

Le projet EVOLEX (Outillage pour l') étude de la fluence et de l'accès au lexique

Xavier de Boissezon⁴ Lola Danet⁴ Cécile Fabre¹ Jérôme
Farinas² Bruno Gaume¹ Nabil Hatout¹ Lydia-Mai Ho-Dac¹
Mélanie Jucla³ Patrice Péran⁴ Bénédicte Pierrejean¹ Julien
Pinquier² Ludovic Tanguy¹

UE TAL, 18.12.2017



Plan

- 1 EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 Protocole : "Cahier de passation" pour recueillir des "accès au lexique"
- 3 Données à disposition : des paires de mots et des fluences
- 4 Caractérisation linguistique des paires de mots
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)

Plan

- 1 **EVOLEX** : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 Protocole : "Cahier de passation" pour recueillir des "accès au lexique"
- 3 Données à disposition : des paires de mots et des fluences
- 4 Caractérisation linguistique des paires de mots
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)

EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique

Objectif premier (TONIC, Octogone-Lordat) : Fluence et Accès au lexique

- Mesurer la Fluence i.e. rapidité avec laquelle
 - ① nous accédons aux informations présentes dans notre lexique mental
 - ② nous associons ces informations les unes avec les autres
- La fluence reflèterait l'organisation du lexique (mental)

Objectifs secondaires

IRIT Améliorer les systèmes de reconnaissance vocale

CLLE-ERSS Tester et valoriser nos modèles et techniques de sémantique distributionnelle

Tradition : Mesurer la *fluence verbale* par des exercices d'*Évocation verbale*

Fluence *capacité d'un individu à évoquer et donner une série de mots à partir d'un champ sémantique (encore appelé champ lexical), ou d'un terme générique, ou du thème qui lui est proposé, exemples : énoncer le plus grand nombre de noms d'animaux, d'instruments de musique [...]*

Fluence moyenne pour un locuteur sain : 90 mots/minute

Évocation *épreuve ou exercice consistant à donner oralement ou par écrit plusieurs mots se rapportant par exemple à un champ lexical déterminé (exemple : dire tous les noms d'animaux connus, les jours de la semaine...) ou consistant à trouver un synonyme, un antonyme. L'évocation permet de tester ou de travailler l'accès au stock lexical et sa constitution*

Brin-Henry F., Courrier C., Lederlé E., Masy V. (2011) Dictionnaire d'orthophonie, Isbergues : Ortho Edition.

Problèmes de fluence verbale : trouble des fonctions exécutives ou "démence" sémantique

Une faible fluence (< 14 mots/minute) peut être due à

Des déficits exécutifs

fonctions exécutives *"fonctions de direction" permettant, lors de la réalisation d'une tâche, la définition du but ou des objectifs à atteindre, d'une stratégie pour y parvenir, le contrôle de sa mise en oeuvre et des résultats. Elles correspondent donc à des fonctions de haut niveau (des fonctions de contrôle), impliquées dans de nombreuses formes d'activités cognitives*

Une perturbation des représentations sémantiques, altération de la mémoire sémantique

mémoire sémantique mémoire des mots, des concepts, des connaissances que nous avons sur le monde et sur soi, affranchies de leur contexte spatial et temporel d'acquisition)

Godefroy O. et le Grefex. (2008). "Fonctions exécutives et pathologies neurologiques et psychiatriques – évaluation en pratique
Dubois B., Michon A. et al. (2015) Démences, Montrouge : Doin

« clinique », Marseille : Solal.

Tests d'évocation verbale

Deux types d'accès, deux types de tests

- 1 Formel (ex : mots commençant par "V", "R", etc.)
- 2 Sémantique (ex : noms d'animaux, synonymes du mot X, idées associées, etc.)

