

SL00405T_SED – Outils d'analyse en SDL 4

Collecte de données - Dossier

Mathilde Viana

N°22010624

Ce dossier a été conçu dans le cadre de l'UE SL00405T - Outils pour l'analyse des SDL 4 afin de nous familiariser avec la collecte de données langagières. En s'appuyant sur le projet Evolex¹, il nous a été demandé de créer un questionnaire afin que des personnes parlant français, caractérisent cinq paire de mots.

I. Description détaillée du questionnaire

Structure du questionnaire

Le questionnaire est composé de 32 questions divisées en 10 écrans :

Écran 1 : Introduction

- présentation du projet (texte)
- bon pour accord pour le traitement anonyme des données (oui/non)

Écran 2 : Profil du répondant

- phrase d'introduction (texte)
- année de naissance (liste déroulante)
- niveau d'études (boutons radio)
- parcours en lien avec les sciences du langage (bouton radio : oui ou non)²
- français comme langue maternelle (bouton radio : oui ou non)
- troubles du langage (bouton radio : oui ou non)
- appétence pour les loisirs suivants (choix multiples : mots croisés / fléchés, Scrabble, poésie, apprentissage de langues étrangères, voyages, lecture, cinéma)

Écran 3 : A lire avant de commencer

- définition de chaque terme utile pour les questions suivantes : synonymie, antonymie, hyperonymie & hyponymie, métonymie & holonymie, co-hyponymie, collocation, jeux de mots, association d'idées avec une mise en forme attractive et des exemples clairs (texte)
- validation de la bonne lecture de ces définitions (oui/non)

Écran 4 à 8 : Caractérisation des paires de mots

Chaque écran permet de caractériser une seule et unique paire de mots : ils sont tous construits selon la même structure :

- phrase rappelant les mots à analyser (texte)
- invitation à écrire une phrase avec les deux termes (zone de texte long)
- choix du lien entre les deux mots (bouton radio avec les choix suivants : synonymie, antonymie, hyperonymie & hyponymie, métonymie & holonymie, co-hyponymie, collocation, jeux de mots, association d'idées). Un rappel des définitions se trouve en aide contextuel de cette question afin que l'utilisateur puisse s'y référer si besoin.

Écran 9 : Force des liens sémantiques entre les paires de mots

- définition de la proximité de sens entre les deux mots de chaque paire (tableau dans lequel 0 signifie qu'il n'y a pas de lien sémantique entre les deux termes, légère pour 1, très forte pour 5)

¹ de Boissezon, X. et al. (2019). EVOLEX : approches psycholinguistique et computationnelle de l'accès au lexique et de la proximité sémantique entre paires de mots. FORUM À LA CROISÉE DES SCIENCES : Interagissez, Imaginez, Innovez - FACS3I, Jan 2019, Toulouse, France. ([hal-02047651](https://hal.archives-ouvertes.fr/hal-02047651))

² Le choix entre (oui/non) et les boutons radios dotés des réponses oui/non est purement d'ordre esthétique.

Écran 10 : Évaluation

- phrase de remerciements (texte)
- notation de l'intérêt du questionnaire (5 boutons radios, 1 étant très ennuyeux et 5 très intéressant)
- notation de la facilité du questionnaire (5 boutons radios, 1 étant difficile et 5 facile)
- notation de la rapidité du questionnaire (5 boutons radios, 1 étant lent et 5 rapide)
- Invitation à laisser un commentaire (zone de texte court)

Structure de l'échantillon

Le questionnaire a obtenu 175 réponses complètes grâce à l'appel lancé via mes réseaux sociaux. J'ai pu toucher mes proches via Facebook mais aussi un grand nombre de personnes inconnues grâce à Twitter. Cela m'a permis d'accéder à des réponses hors de notre cercle étudiant et d'avoir des retours de personnes n'ayant jamais étudié les sciences du langage (84.57% des répondants).

