

ÉVOLEX: à la croisée du TAL et de la psycholinguistique

Lydia-Mai Ho-Dac¹ Mélanie Jucla³

avec toute l'équipe du projet EVOLEX :

*Xavier de Boissezon⁴ Lola Danet⁴ Cécile Fabre¹ Jérôme Farinas² Bruno Gaume¹
Nabil Hatout¹ Patrice Péran⁴ Bénédicte Pierrejean¹ Julien Pinquier² Ludovic
Tanguy¹*

CEPEL, 20.03.2019



Plan

- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX
- 3 Données à disposition pour les traitements automatiques
- 4 Analyse des données
 - Analyse qualitative des données
 - Analyse qualitative des données
- 5 Analyse linguistique des paires de mots
- 6 Conclusion

Plan

- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX
- 3 Données à disposition pour les traitements automatiques
- 4 Analyse des données
- 5 Analyse linguistique des paires de mots
- 6 Conclusion

ÉVOLEX : méthode assistée par ordinateur pour évaluer l'accès au lexique et caractériser les relations lexicales

Un objectif triple

- Développer un outil pour évaluer auprès de populations avec ou sans trouble du langage l'accès lexical, i.e.
 - ① comment accédons-nous aux informations présentes dans notre lexique mental ?
 - ② comment associons-nous ces informations les unes avec les autres ?
- Proposer une caractérisation psycholinguistique de l'accès lexical (analyses qualitatives ET quantitatives)
- Automatiser la caractérisation linguistique des relations lexicales

⇒ **Lexique mental** ... ou comment décrire les relations entre les mots



Du point de vue de la psycholinguistique

Questions de base

- Qu'est-ce que deux mots "proches" dans le lexique mental d'un sujet ?
- Pourquoi deux mots sont "proches" dans le lexique mental d'un sujet ?
- Quels sont les mots les plus proches (les plus proches *voisins*) d'un mot donné dans le lexique mental d'un sujet ?
- Y a-t-il une incidence de l'âge (Burke and Peters, 1986), du genre, du statut social, de certains troubles du langage (Péran et al., 2004) sur le traitement cognitif des relations lexicales ?

Du point de vue de la psycholinguistique

Besoin d'un jeu d'évaluation où toutes les relations entre mots sont qualifiées

- Deux mots proches chez un sujet sont ils proches dans la langue de façon générale ?
- Quelle est la probabilité pour un mot d'être relié à un autre ?

Deux problèmes à résoudre

- 1 On ne pourra jamais avoir un jeu d'évaluation exhaustif couvrant tous les mots connus par un sujet et tous de la langue
- 2 Difficultés à minimiser la part de subjectivité dans la qualification des relations

Du point de vue du TAL

Des méthodes inductives basées sur l'hypothèse distributionnelle

Des mots apparaissant dans contextes similaires ont des sens proches (Harris, 1954)

“You shall know a word by the company it keep”, (Firth, 1957)

Principe de l'approche distributionnelle

(6.1) Ongchoi is delicious sauteed with garlic.

(6.2) Ongchoi is superb over rice.

(6.3) ...ongchoi leaves with salty sauces...

(6.4) ...spinach sauteed with garlic over rice...

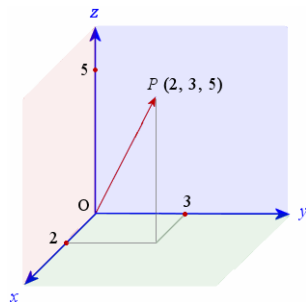
(6.5) ...chard stems and leaves are delicious...

(6.6) ...collard greens and other salty leafy greens

Speech and Language Processing, Jurafsky & Martin, 3rd ed. draft

Modèles distributionnels

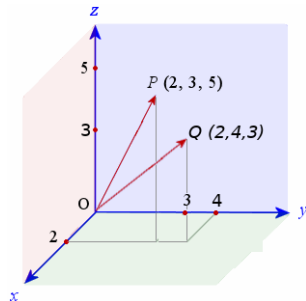
Représentation du sens des mots sous forme de vecteurs compilant des informations liées aux contextes d'apparition des mots



| | dimension x | dimension y | dimension z |
|-------|-------------|-------------|-------------|
| mot P | 2 | 3 | 5 |

Modèles distributionnels

Représentation du sens des mots sous forme de vecteurs compilant des informations liées aux contextes d'apparition des mots



| | | leaves (x) | salty (y) | garlic (z) |
|----------|---------|------------|-----------|------------|
| <i>P</i> | ongchoi | 2 | 3 | 5 |
| <i>Q</i> | spinach | 2 | 4 | 3 |

La proximité entre deux mots est calculée par une "mesure de similarité" basée sur la distance entre les deux vecteurs (cosinus de l'angle = 1 si vecteurs similaires)

Modèles distributionnels

Comment obtenir ces vecteurs ?

