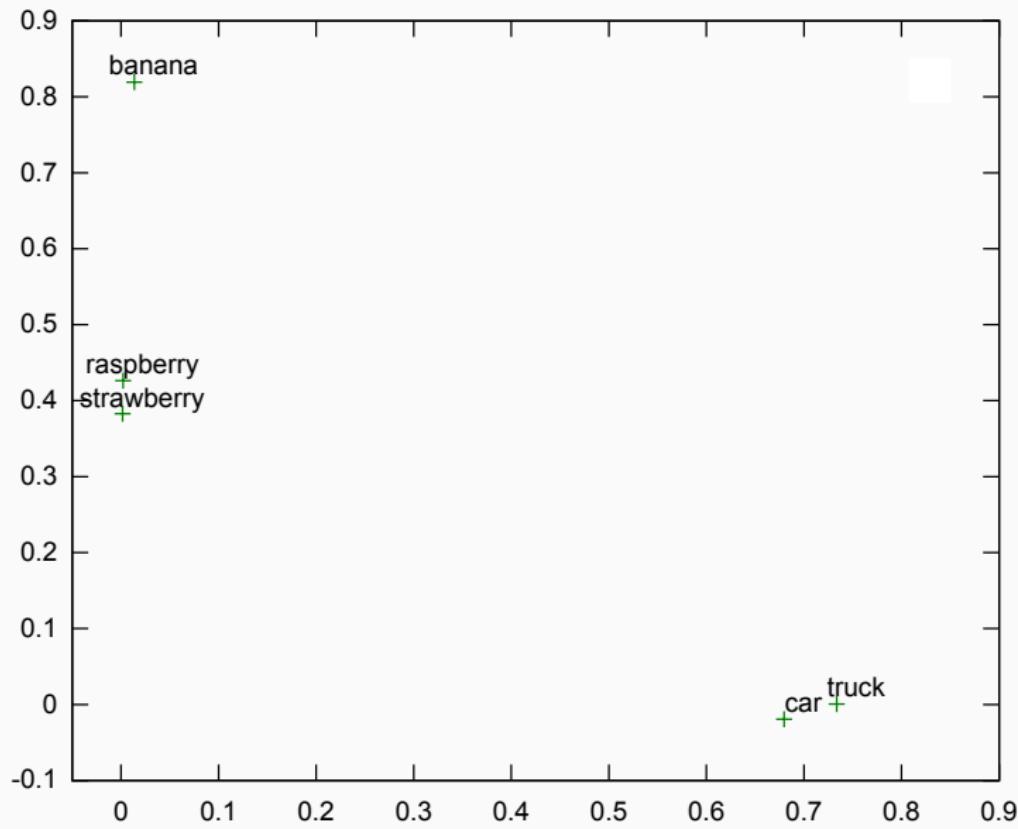
- Les calculs sont intractables
 - Le nombre d'instances et de traits est trop grand pour faire des calculs de similarité
 - une réduction de dimensionalité rend cela de nouveau traitable
- Capacité de généralisation
 - La réduction de dimensionalité est capable d'encoder les données d'une meilleure façon, en capturant des traits sémantiques intrinsèques
 - une réduction de dimensionalité améliore ainsi les résultats (en corrigeant la parsimonie et le bruit dans les données)

Décomposition en valeurs singulières (svd)

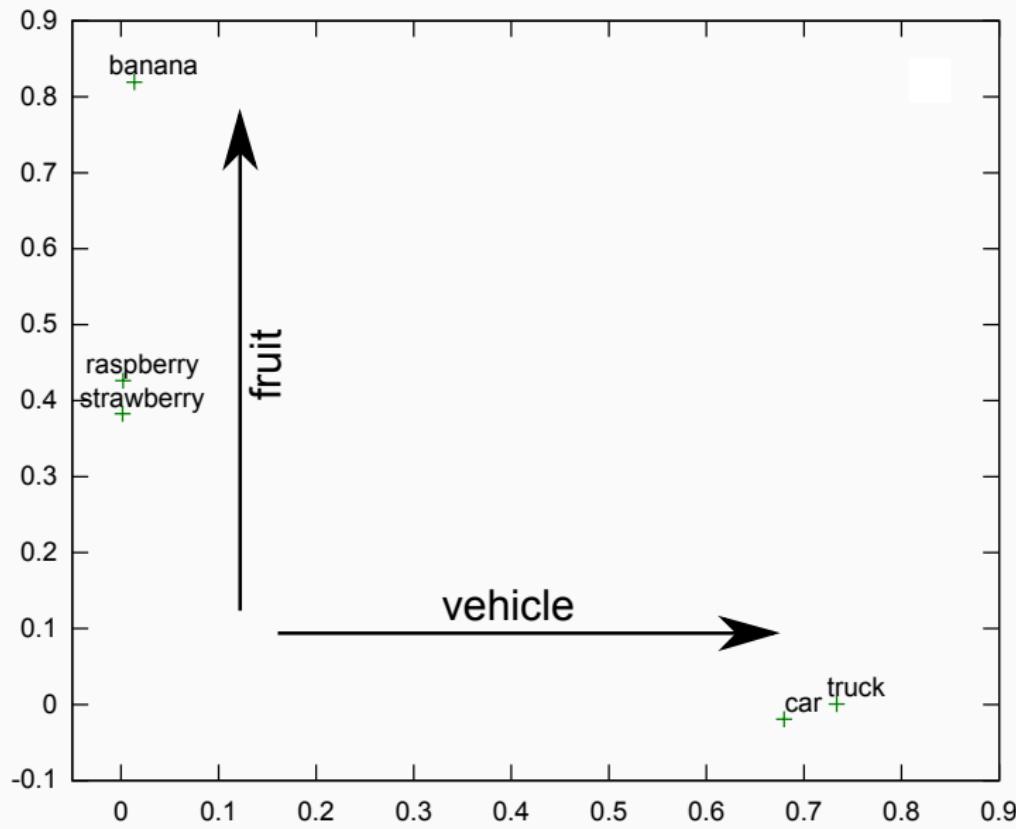
- Technique populaire pour une réduction de dimensionnalité
- Avec svd on va décomposer une matrice \mathbf{X} en un produit de trois matrices $\mathbf{U}\mathbf{S}\mathbf{V}^T$
- \mathbf{U} contient les vecteurs de base pour les instances, \mathbf{V} contient les vecteurs de base pour les traits, \mathbf{S} est une matrice diagonale avec les valeurs singulières
- technique mathématique pour trouver des nouvelles dimensions qui expliquent le plus la variance