Tableau 1 : description des variables concernant l'échantillon du questionnaire

Variables étudiées	Nature	Étendue	Mode	% des réponses associées
Année de naissance*	Variable qualitative ordinale	1900 - 2022	1999	8 %
Niveaux d'étude	Variable qualitative ordinale	BAC +5 / Master ou plus ; BAC +3 / Licence ; BAC +2 / BTS, DUT, DEUG ; BAC ; CAP / BEP	Bac +5 ou plus	48 %
Connaissance des sciences du langage	Variable qualitative nominale	Oui / Non	Non	84.57%
Français comme langue maternelle	Variable qualitative nominale	Oui / Non	Oui	98.86% (2 personnes ont répondu non)
Troubles du langage	Variable qualitative nominale	Oui / Non / Autre	Non	89.14%
Loisirs	Variable qualitative nominale	Lecture, Cinéma, Mots croisés / mots fléchés, Scrabble, Poésie, Apprentissage de langues étrangères, Voyages	Lecture	89.71%

* Cette donnée ne me semble pas exploitable en l'état. J'ai donc converti les années de naissance en tranche d'âge. Les 30-44 ans ont le plus participé à 41,71%.

Avis sur le questionnaire

Tableau 2 : description des variables concernant l'avis des répondants sur le questionnaire

Variables étudiées	Nature	Étendue	Moyenne arithmétique	Ecart type	Médiane	% des réponses associées *
Intérêt	Variable quantitative discrète	2-5	4,03	0,83	4	42,29 %
Facilité	Variable quantitative discrète	1-5	3,37	1,05	3	32,57 %
Rapidité	Variable quantitative discrète	1-5	3,98	0,9	4	37,71 %

* % des réponses associées à la médiane

La majorité des personnes qui n'ont pas complété leur réponse se sont arrêtées dès le début (66 personnes) ou au troisième écran : « A lire avant de commencer » (71 personnes). Cela représente donc au total 84,5% des non répondants qui, selon mes suppositions, ont trouvé ça trop contraignant avant de s'engager à répondre.

Le panel des répondants se voulant volontairement large, la difficulté ressentie est assez hétérogène. Néanmoins, 74,29% et 70,85% ont répondu 4 ou 5 respectivement en faveur de l'intérêt et la rapidité du questionnaire. Sur les 27 commentaires laissés spontanément, 24 sont très positifs / encourageants.

II. Analyse des caractéristiques attribuées à chaque paire de mot

Types de relations

Tableau 3 : nombre de répondants pour chaque caractéristique et chaque paire (sur 175 répondants au total). En gras, les relations majoritaires ou principales identifiées.

	Synonymie	Antonymie	Hyperonymie & hyponymie	Méronymie & holonymie	Co-hyponymie	Collocation	Jeux de mots	Association d'idées	Aucune relation
Chauve-souris & Halloween	0	0	7	2	1	21	0	133	11
Écureuil & chèvre	0	0	4	1	157	0	1	1	11
Maladie & mort	0	0	6	10	6	33	1	104	15
Plume & oiseau	0	0	9	148	0	8	0	9	1
Sucre et glucide	67	0	67	22	9	3	0	6	1

L'association d'idées a été désignée comme relation majoritaire pour les paires « chauve souris & Halloween » ainsi que « maladie & mort ».

Dans les exemples, la chauve souris a souvent été désigné comme déguisement (59 occurrences) ou décoration (51 occurrences) pour Halloween : « Pour Halloween j'ai décoré ma maison avec des motifs de chauve-souris. » ou « Je me suis déguisée en chauve-souris pour Halloween ». Quant à la mort et la maladie, diverses associations ont été décrites : un lien de cause à effet ou une absence de cause à effet. Les exemples sont de type « Certaines maladies mènent inévitablement à la mort. » ou à l'inverse « Toute maladie n'entraîne pas la mort »

«Écureuil & chèvre » ont pour relation majoritaire **la co-hyponymie**. Les exemples évoquent souvent les termes « animaux » ou « mammifères » comme hyperonymes : « Un écureuil et une chèvre sont des animaux » ou « La chèvre et l'écureuil sont des mammifères »

