HLIN512I Projet CMI Encadré par Christian Retoré

_

Étude des préférences dans l'expression de la quantification

_

Alexandre CANTON CONDES, Ambre LAMOUCHI, Jérémie ROUX et Yasmine SMAILI 2019 - 2020







Table des matières

Ι	Introduction	3
II	Présentation du problème	4
1	Quantification universelle	4
2	Quantification existentielle	4
3	Ambiguïté de portée des quantificateurs	5
III	Diagramme de Gantt	6
IV	Description du travail réalisé	7
	Structure et développement 1.1 Hébergement du site web et dialogue avec la base de données	7 7 7 7
2	Phrases et propositions à évaluer	8
3	Diffusion du formulaire	9
	Résultats de l'étude 4.1 Sens donné aux propositions	11 11
\mathbf{V}	Perspectives pour le second semestre	16
1	Diversité des phrases	16
2	Autres améliorations	16
VI	Annexe	17
	Compte rendu des réunions 1.1 Lundi 14/10/19 1.2 Lundi 04/11/19 1.3 Lundi 18/11/19 1.4 Lundi 02/12/19 1.5 Jeudi 18/12/19	18 18
2	Table Utilisateur - Schéma de la base de données	19
3	Données récoltées	20
Ré	férences	21

Première partie

Introduction

Nous sommes 4 étudiants de 3ème année de licence informatique de la Faculté des Sciences (Université de Montpellier) :

- Alexandre CANTON CONDES (non CMI)
- Ambre LAMOUCHI
- Jérémie ROUX
- Yasmine SMAILI

Dans le cadre de l'UE HLIN512I (Projet CMI), nous avons choisi le sujet « Étude des préférences dans l'expression de la quantification » proposé par M. Christian Retoré ¹, enseignant-chercheur à la Faculté des Sciences et responsable de l'équipe TEXTE ² du LIRMM ³ que nous remercions pour son aide et sa présence lors de nos réunions quinzomadaires (voir comptes-rendus des réunions en Annexe).

Durant ce semestre, nous avons développé un formulaire web qui permet d'étudier les préférences de différentes personnes dans l'expression de phrases comportant des quantificateurs (« tout », « chaque », « tous les », « un », « certain(s) », « quelque(s) », « des »). Notre formulaire est composé de 12 phrases (que l'on appellera phrases exemples) comportant ou non des ambiguïtés sémantiques (défini plus tard) et de portées ⁴ (voir Fig. 1). Deux à quatre propositions de phrases par phrase exemple doivent être évaluées par l'utilisateur en fonction de si elles lui paraissent avoir le même sens que la phrase exemple. L'utilisateur a trois choix de notations : « oui », « peut-être » et « non » que l'on désignera maintenant par OUI, PEUT-ÊTRE et NON.

Fig. 1 – Exemple d'ambiguïté de portée avec l'opérateur de négation

Le temps de réponse aux questions est chronométré de façon à pouvoir évaluer l'éventuelle hésitation de la personne sondée lorsqu'elle répond. Les ordres des phrases et des propositions de phrases sont aléatoirisés afin d'évaluer si cela joue un rôle dans le choix de réponse. À la fin du questionnaire, on demande au sondé de renseigner si le français est sa langue maternelle, son statut et son secteur d'activité ou d'étude. Ces données vont nous permettre de savoir si le milieu et la maîtrise de la langue française par la personne qui rempli le formulaire joue un rôle dans ses choix de réponse.

Ce formulaire web est développé en HTML/CSS sur une base de modèle ⁵ Bootstrap. La programmation côté serveur est réalisée en PHP avec une base de données MySQL. La programmation côté client est réalisée en JavaScript.

^{1.} Site Web de M. Christian RETORÉ: https://www.lirmm.fr/retore/

^{2.} TEXTE: Exploration et exploitation de données textuelles

^{3.} LIRMM: Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier

^{4.} Portée: Terme utilisé en linguistique pour désigner le domaine sur lequel un opérateur fait effet.