Modèles explicites fondés sur le décompte des contextes

- 1 Extraction des contextes
- 2 Décompte et pondération
- 3 (éventuellement) Réduction de dimensions
- 4 Calcul de similarité

Modèles prédictifs (*Word embeddings*)

- Apprentissage automatique de la représentation vectorielle : le système apprend à assigner des vecteurs similaires à des mots similaires (utilisation d'un "réseau de neurones")
- Les contextes qui ont rapprochés des mots ne sont plus explicites

Du point de vue du TAL

Problèmes liés à l'utilisation des modèles distributionnels

- Caractérisation grossière qui confond synonymes et associations lexicales (*canapé - sofa* vs. *canapé - sieste*)
- Manque de jeux d'évaluation pour caractériser la nature des relations mises en évidence (Hill et al. 2005, Baroni and Lenci 2015)

ÉVOLEX, une opportunité pour ...

- Utiliser des modèles de sémantique distributionnelle pour étudier les relations lexicales et comprendre l'accès au lexique
- Utiliser un jeu d'évaluation obtenu lors d'une expérience psycholinguistique qui permet d'aborder la "compétence lexicale" en bénéficiant du profil du répondant

Plan

- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX**
- 3 Données à disposition pour les traitements automatiques
- 4 Analyse des données
- 5 Analyse linguistique des paires de mots
- 6 Conclusion

Mesurer l'accès au lexique par des exercices d'*Évocation verbale*

Des tests variés et complémentaires permettant de récolter des mots et des temps d'exécution

Dénomination Dire un mot dénommant une image

Génération Proposer un mot "en lien" avec un autre mot

Fluences verbales Dire le plus de mots dans un temps imparti

- Deux types de fluences, deux types de tests :
 - 1 Formel (ex : mots commençant par "V", "R", etc.)
 - 2 Sémantique (ex : noms de fruits, synonymes du mot X, idées associées, etc.)

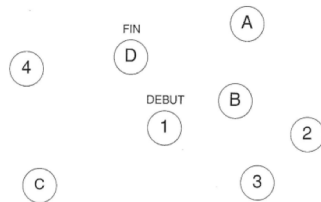
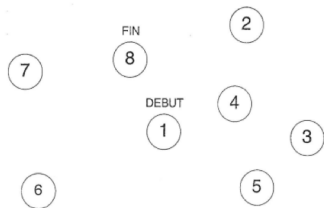
Le protocole ÉVOLEX

5 tests de contrôle et de calibrage

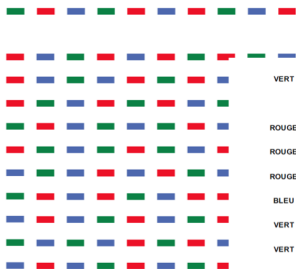
- 1 **Lecture** d'un texte à haute voix : mesure des capacités langagières et calibrage de la reco. vocale
- 2 **MoCA** (Montréal Cognitive Assessment, plusieurs exercices) : mesure des capacités cognitives générales
- 3 **Empans mnésique** (répéter des séries de chiffres) : mesure des capacités de mémoire à court terme
- 4 **TMT** (Trail Making Test, relier des points) : mesure la flexibilité mentale
- 5 **Stroop** (e.g. ●, bleu, **bleu**) : évalue les capacités d'inhibition de processus automatiques



TMT (Trail Making Test)



Stroop (John Stroop, 1897 – 1973)