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

Fluences verbales Dire le plus de mots dans un temps imparti

Dénomination Dire un mot dénommant une image

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

Tests de fluence verbale

écrit	Thurstone, L. L. & Thurstone, T. G. (1938). Primary Mental Abilities. Chicago : Univ. of Chicago Press.
oral, accès formel	Benton, A. L. (1967). Problems of test construction in the field of aphasia Cortex, 3, 32–58. Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. Neuropsychologia, 5, 135-140.
oral, accès sémantique	Isaacs, B., & Kennie, A. T. (1973). The Set test as an aid to the detection of dementia in old people. The British Journal of Psychiatry, 123, 467-470.

Proposition du projet EVOLEX

- 1 Recueil de données auprès de locuteurs "sains"
 - Un protocole en 8 étapes 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.)

Plan

- 1 EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 **Protocole : "Cahier de passation" pour recueillir des "accès au lexique"**
- 3 Données à disposition : des paires de mots et des fluences
- 4 Caractérisation linguistique des paires de mots
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)

Protocole : "Cahier de passation" en 8 étapes (1h)

5 étapes de contrôle et de calibrage

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

Protocole : "Cahier de passation" en 8 étapes (1h)

3 étapes dédiées Evolex

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* → ? ...

RQ 30 stimuli commun entre (7) et (8)

Plan

- 1 EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 Protocole : "Cahier de passation" pour recueillir des "accès au lexique"
- 3 **Données à disposition : des paires de mots et des fluences**
- 4 Caractérisation linguistique des paires de mots
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)

EVOLEx : des groupes de mots et des fluences

Récolte des données

- 1 Recueil auprès de locuteurs "sains" par des binôme d'étudiants de M1 de l'école d'orthophonie de Toulouse
 - Trouver des locuteurs présentant des "catégories" variées (âge, catégorie socio-prof., etc.)
 - Entre 15 et 20 passations par binôme
 - Recueil manuel des réponses pour les étapes contrôle
 - Recueil automatique des réponses pour les étapes de lecture de texte, stroop, fluences verbales, dénomination et génération
 - 2015, version expérimentale : 30 passations réalisées
 - 2016, 2017, nouvelle version du protocole et amélioration du logiciel EvoLex : 103 passations réalisées (Des tranches d'âge surreprésentées (> 30% de 18-29 ans et 15% de 50-59))
- 2 Correction des sorties logiciel EvoLex (stagiaires ortho.) : erreurs de reco. vocale, erreurs d'alignement temporel

60 Stimuli ordonnés pour la tâche de Génération

V0 → ~~V1~~ → V2

1- Entrecôte	13- Arrosoir	Pied	Chauve-Souris	Caillou
2- Escargot	14- Abricot Acteur	Concombre	Thermomètre	Ortie
3- Puzzle	Parc	Rail	Ballon	Stéthoscope
4- Ecureuil	Etoile	Plume	Sucre	Dame Jambes
5- Entonnoir	Siège Vitre	Kiwi	Couverture	Poire
6- Biberon	Scarabée	Terrier	Noce	Marionnette
7- Sœur	Nœud	Brochette	Echelle	Palmier
8- Clef	Tournevis	Trou	Démon	Squelette
9- Couvercle	Hirondelle	Chapiteau	Marmite	Siège Robot
10- Maladie	Tomate	Spaghettis	Cigarette	Volcan
11- Cactus	Sac	Chaise	Crabe	Fusée
12- Osier	Lézard	Varicelle	Canard	60- Escalier

Stimuli de V0 absents de V1 et V2 : Animal, Aube, Balançoire, Bec, Bidon, Bonbon, Bougie, Canapé, Champignon, Chapiteau, Chat, Chien, Clou, Couverture, Crocodile, Doigt, Fleur, Igloo, Lavabo, Lion, Légume, Magicien, Marteau, Miel, Mouton, Oiseau, Olive, Papillon, Perroquet, Pharmacie, Pinceau, Poil, Pôle, Raisin, Sapin, Serpent, Vélo

Correction des sorties logiciel EvoLex

Passation 490DF9113D1E11B7FBD32A1E704C7565

Date | 8 juillet 2017 10:54
Catégorie | Génération
Test | Génération sémantique (60 stim)