^{5.} Modèle: de l'anglais template

Deuxième partie

Présentation du problème

La thématique de notre sujet se concentre sur le fait de savoir s'il existe (ou pas) différentes manières de comprendre une même phrase. Cette différence de compréhension est appelée ambiguïté. Ce sujet est à cheval entre la logique et la linguistique dans les sciences cognitives. Les résultats obtenus grâce aux données collectées vont nous servir à démontrer l'ambiguïté de l'emploi des quantificateurs dans la langue française. En quoi le langage naturel est-il ambiguë?

Il existe différents types d'ambiguïtés:

- Ambiguïtés de prononciations liés aux graphèmes ⁶
- Ambiguïtés de portées (déjà défini)
- Ambiguïtés de la grammaire et de la sémantique ⁷

Nous nous concentrons sur l'ambiguïté sémantique liée à l'emploi de quantificateurs. Il existe différents types de quantificateurs.

1 Quantification universelle

La quantification universelle peut signifier intuitivement « attribuer la propriété Q à toutes les entités d'un ensemble P » [2]. En mathématique, cela s'écrit comme tel : « $\forall x \in P, Q(x)$ »

- Le quantificateur universel « Tout »: « Tout » est un quantificateur universel du premier ordre, il assigne une propriété à chacun des individus atomique et est incompatible avec une situation dans laquelle des individus particuliers satisfont la propriété. Selon Brissonet Lasersohn, il n'est pas analysé comme un véritable quantificateur, mais plutôt comme un «réducteur de marge». Son emploi repose sur l'observation de l'existence d'une règle: $P(x) \implies Q(x)$. [4]
- Le quantificateur universel « Chaque » : Contrairement à « Tout », ce quantificateur demande à ce que chaque entité du domaine soit observée indépendamment. Il a une sémantique distributive et est malvenu dans des phrases génériques. Son emploi est plutôt descriptif. [1] [3]

2 Quantification existentielle

La quantification existentielle affirme ou nie l'existence d'au moins une chose. En mathématiques : « $\exists x \in E$ » se lit « Il existe au moins un élément x de l'ensemble E ».

Parmi les quantificateurs existentiels on peut compter:

- Certain: Il est traité comme indéfini au sens strict dans de nombreuses approches, alors que plusieurs auteurs l'excluent de cette classe. Il semblerait que « Certain » ait en général une prédilection marquée pour les portées larges et intermédiaires.
- Un: Dans la langue française le sens du quantificateur « Un » n'est jamais compris dans les phrases, sa compréhension est toujours associé au chiffre 1. Son interprétation réelle équivaut à « au moins » dans les phrases dans lesquelles il est utilisé. Par exemple, « Pierre a un diplôme en Informatique » doit être compris « Pierre a au moins un diplôme en informatique » et non « Pierre a un seul diplôme en informatique ».

^{6.} Graphème: lettre ou groupe de lettres transcrivant un phonème (élément sonore du langage parlé).

^{7.} Sémantique: Étude du sens, de la signification des signes, notamment dans le langage.

Il faut noter qu'en français comme en logique, l'utilisation de tout équivaut à celui de chaque mais la réciproque ne fonctionne pas car dans la plupart des énoncés avec « chaque », « tout » induit une relation causale peu évidente ou même d'ailleurs quasi impossible.

3 Ambiguïté de portée des quantificateurs

Les quantificateurs disposent d'une portée dans leur domaine de définition. Elle est explicite dans une formule mathématique (exemple: « $\forall x \in \mathbb{N} : (x+4) \in \mathbb{N}$ » se lit « Quel que soit l'entier naturel x, (x+4) est un entier naturel. »). Cependant, cette portée est implicite et ambiguë dans le langage naturel.

Par exemple, la phrase « Un bus ramènera les invités », peut être retranscrite par les formules logiques:

$$\exists b, \, \forall i, \, \text{ramener}(b,i)$$
 (5)

$$\forall i, \exists b, \, \text{ramener}(b,i)$$
 (6)

Les formules (5) et (6) sont sémantiquement différentes. L'énoncé est plus ambiguë quand le sens est obscur pour celui qu'il l'interprète, ce qui est tout à fait naturel.