Exemple



Exemple



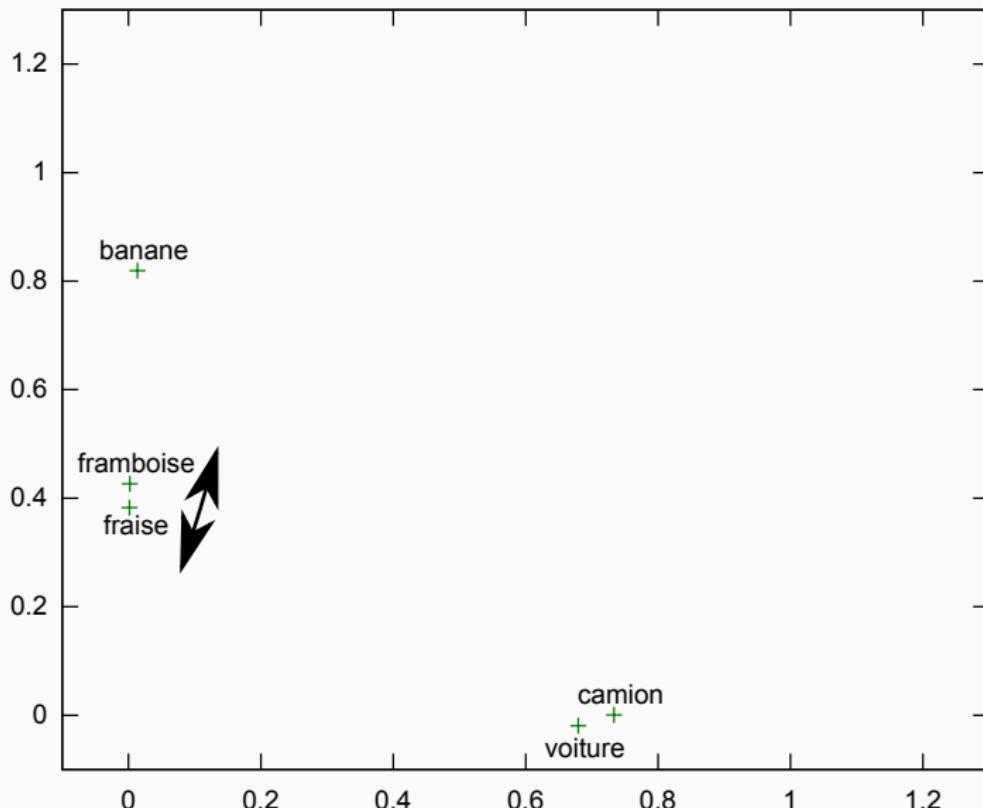
Dimension latente: exemple

words	context words	dependency relations
virus/NN	security/NN	amod#malicious/JJ
software/NN	Microsoft/NNP	nn-1#vulnerability/NN
security/NN	Internet/NNP	conj_and/cc#worm/NN
firewall/NN	Windows/NNP	nn-1#worm/NN
spam/NN	computer/NN	nn-1#flaw/NN
Security>NNP	network/NN	nn#antivirus/NN
vulnerability/NN	attack/NN	nn#IM/NNP
system/NN	software/NN	prep_of/in#worm/NN
Microsoft/NNP	protect/VB	nn#Trojan/NNP
computer/NN	protection/NN	conj_and/cc#virus/NN

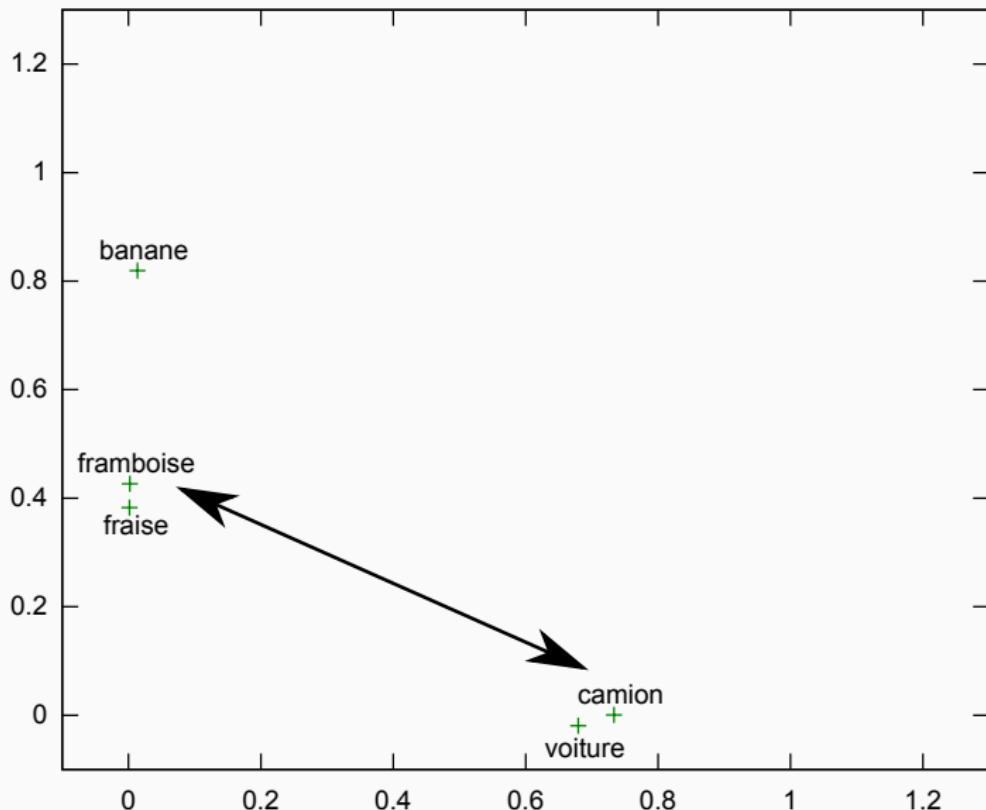
Dimension latente: exemple

words	context words	dependency relations
virus/NN	brain/NN	dobj-1#infect/VB
disease/NN	animal/NN	nsubjpass-1#infect/VB
bacterium/NN	disease/NN	rcmod#infect/VB
infection/NN	human/JJ	nsubj-1#infect/VB
human/NN	blood/NN	prep_with/in-1#infect/VB
rat/NN	cell/NN	conj_and/cc#rat/NN
cell/NN	cancer/NN	prep_of/in#virus/NN
animal/NN	skin/NN	amod#infected/JJ
mouse/NN	scientist/NN	prep_of/in#flu/NN
cancer/NN	drug/NN	nn#monkey/NN