VERT BLEU ROUGE ROUGE VERT BLEU VERT ROUGE VERT BLEU

ROUGE BLEU ROUGE VERT BLEU ROUGE VE

ROUGE VERT BLEU ROUGE VERT ROUGE BL

ROUGE BLEU ROUGE VERT BLEU ROUGE VE

BLEU VERT ROUGE ROUGE BLEU VERT RO

VERT ROUGE BLEU VERT ROUGE BLEU RO

VERT BLEU ROUGE BLEU VERT ROUGE VE

ROUGE BLEU BLEU ROUGE VERT BLEU VE

VERT ROUGE VERT BLEU ROUGE VERT RO

ROUGE VERT BLEU ROUGE VERT ROUGE BL

BLEU VERT ROUGE VERT ROUGE BLEU RO

ROUGE VERT BLEU ROUGE ROUGE VERT BLEU VERT ROUGE BLEU

ROUGE BLEU VERT ROUGE BLEU ROUGE VERT VERT BLEU BLEU

ROUGE BLEU VERT BLEU ROUGE VERT ROUGE BLEU VERT ROUGE

ROUGE VERT BLEU ROUGE VERT ROUGE BLEU VERT ROUGE VERT

VERT ROUGE BLEU VERT BLEU VERT ROUGE BLEU BLEU ROUGE

ROUGE VERT BLEU VERT ROUGE VERT BLEU ROUGE BLEU VERT

BLEU VERT BLEU ROUGE BLEU ROUGE VERT BLEU ROUGE VERT

VERT ROUGE BLEU ROUGE VERT BLEU ROUGE ROUGE VERT BLEU

BLEU ROUGE BLEU VERT ROUGE BLEU VERT ROUGE BLEU ROUGE

VERT BLEU VERT BLEU ROUGE VERT ROUGE BLEU VERT ROUGE

BLEU ROUGE BLEU VERT ROUGE BLEU VERT ROUGE BLEU ROUGE

Le protocole ÉVOLEX

3 test dédiés ÉVOLEX

1 **Fluences verbales** (2 min.)

4 séries : v.*, r.* (accès formel), animaux, **fruits** (accès sémantique)

2 **Dénomination** (une image affichée → un mot, 60 stimuli)



→ ?



→ ?



→ ? ...

3 **Génération** (un mot entendu (synthèse vocale) → un mot "proche", 60 stimuli, ordre figé)

Vous allez entendre des noms communs de la langue Française, vous devrez alors, sans réfléchir, prononcer le premier mot qui vous vient à l'esprit ayant rapport avec celui que vous avez entendu. Par exemple, si vous entendez - TABLE -, alors vous pourrez répondre - CHAISE -.

Nous allons procéder à un test, si vous entendez - ARBRE -, que répondez-vous ?

audio → ? **audio** → ? **audio** → ? ...

Plan

- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX
- 3 Données à disposition pour les traitements automatiques**
- 4 Analyse des données
- 5 Analyse linguistique des paires de mots
- 6 Conclusion



Paires de mots issues de la tâche de Génération

| | A | AX | AY | AZ | BA | BB | BC | BD | |
|----|------------|------------|----------------|--------------|----------------|------------|----------------|----------|-------|
| 1 | Stimulus | réponse | latence chrono | réponse | latence chrono | réponse | latence chrono | réponse | laten |
| 39 | Ortie | feuille | 2,39 | pissenlit | 2,12 | plante | 2,05 | pique | |
| 40 | Osier | plante | 2,87 | panier | 1,41 | panier | 1,89 | panier | |
| 41 | Papillon | fleur | 1,54 | insecte | 1,77 | X | X | de nuit | |
| 42 | Parc | chien | 1,48 | jardin | 1,46 | jeux | 1,25 | cygne | |
| 43 | Perroquet | animal | 1,52 | oiseau | 2,85 | oiseau | 1,68 | bavard | |
| 44 | Pharmacie | médicament | 1,39 | médicament | 1,96 | médicament | 1,74 | sirop | |
| 45 | Pinceau | peinture | 1,28 | tableau | 1,64 | peinture | 3,28 | peintre | |
| 46 | Poil | barbe | 1,92 | rasoir | 3,9 | épilation | 1,86 | œuf | |
| 47 | Pôle | nord | 2,73 | X | X | nord | 1,39 | nord | |
| 48 | Rail | X | X | train | 1,72 | Faudel | 2,18 | train | |
| 49 | Raisin | figue | 1,51 | grappe | 1,77 | sec | 1,19 | vin | |
| 50 | Sac | réceptient | 2,06 | portefeuille | 3,51 | à main | 3,9 | à main | |
| 51 | Sapin | noël | 1,05 | noël | 1,31 | noël | 0,97 | noël | |
| 52 | Serpent | reptile | 1,48 | sonnette | 2,52 | langue | 1,26 | langue | |
| 53 | Siège | fauteuil | 1,64 | trône | 2,5 | fauteuil | 1,48 | fauteuil | |
| 54 | Sœur | frère | 1,36 | frère | 1,54 | frère | 1,1 | frère | |
| 55 | Spaghettis | bolognaise | 1,75 | bolognaise | 2,19 | pâtes | 1,17 | pâtes | |
| 56 | Terrier | lapin | 0,96 | renard | 1,59 | lapin | 1,17 | lapin | |
| 57 | Trou | vide | 5,48 | puits | 2,85 | puits | 1,17 | puits | |
| 58 | Truc | machin | 1,08 | truc | 1,08 | truc | 1,08 | truc | |
| 59 | Varicelle | maladie | 1,77 | maladie | 1,77 | maladie | 1,77 | maladie | |