Troisième partie

Diagramme de Gantt

Afin de nous organiser pour ce TER, nous avons réalisé le diagramme de Gantt de la Fig. 2. Dès lors que nous avons choisi un sujet et formé le groupe, nous avons contacté M. Retoré afin de pas perdre de temps et créé le Git du TER⁸. La première étape a été de faire des recherches bibliographiques afin de comprendre le but du TER. Parallèlement, nous avons commencé la rédaction du rapport avec l'explication des articles que nous lisions. Une fois ces pré-requis passés, nous avons commencé le développement avec le CSS et le HTML qui constituent la base du formulaire. Nous avons ensuite développé la partie PHP et base de données. Puis, il a fallut créer des messages d'erreurs si tous les champs ne sont pas rempli par l'utilisateur et des protections pour ne pas pouvoir injecter n'importe quoi dans la base de données.

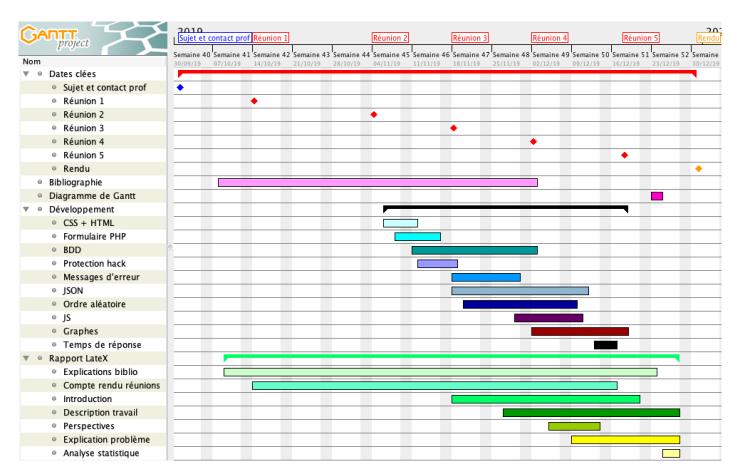


Fig. 2 – Diagramme de Gantt du TER

Ensuite, les 12 questions que nous voulions poser on été rassemblées dans un fichier JSON. A partir de ce fichier, nous avons aléatoirisé l'ordre dans lequel les phrases sont proposées à l'utilisateur (codé en JavaScript). Enfin, nous avons réalisé des graphes (également en JavaScript) afin d'afficher les résultats obtenus. Nous avons aussi utilisé du JS pour calculer le temps de réponse de l'utilisateur à toutes les questions qui lui sont posées pour exploiter ces données supplémentaires et afficher d'autres graphes. Nous avons terminé la rédaction du rapport et réalisé une analyse statistique après la diffusion du formulaire.

^{8.} Git du TER: https://gitlab.info-ufr.univ-montp2.fr/e20170000656/TER L3

Quatrième partie

Description du travail réalisé

Nous devons réaliser une étude statistique simple pour étudier les différences de sens et les préférences des locuteurs entre les différentes expressions de la quantification universelle (quel que soit, tout, tous les, chaque,...) et existentielle (il existe, un, certains, des...). Pour cela, nous réalisons un formulaire web concis (15 minutes maximum) et ergonomique pour les utilisateurs afin de récolter leurs réponses et les stocker dans une base de données.

1 Structure et développement

1.1 Hébergement du site web et dialogue avec la base de données

Afin de faciliter l'accès au site, nous faisons le choix de l'héberger chez l'un des membre du groupe. Les pages web sont stockées sur une serveur de stockage en réseau NAS (de l'anglais Network Attached Storage). Ce serveur est lié à un nom de domaine OVH (déjà acheté) et non référencé sur les moteurs de recherche. L'adresse web du formulaire est https://jeremieroux.fr/TERL3 et les résultats sont disponibles à l'adresse https://jeremieroux.fr/TERL3/resultats.php.

La base de données est créée grâce à l'application web *phpMyAdmin*. Le serveur de base de données relationnel MySQL (sous double licence GPL et propriétaire) utilise le langage de requête structuré SQL, il s'agit du deuxième logiciel de gestion de bases de données le plus utilisé après Oracle. En hébergeant de cette manière le site web, nous garantissons la sécurité et l'accessibilité des données et des pages web.