Pour la plupart des répondants, « plume » et « oiseau » entretiennent majoritairement **une relation d'holonyme / méronyme** ; la plume étant une partie de l'oiseau comme le montre les exemples suivants : « Certains oiseaux ont des plumes de toutes les couleurs. » ou « L'oiseau est couvert de plumes ». Certains répondants se sont laissés aller à de la poésie en utilisant l'image de la plume pour écrire : « Il écrivit le mot "Oiseau" avec sa plus belle plume ! » ou « Pourquoi écrire de sa plume alors qu'il suffit d'écouter l'oiseau ? » Mais cela n'a empêché à ce que la tendance soit très nette : 84,57% ont choisi « méronymie & holonymie »

« Sucre » et « glucide » sont une exception. Deux relations principales arrivent ex-aequo pour cette paire : la relation de **synonymie** et la **relation hyperonymie / hyponymie**. Les exemples « Le sucre est un glucide » ou « le glucide est un sucre » reviennent très souvent (63 occurrences) et sont en faveur de la synonymie. De nombreux exemples vont aussi dans le sens de la relation hyperonyme / hyponyme : « 20 g de glucides dont 19 de sucre. », « Les glucides sont un type de sucre »; etc. Le choix de la relation semble dépendre du sens que l'on donne au lexème « sucre ». La définition chimique entraîne une relation de synonymie

Degré de proximité

Tableau 4 : analyse du degré de proximité des paires selon les répondants

Rappel : 0 signifie qu'il n'y a aucun lien de proximité entre les deux mots ; 5 que le lien est extrêmement fort.

Accord sur le degré de proximité	Paires	Degré de proximité médian	Degré de proximité moyen*1	Pourcentage d'accord sur le degré de proximité moyen	Nombre d'autres degrés répondus*2					
					0	1	2	3	4	5
Moyen < 40%	Maladie & mort	3	2,73 (3)	38,29 %	8	14	30	67	39	17
	Sucre & glucide	5	4,36 (4)	26,29 %	6	1	4	10	46	108
	Chauve-souris & Halloween	1	1,39 (1)	22,29 %	62	39	29	28	13	2
	Écureuil & chèvre	2	1,41 (1)	21,14 %	41	37	40	31	18	8
Faible < 20%	Plume & oiseau	4	3,49 (3)	15,43 %	11	12	18	27	53	54

*1 Les données sont arrondies à deux dixièmes. Le chiffre entre parenthèse correspond au degré arrondi sans décimale. *2 Le mode est souligné.

Malgré des tendances assez nettes sur les types de relations, aucun accord fort n'a été trouvé en ce qui concerne les degrés de proximité. Une bonne connaissance des relations entre les termes et une bonne utilisation en contexte ne permet pas d'assurer un accord fiable sur le degré de proximité.

Quand l'accord de proximité est faible, la répartition des réponses est assez similaire :

Exemple pour écureuil & chèvre > **0:41, 1:37, 2:40**, 3:31, 4:18, 5:8

Exemple pour plume & oiseau > 0:11, 1:12, 2:18, 3:27, **4:53, 5:54**

En comparant ces données avec celles des relations, il semble que les paires étant vues comme des associations d'idées ont le meilleur taux d'accord de proximité (maladie & mort pour 38,29% et chauve souris & Halloween). Sucre & glucide vu comme un synonyme pour beaucoup occupe la deuxième place avec 26,29%.

A contrario, malgré des liens fort et reconnus par les répondants (89,71% d'accord pour la relation de co-hyponymie pour écureuil & chèvre, 84,51% pour la relation méronyme / homonyme de plume & oiseau), l'accord sur le degré de proximité est faible. Mon hypothèse est que ce sont des relations moins connues qui font hésiter le répondant sur le degré à déterminer.

Conclusion

Travailler sur ce questionnaire m'a permis d'entr'apercevoir la puissance de l'outil Lime Survey. J'ai pu voir des pistes d'améliorations sur les données (notamment l'année de naissance) et je suis très curieuse de lire les conclusions de mes camarades de promotion afin de valider ou non ma problématique.