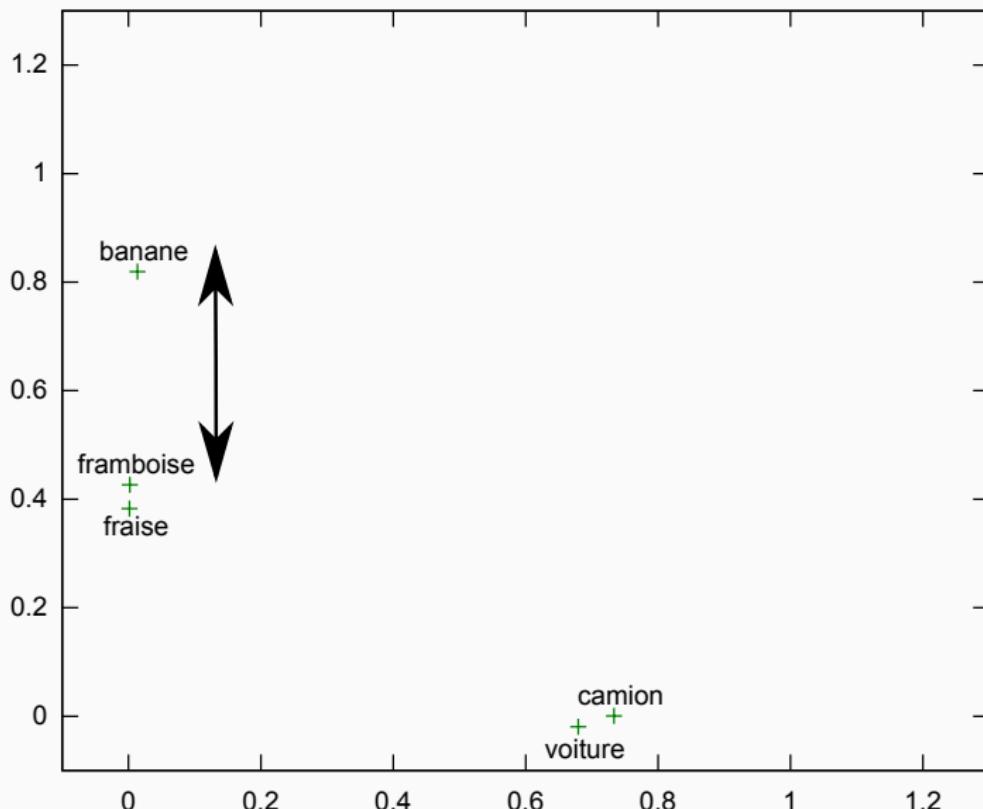
Mesure de similarité



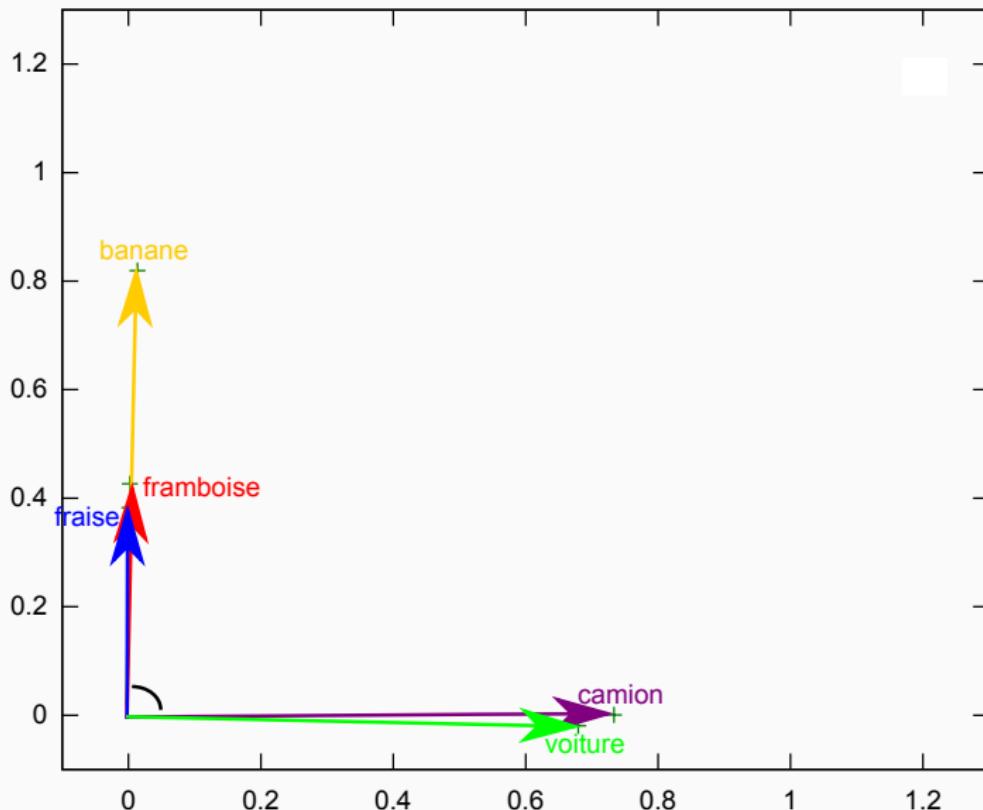
Mesure de similarité



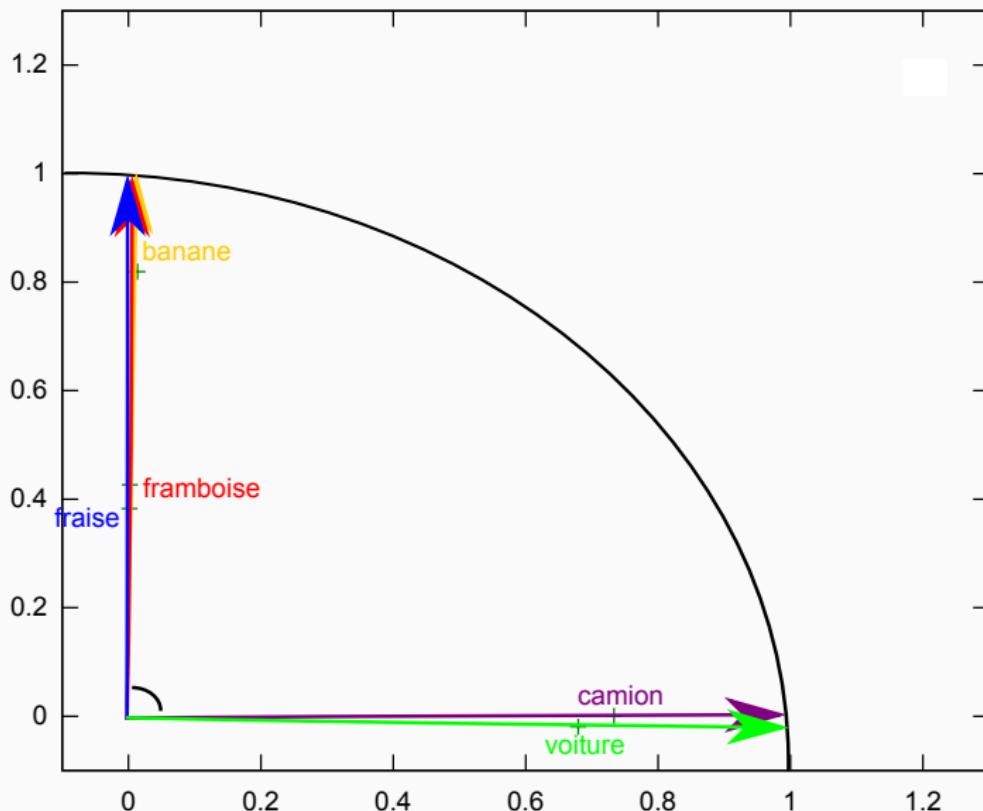
Mesure de similarité



Mesure de similarité



Mesure de similarité



TD : word embeddings

TD: Static Word embeddings and distributional similarity

Similarité lexicale

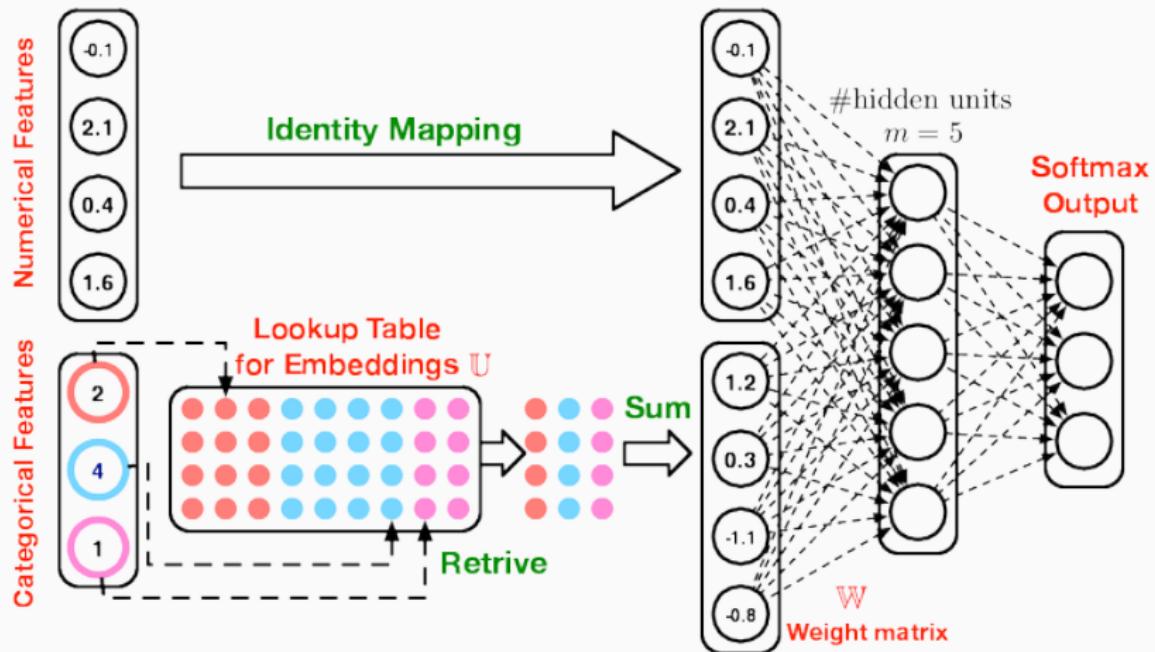
- Méthode statique → A partir de vecteurs sparses (matrice de co-occurrences)
- Méthode statique → A partir de vecteurs denses (Singular Value Decomposition-SVD)
- **Méthode prédictive → A partir de vecteurs denses (Word embeddings – plongement de mots)**

Un autre moyen de résoudre le problème d'éparpillement lexical

Plongements lexicaux / Word embeddings

- dans un modèle neuronal, les informations de catégories peuvent être codées par des représentations vectorielles intermédiaires apprises : des "embeddings" (plongement en français).
- pourrait-on trouver une tâche qui permettent d'avoir des représentations des mots intéressantes en général ?
- en s'inspirant de l'hypothèse distributionnelle : essayer de prédire l'apparition d'un mot dans un contexte

Illustration: encodage de catégories dans une représentation numérique



Plongements lexicaux / Word embeddings

- Chaque mot i est représenté par un vecteur plus petit, dense $\mathbf{v}_i \in \mathbb{R}^d$
- d typiquement $\approx 50\text{--}1000$
- Matrice de taille V (taille du vocabulaire) $\times d$ (taille de l'embedding)
- les mots sont ‘plongés’ (embedded) dans un espace réel de dimension (relativement) basse
- **les mots similaires auront des embeddings similaires**

Plongements lexicaux / Word embeddings

- Chaque mot i est représenté par un vecteur plus petit, dense $\mathbf{v}_i \in \mathbb{R}^d$