1.2 Conception de la base de données

La base de donnée est composée d'un unique table UTILISATEUR (voir en Annexe) qui a 153 colonnes:

- L'id de l'utilisateur auto-incrémenté (clef primaire)
- La date de soumission du formulaire
- La réponse à la question : « Le français est-il votre langue maternelle? »
- La réponse à la question : « Quel est votre statut? »
- La réponse à la question : « Quel est votre secteur d'activité? »
- L'adresse mail si l'utilisateur l'a renseignée
- Les réponses données au 33 propositions de phrases (oui = « o », peut-être = « p » ou non = « n »)
- L'ordre des 12 phrases car l'ordre des phrases est aléatoirisé (de 1 à 12).
- L'ordre des 33 propositions de phrases car l'ordre des propositions est aléatoirisé pour une phrase donnée.
- Le temps de réponse à chacune des 12 phrases.
- Le temps de réponses à chacune des 33 propositions.
- Le temps de lecture de chacune des 12 phrases.
- Le temps de réponse moyen aux propositions de chacune des 12 phrases.

1.3 Conception du formulaire

Le but de cette étude est de rassembler un maximum d'informations sur les préférences des locuteurs en fonction de si le français est leur langue maternelle, de leur domaine d'étude (art, droit et science politique, économie et gestion, lettres et langues, santé, sciences appliquées, sciences humaines et sociales, autre) et de

leur statut professionnel (étudiant, enseignant, autre). Ces trois questions simples ont été convenues car elles préservent l'intimité des sondés et elles permettent de les classer dans des catégories qui indiquent potentiellement une familiarité avec la logique.

2 Phrases et propositions à évaluer

Voici les 12 phrases que nous avons choisi avec à chaque fois les 2 à 4 propositions à évaluer. Cela fait 33 notes à donner pour chaque utilisateur. Cela prend à peu près 10 à 15 minutes pour répondre au questionnaire.

- 1. Un gardien surveille chaque salle du musée.
 - Un seul gardien est dans une salle dans laquelle il a accès à toutes les caméras du musée.
 - Un gardien différent doit surveiller la salle qui lui a été assignée.
 - Un gardien doit faire une ronde dans le musée pour surveiller chaque salle.
- 2. Toutes les écoles sont fermées un certain jour.
 - Toutes les écoles sont fermées le même jour.
 - Toutes les écoles sont fermées un jour différent.
 - Toutes les écoles sont fermées un jour (pas forcément différent des autres).
- 3. Chaque magasin a son propre jour de fermeture.
 - Les magasins ont un jour de fermeture forcément différents les uns des autres.
 - Deux magasins peuvent être fermés le même jour.
- 4. Un bus ramènera les invités.
 - Un bus pour tous les invités.
 - Un bus pour chaque invité.
 - Un même invité dans tous les bus.
- 5. Les enfants ont reçu un cadeau lors du dernier jour d'école.
 - Chaque enfant a reçu un cadeau différent.
 - Les enfants n'ont pas eu de cadeau.
 - Tous les enfants ont reçu le même cadeau.
- 6. Les entreprises vont faire faillite.
 - Certaines entreprises vont faire faillite.
 - Toutes les entreprises vont faire faillite.
- 7. Un cercle touche chacun des carrés.
 - Un même cercle touche tous les carrés.
 - Un cercle différent touche chaque carré.
 - Deux carrés peuvent être touchés par le même cercle ou par des cercles différents.
- 8. Des crayons gratuits seront distribués aux adhérents.
 - Chaque adhérent aura droit à maximum un crayon.
 - Les adhérents auront droit à autant de crayons qu'ils le souhaitent.