Jeu de données issu de la tâche de Génération

Résultats d'une première récolte

- 30 passations, 60 stimuli, 1800 paires de mots
- des paires à corriger, normaliser et filtrer :
 - deux stimuli ambigus non pris en compte (bec / mec ; poil / poele)
 - sélection des paires des NOMS (communs et propres) simples
 - correction orthographique
 - normalisation : flexion et casse
- 1544 paires exploitables et 559 paires différentes

Plan

- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX
- 3 Données à disposition pour les traitements automatiques
- 4 Analyse des données**
- 5 Analyse linguistique des paires de mots
- 6 Conclusion



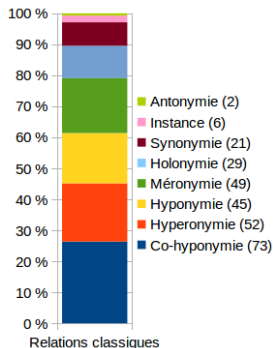
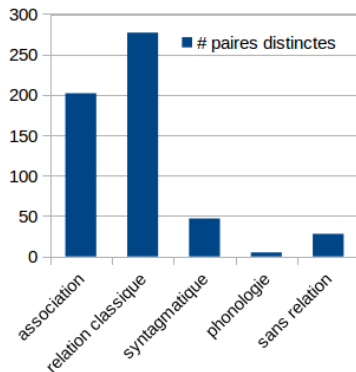
Double annotation manuelle puis adjudication

Guide d'annotation

| Relation | Exemple |
|---------------------------|------------------------------|
| antonyme | <i>aube / crépuscule</i> |
| association | <i>balançoire / enfant</i> |
| co-hyponyme | <i>balançoire / toboggan</i> |
| holonyme | <i>doigt / main</i> |
| hyperonyme | <i>balançoire / jeu</i> |
| hyponyme | <i>animal / chat</i> |
| instance | <i>magicien / Merlin</i> |
| méronyme | <i>balançoire / corde</i> |
| rel. phonologique | <i>chapiteau / château</i> |
| synonyme | <i>canapé / sofa</i> |
| rel. syntagmatique | <i>fleur / peau</i> |
| sans relation | <i>perroquet / placard</i> |



Résultats



⇒ 50% de **relations "classiques"**

- des relations "autres" : association d'idées (36%), syntagmes (9%), pas de relation (5%), lien phonologique (1%)



Quatre mesures de similarités sémantique

2 SIMILARITÉS issue de corpus

- **Corpus1st** : Plus deux mots apparaissent fréquemment ensemble en corpus, plus ces deux mots sont similaires (= collocations)
- **Corpus2nd** : Plus deux mots partagent les mêmes collocations, plus ces deux mots sont similaires (= substituabilité)
- corpus = FrWac (pages Web *.fr*, 2 milliards de mots)

2 SIMILARITÉS issues de dictionnaires

- **Dict1st** : Si un mot apparaît dans la définition d'un autre mot, ces deux mots sont similaires
- **Dict2nd** : Plus les définitions de deux mots partagent de mots, plus ces deux mots sont similaires
- dict. = TLF – Trésor de la Langue Française