- d typiquement $\approx 50\text{--}1000$
- Matrice de taille V (taille du vocabulaire) $\times d$ (taille de l'embedding)
- les mots sont ‘plongés’ (embedded) dans un espace réel de dimension (relativement) basse
- **les mots similaires auront des embeddings similaires**

	d_1	d_2	d_3	...
apple	-2.34	-1.01	0.33	
pear	-2.28	-1.20	0.11	
car	-0.20	1.02	2.44	
...				

Plongements lexicaux / Word embeddings

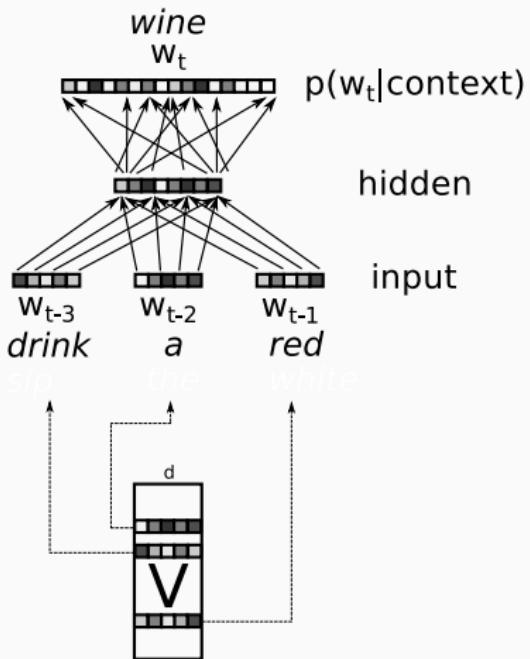
- Chaque mot i est représenté par un vecteur plus petit, dense $\mathbf{v}_i \in \mathbb{R}^d$
- d typiquement $\approx 50\text{--}1000$
- Matrice de taille V (taille du vocabulaire) $\times d$ (taille de l'embedding)
- les mots sont ‘plongés’ (embedded) dans un espace réel de dimension (relativement) basse
- **les mots similaires auront des embeddings similaires**

Neural word embeddings

- les plongements de mots dans un espace ne sont pas une idée récente
- mais le terme 'embedding' a été créé par la communauté (récente) des réseaux de neurones appliqués au langage naturel, avec de nouvelles méthodes pour les calculer
- Idée: on alloue un certain nombre de paramètres pour chaque mot, et le réseau apprend quelles sont les valeurs "utiles"
- Dirigé par la prédiction sur une tâche spécifique : par exemple prédire le mot suivant une suite de mot
- on retrouve l'idée des modèles de langue (cf début du cours)des

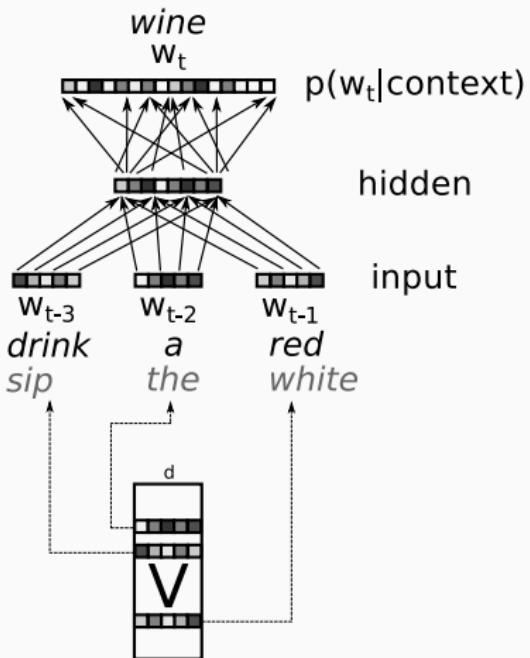
Embeddings obtenus par modèles de langue

- Prédire le mot suivant une séquence, à partir du mot précédent
- une seule couche non-linéaire + une couche de classification softmax
- les paramètres donnés aux embeddings doivent optimiser la probabilité du mot vraiment observé