- 9. Tous les musées de la ville ferment certains jours de la semaine.
 - Parmi tous les musées de la ville, au moins un est fermé chaque jour.
 - Tous les musées de la ville ferment le ou les même jours.
 - Les musées de la ville ferment indépendamment les uns des autres.
- 10. Des étudiants cherchent du travail.
 - Tous les étudiants cherchent le même travail.
 - Tous les étudiants cherchent un travail différent.
 - Certains étudiants cherchent le même travail.
 - Certains étudiants cherchent un travail différent.
- 11. Tous les touristes jettent leurs mégots dans la nature.
 - Tous les touristes fument et certains jettent leurs mégots.
 - Tous les touristes qui fument jettent leurs mégots.
- 12. Il existe des gens qui détestent les pizzas à l'ananas.
 - Personne n'aime les pizzas à l'ananas.
 - Certaines personnes aiment les pizzas mais pas à l'ananas.
 - Tout le monde aime les pizzas mais certaines personnes détestent quand elles sont à l'ananas.

3 Diffusion du formulaire

Le formulaire web a été diffusé par les membres du groupe à certains de leurs proches et leurs amis afin d'acquérir le plus de données possible. Nous avons diffusé le lien sur le serveur Discord des Licences Informatiques de l'Université de Montpellier⁹. Nous avons ainsi récolté les réponses au questionnaire de 52 personnes qui ont chacune donné leur avis sur les 12 phrases à travers les 33 propositions de phrases (voir section précédente).

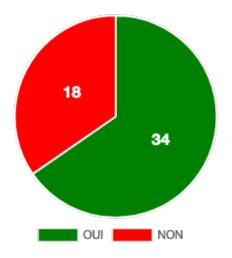


Fig. 3 – Réponses à la question : « Le français est-il votre langue maternelle? »

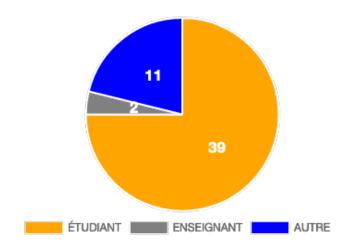


Fig. 4 – Réponses à la question : « Quel est votre statut? »

^{9.} Lien du serveur discord: https://discord.gg/GkrGZGn

On constate à travers l'analyse des Figs. 3 et 4 que la majorité des personnes ayant rempli le questionnaire sont des étudiants parlant couramment français. De plus, la plupart des sondés étudient dans le secteur des sciences appliquées (et en particulier en informatique) ce qui est logique étant donné que le lien du formulaire a été diffusé à des étudiants en informatique (voir Fig. 5).

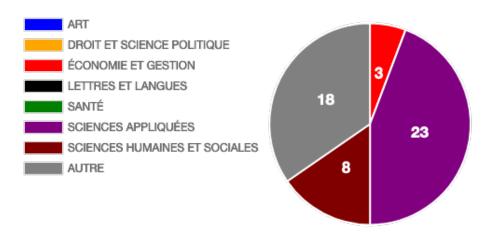


Fig. 5 – Réponses à la question : « Quel est votre secteur d'activité ou d'étude ? »

4 Résultats de l'étude

4.1 Sens donné aux propositions

On constate on observant la Fig. 6 que le nombre de OUI, NON et PEUT-ÊTRE sont à peu près les mêmes laissant suggérer qu'aucun de ces moyens de notation ne semble être privilégié par les utilisateurs. D'autre part, on constate que le temps moyen de réflexion lorsque le sondé répond PEUT-ÊTRE est de 8,9 secondes alors qu'il est seulement de 7,3 secondes quand il répond NON et 6,1 secondes quand il répond OUI (Fig. 7). On peut donc conclure que le temps de réflexion est plus long quand l'utilisateur répond PEUT-ÊTRE, démontrant que celui-ci hésite longtemps avant de le sélectionner. Les propositions comptant un grand nombre de PEUT-ÊTRE peuvent donc être qualifiées comme très ambiguës d'autant plus que le temps de réflexion y est important.