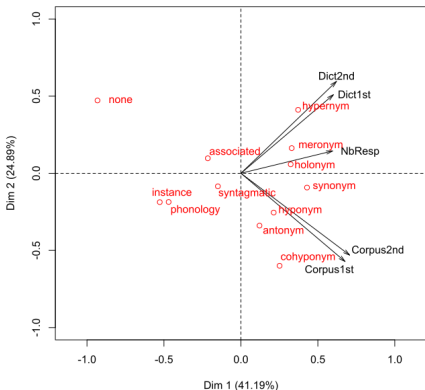


| identifiant | stimulus corrigé | ↑ | SIMILARITE | ASSOC | SYNTAGMA | PHONO/MORPH | ? | POS | polysémie | Njdm_r associated edge [0.34426541] | Njdm_r associated prox [0.46134421] | Njdm_r_syn edge [0.06630699] | Njdm_r_syn prox [0.16639659] | FRMACDEP | FRMACCOOC |
|-------------|---------------------|---|------------|-------|----------|-------------|---|-----|-----------|--|--|---------------------------------|---------------------------------|----------------|----------------|
| SRP-1 | abricot arbre | | holonyme | x | | | | | | 90 : 0.71084663 | 0.00481111 : 1.27760392 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.307984140227 | 0.237063008347 |
| SRP-2 | abricot caramel | | holonyme | x | | | | | | 0 : 0.00000000 | 0.00020092 : 0.05335488 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.716536036546 | 0.58595125267 |
| SRP-3 | abricot confiture | | holonyme | | nden | | | | | 110 : 0.86881255 | 0.00297627 : 0.79035695 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.699835570957 | 0.608131859695 |
| SRP-4 | abricot fruit | | hyperonyme | | | | | | | 407 : 3.21466044 | 0.02385163 : 6.33368805 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.585114703217 | 0.498159446072 |
| SRP-5 | abricot kiwi | | cohyponym | | | | | | | 0 : 0.00000000 | 0.00039997 : 0.10621317 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.851678422504 | 0.601300971603 |
| SRP-6 | abricot noyau | | mero | | | | | | | 193 : 1.52437111 | 0.00698082 : 1.85377657 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.355403225095 | 0.299773918085 |
| SRP-7 | abricot orange | | cohyponym | | | | | | | 110 : 0.86881255 | 0.00566499 : 1.50435418 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.625514479621 | 0.501379154017 |
| SRP-8 | abricot pêche | | cohyponym | | | | | | | 140 : 1.10576143 | 0.01580459 : 4.19695374 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.388458939545 | 0.304255311713 |
| SRP-8 | abricot pêche | | cohyponym | | | | | | | 140 : 1.10576143 | 0.01580459 : 4.19695374 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.388458939545 | 0.304255311713 |
| SRP-9 | ampoule ampère | | | x | | | | x | | 0 : 0.00000000 | 0.00014854 : 0.03944522 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.469875706529 | 0.282082072716 |
| SRP-16 | ampoule baïonnette | | | | | | | | | | | | | | |
| SRP-10 | ampoule brûlure | | | x | | | | x | | 20 : 0.15796592 | 0.00088016 : 0.23372899 | 55.0 : 5.72600053 | 0.01274641 : 2.91072580 | 0.423211278474 | 0.311680853138 |
| SRP-11 | ampoule électricité | | | x | | | | x | | 139 : 1.09786313 | 0.00890664 : 2.36518353 | 0 : 0.00000000 | 0.00024113 : 0.05506361 | 0.333496831652 | 0.314671688215 |
| SRP-12 | ampoule filament | | mero | | | | | x | | 123 : 0.97149040 | 0.00376784 : 1.00056061 | 0 : 0.00000000 | 0.00009200 : 0.02100880 | 0.472884948322 | 0.376733679556 |
| SRP-13 | ampoule lampe | | holonyme | | | | | | | 193 : 1.52437111 | 0.01416701 : 3.76208972 | 81.0 : 8.43283715 | 0.01651294 : 3.77083748 | 0.705806520646 | 0.701339379011 |