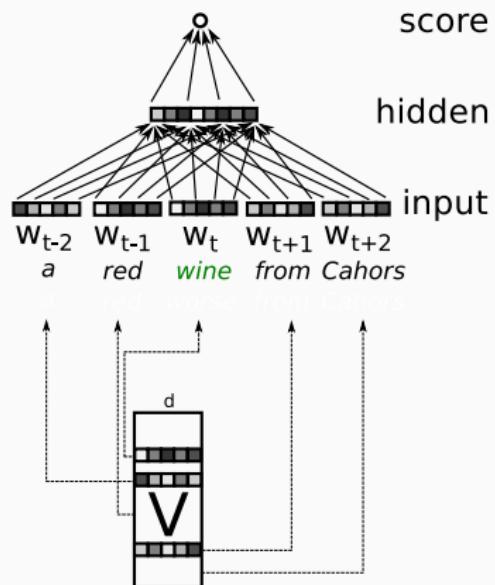
Embeddings obtenus par modèles de langue

- Prédire le mot suivant une séquence, à partir du mot précédent
- une seule couche non-linéaire + une couche de classification softmax
- les paramètres donnés aux embeddings doivent optimiser la probabilité du mot vraiment observé



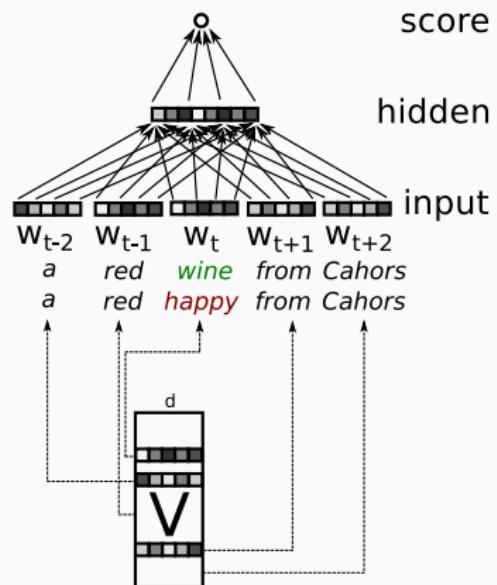
Embeddings obtenus par détection d'erreurs

- Prendre une phrase (ou une partie) correcte, et modifier un mot
- entraîner un réseau à trouver la version correcte contre la version corrompue



Embeddings obtenus par détection d'erreurs

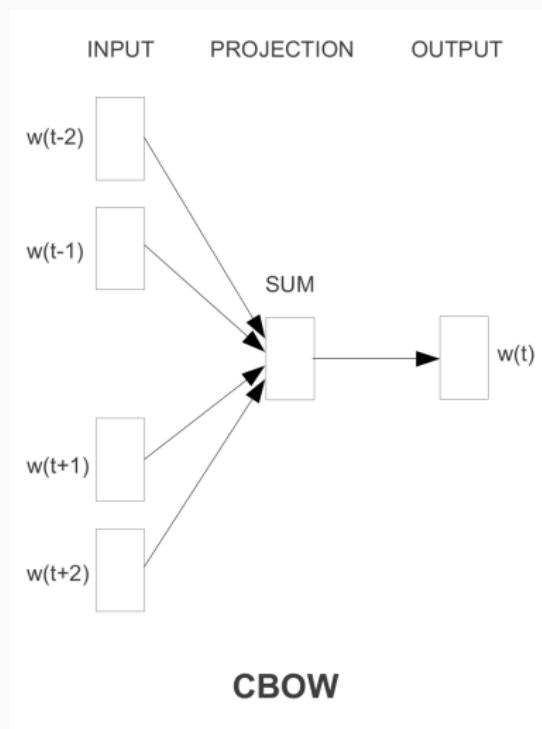
- Prendre une phrase (ou une partie) correcte, et modifier un mot
- entraîner un réseau à trouver la version correcte contre la version corrompue



- Approche par NN fonctionne bien mais assez lourde en pratique
- Une version légère avec moins de paramètres: **word2vec** (≈ 2013)
- seulement une couche: classifieur softmax
- deux versions
 - Continuous bag of words (CBOW): prédit un mot en fonction des mots juste autour
 - Skip-gram: prédit les mots d'un petit contexte autour d'un mot donné

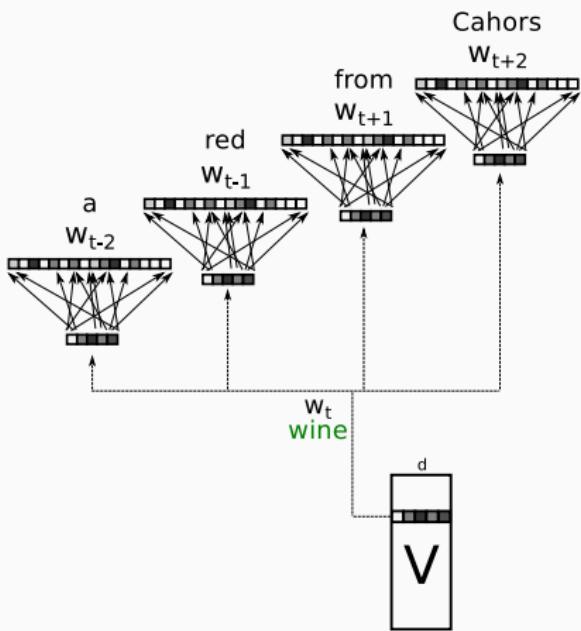
CBOW

- mot courant w_t prédict en fonction des mots du contexte
- on ajoute les embeddings des mots du contexte



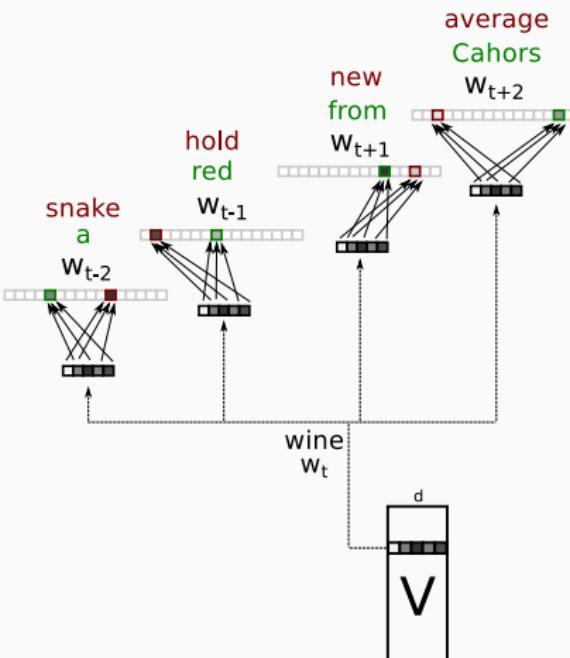
Skip-gram

- Dans une fenêtre fixe, chaque mot du contexte est prédit à partir du mot courant
- les paramètres de la couche de softmax sont partagés entre les prédictions



Negative sampling

- si on prend toutes les paires de mot possibles: calculs très lourds
- on choisit juste les paires "positives" et un échantillon de paires "négatives"