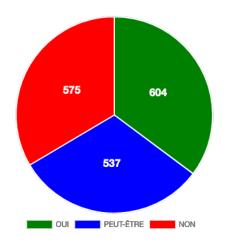


Fig. 6 – Nombre de réponse par sens donné aux propositions

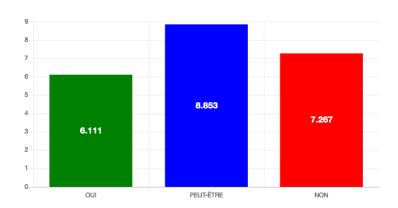


Fig. 7 – Temps de réflexion moyen par sens donné aux propositions (en secondes)

4.2 Ordre des phrases

On veut démontrer la pertinence de l'aléatoirisation de l'ordre des phrases. On constate avec les résultats rassemblés dans la Fig. 8 que les phrases ont un temps de réflexion associé (de lecture de la question et de notation des propositions) plus important quand elles sont abordées au début du questionnaire qu'à la fin. Ce temps de réflexion est globalement décroissant au fur et à mesure que l'on avance dans le questionnaire. L'intérêt d'aléatoiriser l'ordre des phrases est donc essentiel pour garantir la validité des résultats. La variable du temps de réflexion peut donc être exploitée avec un grand nombre de données collectées car chaque sondé aura certes les mêmes phrases mais elles lui seront exposées dans un ordre différent des autres. Dans ce même but de pouvoir exploiter les données de temps de réflexion, nous avons également aléatoirisé l'ordre des propositions pour une même phrase.

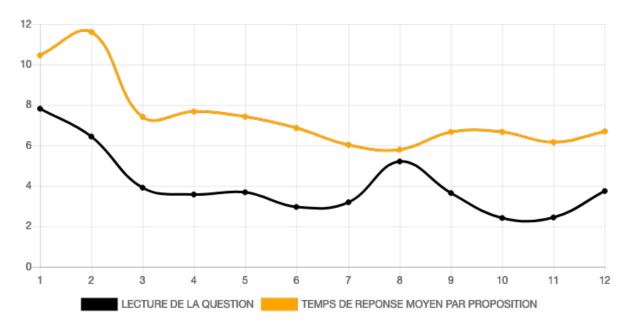


Fig. 8 – Temps de réflexion moyen par propositions (en secondes) en fonction de l'ordre dans lequel les questions sont abordées

4.3 Analyse des résultats

Nous allons étudier toutes les phrases et les propositions associées à ses phrases en mettant cela en relation avec la Fig. 9 qui représente la proportion de OUI, NON et PEUT-ÊTRE pour chaque proposition, la Fig. 10 qui est une simplification de la Fig. 9 et la Fig. 11 qui rassemble les données de temps de réponse pour chaque proposition ainsi que le temps de lecture de la phrase exemple. On peut classer ces phrases en 6 catégories (voir légende de la Fig. 10).

4.3.1 Catégorie n°1 - « OUI » majoritaire

Les propositions pour lesquelles le nombre de sondé ayant répondu OUI est majoritaire induit que cette proposition a le même sens que la phrase exemple selon la majorité des sondés. On peut par exemple citer la P4.1 lo pour laquelle il est vrai selon les sondés d'interpréter « Un bus ramènera tous les invités » comme « Un

^{10.} Px.y désigne la proposition y de la phrase n°x

bus pour tous les invités ». Il est évident ici qu'il y a un seul bus et que tous les invités montent dans le même. On peut également citer la P2.3 qui est majoritairement notée comme étant vraie. En effet, la phrase « Toutes les écoles sont fermées un certain jour. » et la proposition P2.3 « Toutes les écoles sont fermées un jour (pas forcément différent des autres). » met en avant l'ambiguïté du quantificateur « certain » en disant que des écoles ferment le même jour et d'autres pas. Le sondés notent majoritairement cette proposition comme étant vraie car elle englobe toutes les possibilités liées à cette ambiguïté.

On constate que les temps de réponse à la P4.1 et la P2.3 sont respectivement de 4,965 secondes et de 7,65 secondes, ce qui est plutôt rapide quand on compare cela aux autres temps de réponse. La Fig. 7 montre également que le temps de réflexion moyen lorsque le sens de la proposition est OUI est inférieur aux autres : il est de 6,111 secondes alors qu'il est de 8,853 secondes pour PEUT-ÊTRE et 7,267 secondes pour NON.