| SRP-14 | ampoule lumière | | | x | | | | x | | 400 : 3.15931837 | 0.01795581 : 4.76821632 | 0 : 0.00000000 | 0.00414214 : 0.94588467 | 0.366875715353 | 0.284478752000 |
| SRP-15 | ampoule pied | | holonyme | | | | | x | | 139 : 1.09786313 | 0.00397904 : 1.05664537 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.251624256718 | 0.234770000000 |
| SRP-17 | animal bestial | | | | | | | adj | x | | -r_abs | -r_abs | -r_abs | 0.0 | 0.0 |
| SRP-18 | animal bestiole | | syn | | | | | | | 54 : 0.42650798 | 0.00005179 : 0.01375298 | 71.5 : 7.44380070 | 0.02141655 : 4.89060878 | 0.608287387978 | 0.390000000000 |
| SRP-19 | animal chat | | hyponym | | | | | | | 612 : 4.83375710 | 0.00042390 : 1.11256785 | 0 : 0.00000000 | 0.00001859 : 0.00424515 | 0.569137592677 | 0.390000000000 |
| SRP-20 | animal chien | | hyponym | | | | | | | 672 : 5.30765486 | 0.00042859 : 1.11381329 | 0 : 0.00000000 | 0.00010261 : 0.02433166 | 0.80412835226 | 0.390000000000 |
| SRP-21 | animal compagnie | | | | nden | | | | | 55 : 0.43440628 | 0.00003196 : 0.00848707 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.190661582000 | 0.390000000000 |
| SRP-22 | animal fouine | | hyponym | | | | | | | 52 : 0.41071139 | 0.00007716 : 0.02049006 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.515076600000 | 0.390000000000 |
| SRP-23 | animal humain | | hyponym | | | | | | | 0 : 0.00000000 | 0.00014104 : 0.03745257 | 0 : 0.00000000 | 0.00039320 : 0.08978978 | 0.451000000000 | 0.390000000000 |
| SRP-24 | animal lion | | hyponym | | | | | | | 361 : 2.85129483 | 0.00025391 : 0.06742652 | 0 : 0.00000000 | 0.00003101 : 0.00708134 | 0.451000000000 | 0.390000000000 |
| SRP-25 | animal mignon | | | | | | | adj | x | 0 : 0.00000000 | 0.00001578 : 0.00419042 | 0 : 0.00000000 | 0.00012050 : 0.02751888 | 0.451000000000 | 0.390000000000 |
| SRP-26 | animal mouton | | hyponym | | | | | | | 23 : 0.18166081 | 0.00011265 : 0.02991453 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.451000000000 | 0.390000000000 |
| SRP-27 | animal oiseau | | hyponym | | | | | | | 61 : 0.48179605 | 0.00092663 : 0.24606923 | 0 : 0.00000000 | 0.00037876 : 0.00900000 | 0.451000000000 | 0.390000000000 |
| SRP-28 | animal végétal | | cohyponym | | | | | | | 54 : 0.42650798 | 0.00006368 : 0.01691040 | 0 : 0.00000000 | 0.00010000 : 0.00000000 | 0.451000000000 | 0.390000000000 |
| SRP-29 | aube aurore | | syn | | | | | x | | 100 : 0.78982959 | 0.00552794 : 1.46796016 | 78.0 : 8.12050985 | 0.00000000 : 0.00000000 | 0.451000000000 | 0.390000000000 |
| SRP-30 | aube communion | | | x | | | | x | | 0 : 0.00000000 | 0.00032927 : 0.08743858 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.451000000000 | 0.390000000000 |
| SRP-31 | aube crépuscule | | antonyme | | | | | x | | 40 : 0.31593184 | 0.00293856 : 0.79034295 | 46.0 : 4.63000000 | 0.00000000 : 0.00000000 | 0.451000000000 | 0.390000000000 |
| SRP-40 | aube Hiroshima | | | | | | | x | | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.00000000 : 0.00000000 | 0.00000000 : 0.00000000 | 0.451000000000 | 0.390000000000 |
| SRP-32 | aube jour | | holonyme | | | | | | | 97 : 0.71084663 | 0.00481111 : 1.27760392 | 0 : 0.00000000 | 0.00000000 : 0.00000000 | 0.307984140227 | 0.237063008347 |