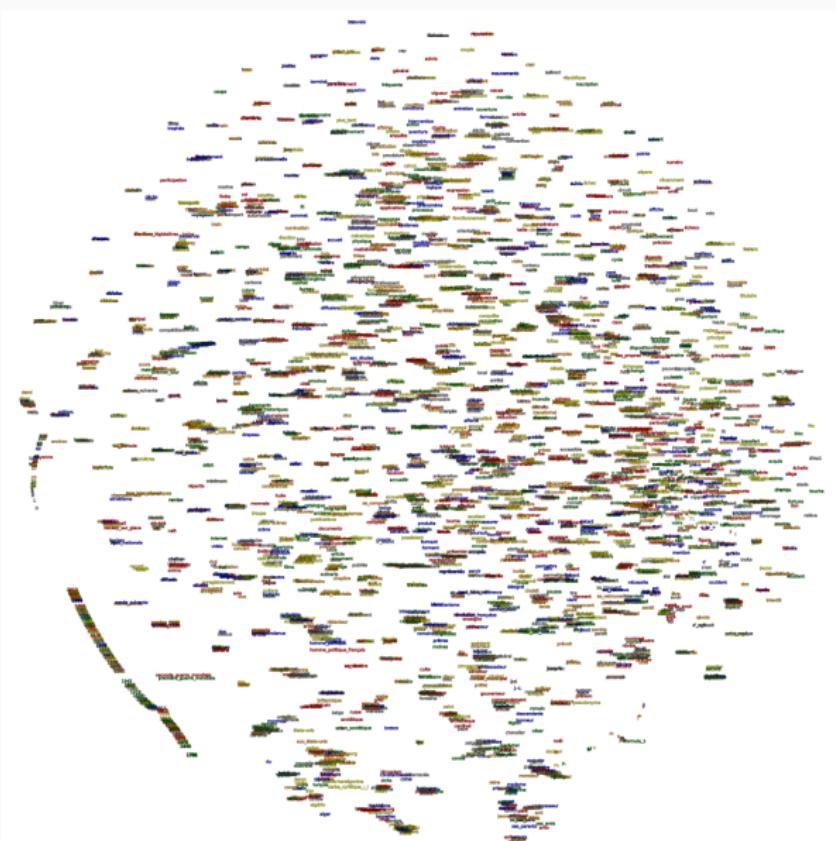
Similarité résultante

on considère que l'embedding de mot représente le mot dans l'espace "sémantique"

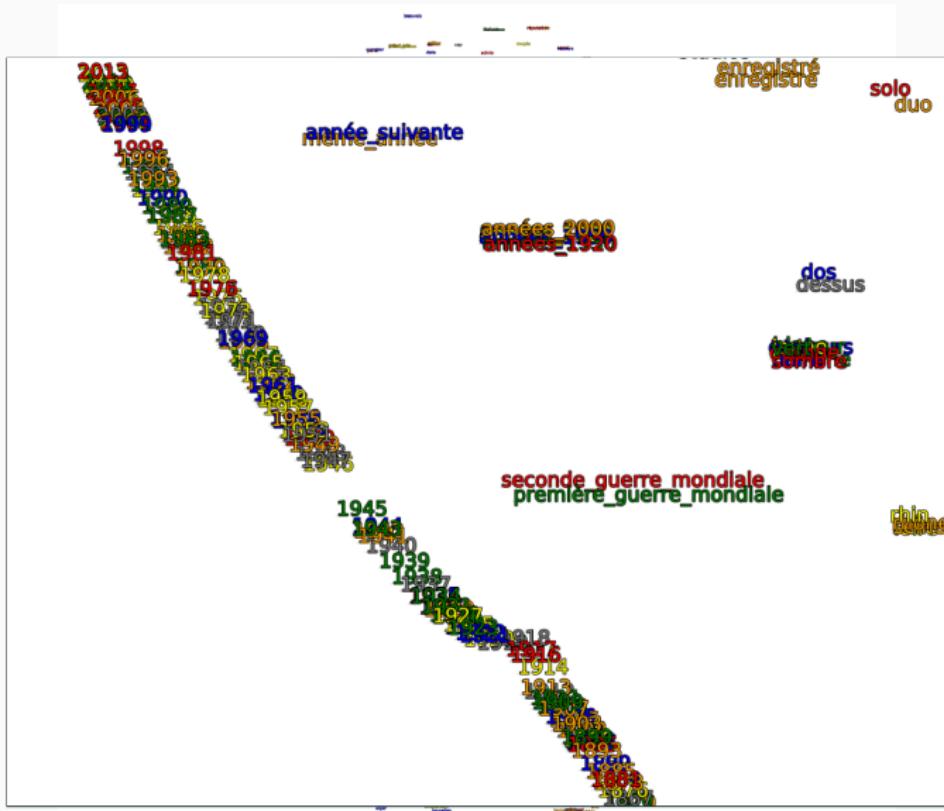
et on regarde les plus proches voisins:

france	jésus	rouge	peinture
grande-bretagne	christ	jaune	sculpture
belgique	jésus-christ	bleu	gravure
espagne	moïse	vert	dessin
angleterre	dieu	blanc	photographie
pologne	prophète	noir	peintures
bretagne	résurrection	bleue	toile
suède	mahomet	gris	céramique
italie	abraham	verte	poésie
roumanie	saints	blanche	peintre
tunisie	anges	noire	art