Entrecôte



Réaction en 1458 ms

Paires de mots issues de la tâche de Génération

Données issues de V0 après correction des sorties du logiciel EvoLex

	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	pissenlît	2,12	plante	2,05	piqûre	
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	réceptier	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	machin	1,17	machin	1,17	machin	
59	Varicelle	maladie	3,34	maladie	1,17	maladie	1,17	maladie	

Paires de mots issues de la tâche de Génération

Données issues de V0 (V1/2 en cours de recueil)

- 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
- 1561 paires conservées, 569 paires distinctes

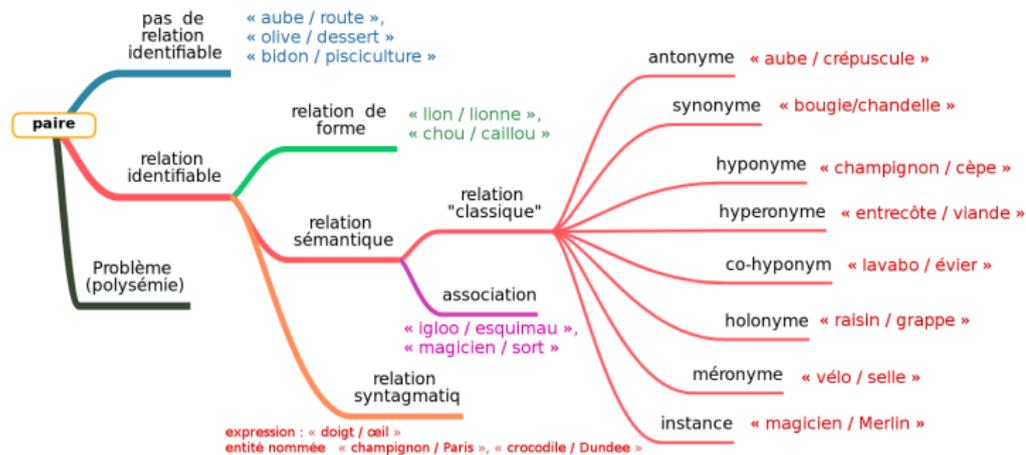
Plan

- 1 EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 Protocole : "Cahier de passation" pour recueillir des "accès au lexique"
- 3 Données à disposition : des paires de mots et des fluences
- 4 Caractérisation linguistique des paires de mots**
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)

Caractérisation des mots issues de la tâche de Génération

Annotation manuelle des relations sémantiques

● Double annotation + adjudication



Mesures de proximité sémantique

Caractérisation des mots issues de la tâche de Génération

6 Mesures de proximité sémantique :
3 méthodes, 2 versions par méthode

Cooccurrence en corpus FRWAC, Pointwise Mutual Information

- fenêtre 3 mots gauche/droite
- fenêtre phrase

Proximité lexicale « Jeux de mots », Complétion et lissage avec l'algorithme Prox (Bruno Gaume)

- Relation « idée associée »
- Relation « synonyme »

Similarité distributionnelle FRWAC, Word2vec

- Contexte = cooccurrence brute
- Contexte = relation syntaxique (Talismane, Urieli 2013)