Plan

- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX
- 3 Données à disposition pour les traitements automatiques
- 4 Analyse des données
- 5 Analyse linguistique des paires de mots**
- 6 Conclusion

Corrélation entre les caractérisations (ACP)



→ Des mesures positivement corrélées aux **relations classiques**

→ Des mesures négativement corrélées dans les cas de

- sens éloigné (sans relation identifiée, phonologie : *chou – caillou*)
- d'instance (*magicien – Merlin*)

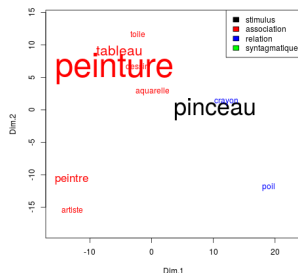
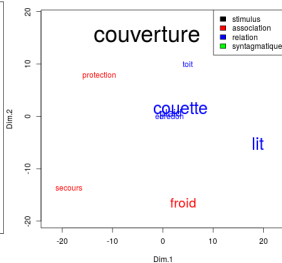
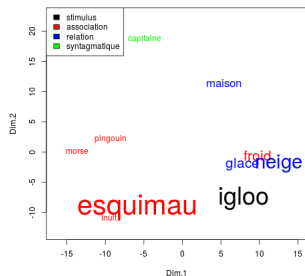
→ Des aspects différents captés par les différentes méthodes

- **dictionnaire** – hyperonymie (*entrecôte – viande*)
- **corpus** – cohyponymie (*balançoire – toboggan*)

Au delà de la caractérisation des paires de mots...

Identifier des groupes (catégories ? profils ?) de réponse

Réponses regroupés dans l'espace vectoriel (corpus – 2nd ordre)



Plan

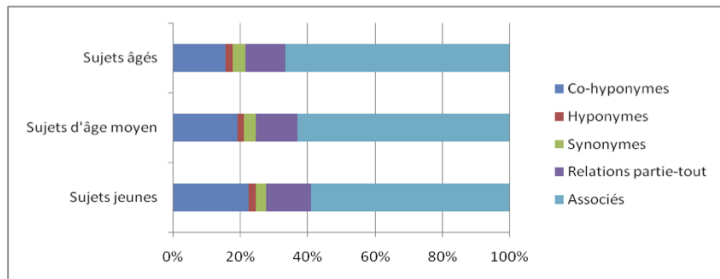
- 1 Objectifs du projet ÉVOLEX
- 2 Le protocole ÉVOLEX
- 3 Données à disposition pour les traitements automatiques
- 4 Analyse des données
- 5 Analyse linguistique des paires de mots
- 6 Conclusion**

En résumé

- 1 Recueil de données auprès de locuteurs "sains"
 - Un protocole en 8 tests permettant de récolter des contextes variés d'évocation verbale
 - Reconnaissance vocale pour enregistrer les réponses et les temps de réaction : logiciel EvoLex
- 2 Analyse des données récoltées
 - Qualifier l'accès
 - "scorer" ces accès
 - (Améliorer la reconnaissance vocale)
- 3 Utilisation des données analysées pour "diagnostiquer" la fluence de locuteurs (sains ou pas)
- 4 Eprouver nos modèles TAL de représentation de la structure lexicale (petits mondes, sem. distrib.)

Et après (parce que c'est cool de faire des trucs ensemble !)

- Reproduire la méthode sur un nouveau jeu de données (100 sujets, 780 paires de mots)
- Croiser les mesures de similarité et les temps de réponse
- Intégrer aux résultats les caractéristiques des répondants





B. Gaume, K. Duvignau, E. Navarro, Y. Desalle, H. Cheung, S.-K. Hsieh, P. Magistry, and L. Prevot.

Skillex : a graph-based lexical score for measuring the semantic efficiency of used verbs by human subjects describing actions.

[Revue TAL](#), 55(3), 2016.



B. Gaume, L. Tanguy, C. Fabre, L.-M. Ho-Dac, B. Pierrejean, N. Hathout, J. Farinas, J. Pinquier, L. Danet, P. Péran, X. De Boissezon, and M. Jucla.

Automatic analysis of word association data from the evolex psycholinguistic tasks using computational lexical semantic similarity measures.

In [13th International Workshop on Natural Language Processing and Cognitive Science \(NLPCS\)](#), Krakow, Poland, 2018.



P. Péran, J.-F. Démonet, C. Pernet, and D. Cardebat.

Verb and noun generation tasks in huntington's disease.

[Movement disorders : official journal of the Movement Disorder Society](#), 19(5) :565–571, 2004.



L. Tanguy, F. Sajous, and N. Hathout.

Évaluation sur mesure de modèles distributionnels sur un corpus spécialisé : comparaison des approches par contextes syntaxiques et par fenêtres graphiques.

[Traitement Automatique des Langues](#), 56(2), 2015.