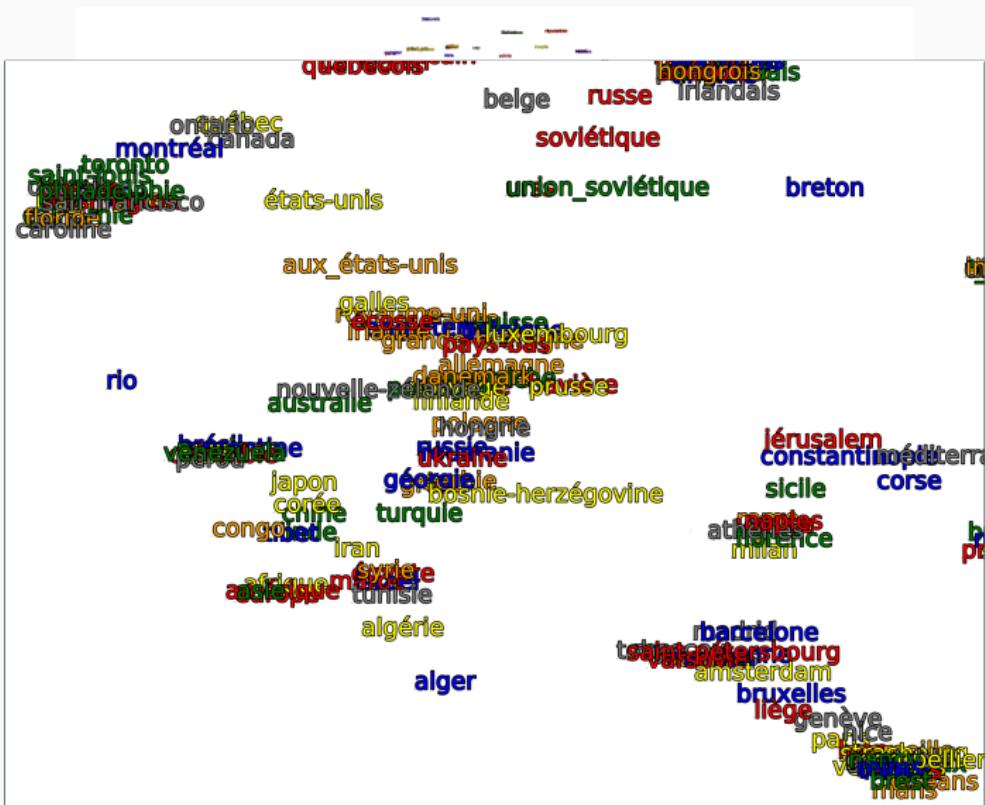
Word similarity



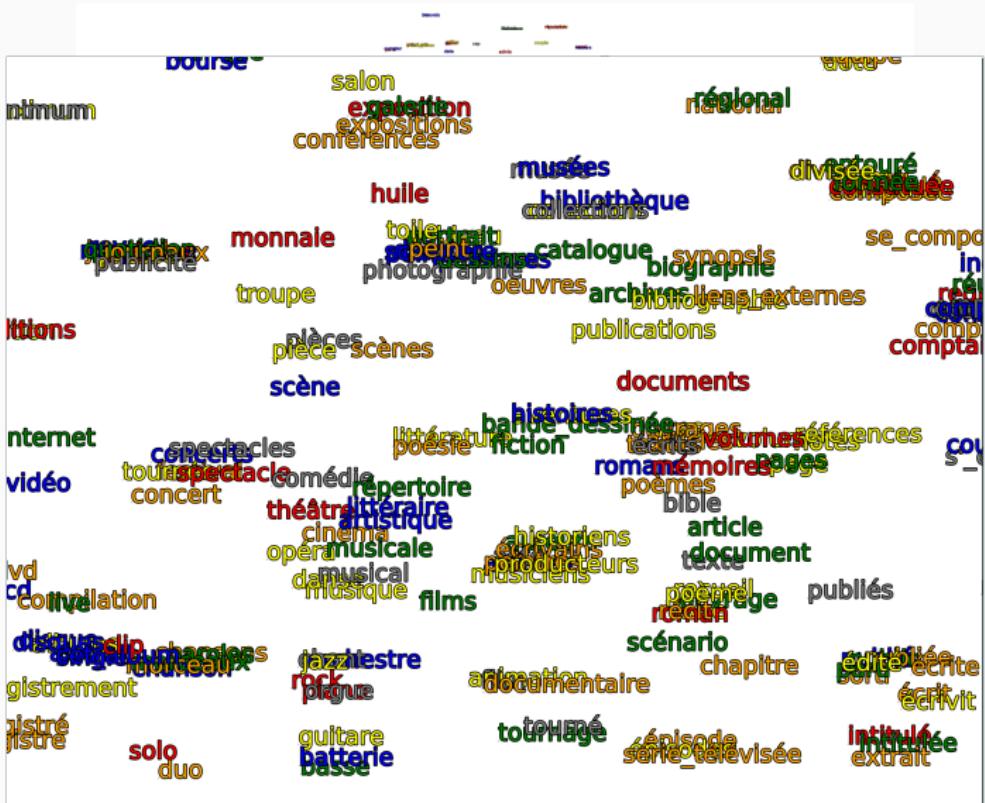
Word similarity



Word similarity



Word similarity



Word similarity



Demo

<http://projector.tensorflow.org/>

Calcul d'analogies

- Questions de la forme :

a est à b ce que c est à d

- on donne à un système *a*, *b*, et *c*; il faut trouver *d*.

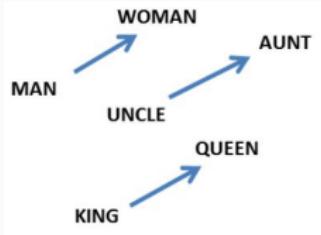
Exemples classiques:

apple is to apples as car is to ?

man is to woman as king is to ?

Calcul d'analogies

- **tache:** a is to b as c is to d
- **idée:** la direction de la relation doit rester la même dans l'espace vectoriel sémantique