Caractérisation des mots issues de la tâche de Génération

identifiant	stimulus corrigé	SIMILARITE	ASSOC	SYNTAGMA	PHONO/MORPH	POS	polyémie	N_jdm_r associated edge [0.34426541]	N_jdm_r associated prox [0.46134421]	N_jdm_r syn edge [0.06630699]	N_jdm_r syn prox [0.10639659]	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		x					0 : 0.00000000	0.00020092 : 0.0535488	0 : 0.00000000	0.00000000 : 0.00000000	0.716536036546	0.58595215267
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 : 3.21460644	0.02385163 : 6.3386805	0 : 0.00000000	0.00000000 : 0.00000000	0.585114703217	0.498159440672
SRP-5	abricot.kiwi	cohyponym						0 : 0.00000000	0.00039997 : 0.10621317	0 : 0.00000000	0.00000000 : 0.00000000	0.851678422504	0.601300971653
SRP-6	abricot.noyau	mero						193 : 1.52437111	0.0698082 : 1.85377657	0 : 0.00000000	0.00000000 : 0.00000000	0.355402325095	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	0.282082027716
SRP-16	ampoule.baionnette												
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					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	0.705806620646	0.701339379011
SRP-14	ampoule.lumière		x			x		400 : 3.15931837	0.01795581 : 4.76821632	0 : 0.00000000	0.00414214 : 0.94588467	0.84587515353	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		-r_abs	-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 : 0.11256785	0 : 0.00000000	0.00001859 : 0.00424515	0.1569137592677	0.000000000000
SRP-20	animal.chien	hyponym						672 : 5.30765496	0.00428259 : 0.11381329	0 : 0.00000000	0.00010261 : 0.02343166	0.80412835526	0.000000000000
SRP-21	animal.compagnie			nden				55 : 0.43440628	0.00003196 : 0.00848707	0 : 0.00000000	0.00000000 : 0.00000000	0.190661582200	0.000000000000
SRP-22	animal.fouine	hyponym						52 : 0.41071139	0.00007716 : 0.02049006	0 : 0.00000000	0.00000000 : 0.00000000	0.513078600000	0.000000000000
SRP-23	animal.humain	hyponym						0 : 0.00000000	0.00014104 : 0.03745357	0 : 0.00000000	0.00003920 : 0.08978978	0.451307860000	0.000000000000
SRP-24	animal.lion	hyponym						361 : 2.85129483	0.00025391 : 0.06742652	0 : 0.00000000	0.00003101 : 0.00708134	0.451307860000	0.000000000000
SRP-25	animal.mignon					adj		0 : 0.00000000	0.00001578 : 0.00419042	0 : 0.00000000	0.00012050 : 0.02751898	0.451307860000	0.000000000000
SRP-26	animal.mouton	hyponym						23 : 0.18166081	0.00011265 : 0.02991453	0 : 0.00000000	0.00000000 : 0.00000000	0.451307860000	0.000000000000
SRP-27	animal.oiseau	hyponym						61 : 0.48179605	0.00092663 : 0.24606923	0 : 0.00000000	0.00037876 : 0.00800000	0.451307860000	0.000000000000
SRP-28	animal.végétal	cohyponym						54 : 0.42650798	0.00006368 : 0.01691040	0 : 0.00000000	0.00010261 : 0.02343166	0.80412835526	0.000000000000
SRP-29	aube.aurore	syn				x		100 : 0.78982959	0.00552794 : 1.46796016	78.0 : 8.12050985	0.00000000 : 0.00000000	0.451307860000	0.000000000000
SRP-30	aube.communion		x			x		0 : 0.00000000	0.00032927 : 0.08743858	0 : 0.00000000	0.00000000 : 0.00000000	0.451307860000	0.000000000000
SRP-31	aube.crépuscule	antonyme				x		40 : 0.31593184	0.00293956 : 0.79041291	40.0 : 4.00000000	0.00000000 : 0.00000000	0.451307860000	0.000000000000
SRP-40	aube.Hiroshima							0 : 0.00000000	0.00000000 : 0.00000000	0.00000000 : 0.00000000	0.00000000 : 0.00000000	0.451307860000	0.000000000000
SRP-32	aube.jour	holonyme											

Plan

- 1 EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 Protocole : "Cahier de passation" pour recueillir des "accès au lexique"
- 3 Données à disposition : des paires de mots et des fluences
- 4 Caractérisation linguistique des paires de mots
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)**
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)

Plan

- 1 EVOLEX : outils et protocole pour l'étude de la fluence et de l'accès au lexique
- 2 Protocole : "Cahier de passation" pour recueillir des "accès au lexique"
- 3 Données à disposition : des paires de mots et des fluences
- 4 Caractérisation linguistique des paires de mots
- 5 Similarité entre structure d'une ressource lexicale et lexique mental (Bruno)
- 6 Analyses statistiques et exploration des embeddings, identification de profils de locuteurs, de mots... (Ludovic)