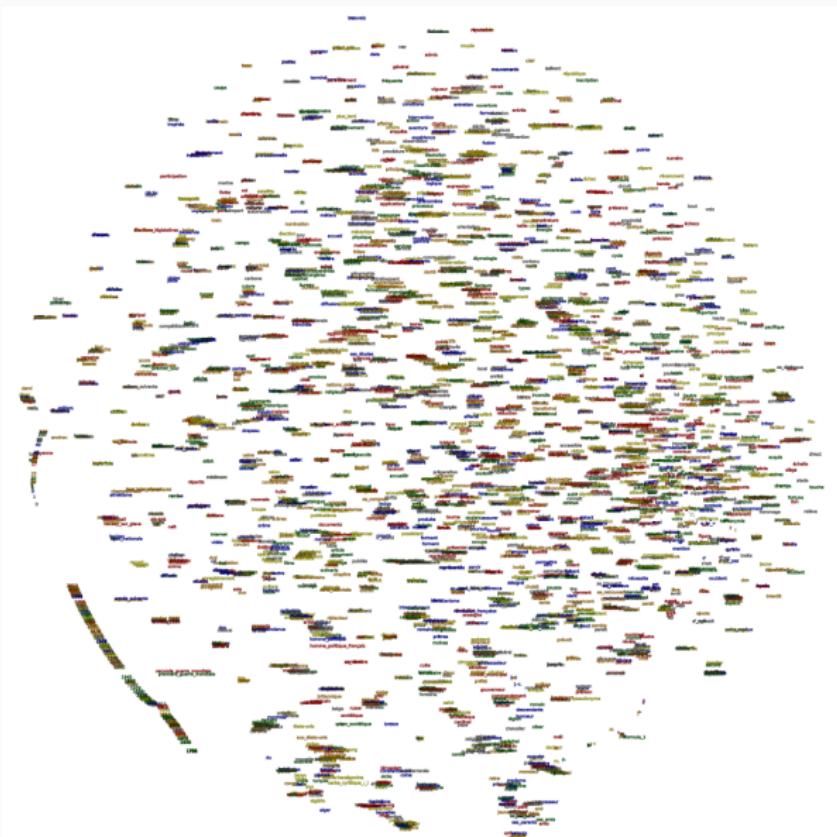
$$a - b \approx c - d$$

$$man - woman \approx king - queen$$

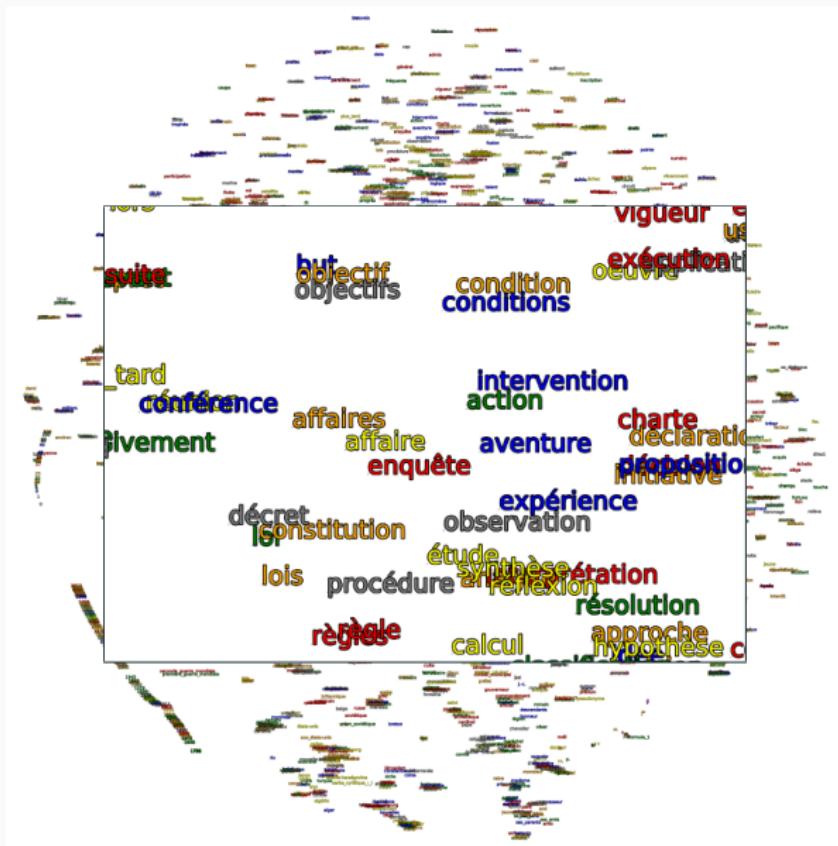
$$queen \approx king - man + woman$$

$$d_w = \underset{d'_w \in V}{\cos}(d', c - a + b)$$

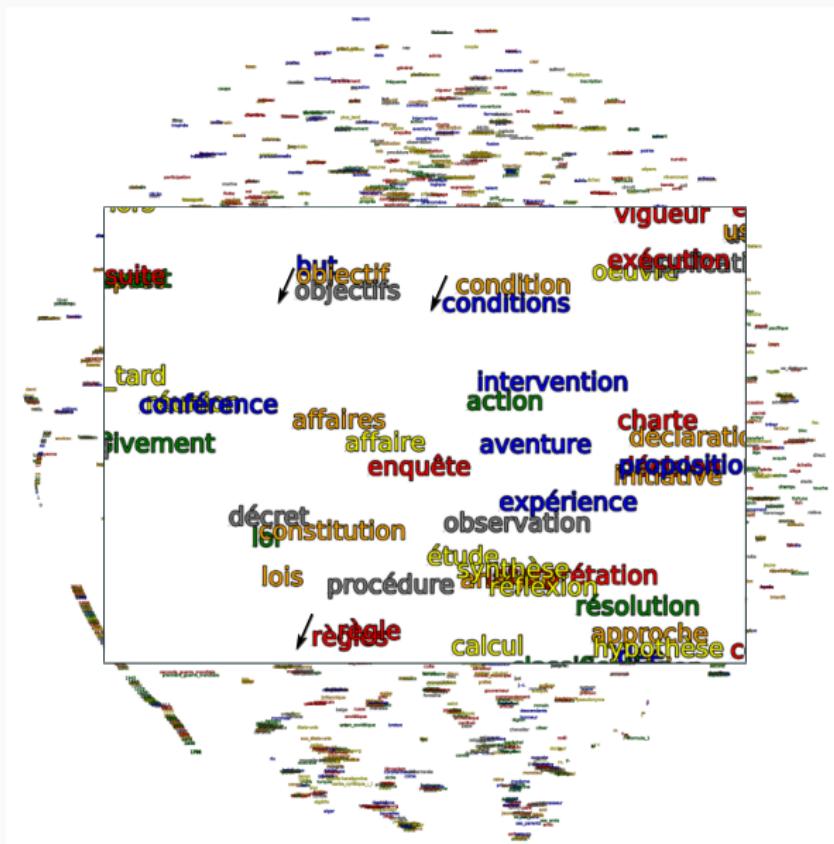
Calcul d'analogies



Calcul d'analogies



Calcul d'analogies



Calcul d'analogies

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Calcul d'analogies

<i>a</i>	is to	<i>b</i>	as	<i>c</i>	is to	<i>d</i>
homme		roi	:	femme		reine
autriche		vienne	:	allemande		berlin
écrivain		livre	:	poète		poème
france		nicolas_sarkozy	:	états-unis		bush
droite		ump	:	gauche		ps
droite		front_national	:	gauche		pcf

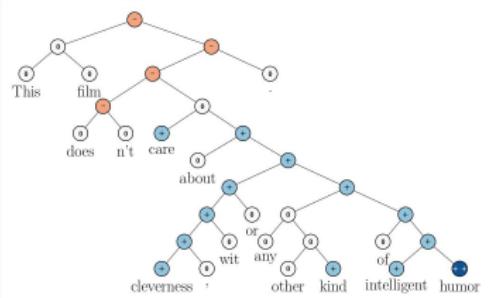
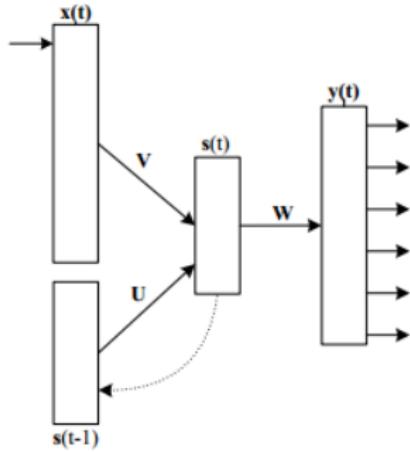
apprentissage "non-supervisé"/auto-supervisé: quels paramètres

- objectif de l'apprentissage
- choix du contexte (bag of words, syntaxe, corpus parallèles)
- taille du contexte (petit, phrase, paragraphe), dynamique, ...
- directionnalité
- Lemmatisation ?, normalization ?, stop words, ...
- unités: modèles de caractères ? sur des parties de mots ? (subwords)

En pratique

- utilisé en "préentraînement"
- non-supervisés: il faut juste du texte, plein de données disponibles facilement
- algorithmes rapides
- Modèle initialisé par le préentraînement, peut être ajusté ensuite selon les tâches
- problème : pas de prise en compte des ambiguïtés et du rôle du contexte d'une instance particulière
- composition au niveau d'une phrase ? (\rightarrow plus tard)

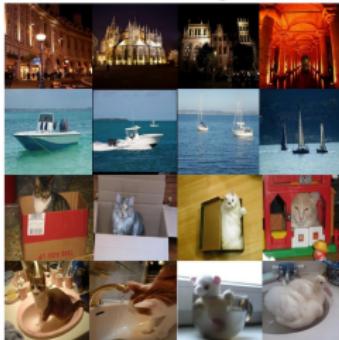
En pratique



Nearest Images



- day + night =



- flying + sailing =



- bowl + box =



- box + bowl =

Conclusion : prédition contre comptage

- **Comptage:** extraire des fréquences de cooccurrences d'un corpus
- **Prédiction:** induire une représentation indirectement des paramètres d'une tâche de prédiction de mots
- deux vues considérées comme différentes, la prédiction semblant plus efficace
- certains résultats indiquent que les deux sont équivalentes sous certaines conditions