4.3.2 Catégorie n°2 - « NON » majoritaire

Il y a des propositions pour lesquelles les sondés ont majoritairement répondu NON. C'est le cas pour la P4.2 et la P4.3 correspondant à la phrase 4 « Un bus ramènera les invités. ». Dans les deux cas, la proposition est absurde. La P4.2 « Un bus pour chaque invité. » sous-entend que chaque invité sera ramené par un bus et qu'il sera le seul dedans, si on suit la logique il faudra autant de bus que d'invités. Pour la P4.3 « Un même invité dans tous les bus. », il s'agit d'une phrase test totalement absurde pour vérifier que les sondés ne répondent pas n'importe quoi. C'est également le cas pour la P5.2 « Les enfants n'ont pas eu de cadeau. » de la phrase exemple « Les enfants ont reçu un cadeau lors du dernier jour d'école. ». Malgré le fait que certains sondés répondent OUI ou PEUT-ÊTRE à ces propositions car ils ont sans doute mal lu les consignes ou les phrases, on constate que la très grande majorité des personnes répondent NON. Ces phrases là ne sont donc pas ambiguës.

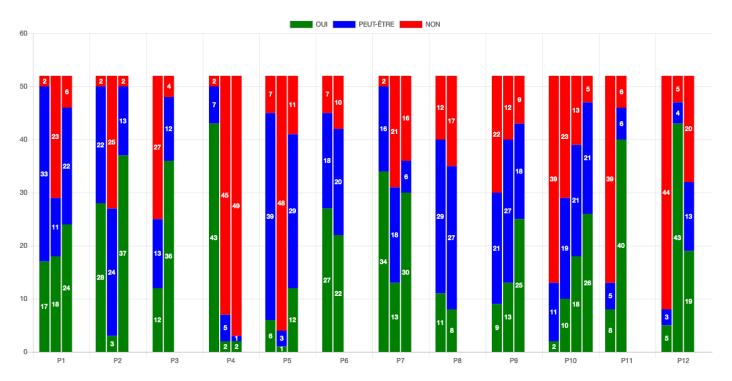


Fig. 9 – Nombre et type de note attribué aux propositions selon le critère : « Les propositions ont-elles le même sens que la phrase exemple ? »

Le temps de réflexion associé à la P4.2 est de 7,65 secondes, celui de la P4.3 est de 9,269 secondes et celui de la P5.2 est de 4,634 secondes. En moyenne, ces propositions notées NON sont évaluées une seconde plus lentement que les propositions notées OUI (voir Fig. 7). On peut interpréter cette différence de temps de réflexion par le fait que certaines propositions sont tellement absurdes et que le sondé se sent obligé de relire les phrases avant de finalement répondre NON car cela n'a effectivement aucun sens.

4.3.3 Catégorie n°3 - « PEUT-ÊTRE » majoritaire

Certaines propositions ont été évaluées majoritairement comme étant ambiguës si bien que les sondés n'ont pas pu dire OUI ou NON, ils ont donc répondu PEUT-ÊTRE. C'est le cas pour la phrase 8 « Des crayons gratuits seront distribués aux adhérents. » et les propositions P8.1 « Chaque adhérent aura droit à maximum un crayon. » et P8.2 « Les adhérents auront droit à autant de crayons qu'ils le souhaitent. ». La phrase exemple 8 ne permet pas de conclure sur le nombre maximum de crayons auquel un adhérent a droit. Pour la phrase 5 « Les enfants ont reçu un cadeau lors du dernier jour d'école. » et les propositions P5.1 « Chaque enfant a reçu un cadeau différent. » et P5.3 « Tous les enfants ont reçu le même cadeau. », les sondés ont préféré répondre PEUT-ÊTRE pour les mêmes raisons. Cependant, pour la P5.3, une dizaine de personnes sont convaincues que OUI et une autre dizaine que NON. On peut donc qualifier ces phrases d'ambiguës mais les sondés en ont majoritairement conscience car ils répondent PEUT-ÊTRE.

Comme on peut le constater avec le graphe de la Fig. 11, le temps de réflexion associé au propositions P8.1, P8.2, P5.1 et P5.3 sont respectivement de 7,33 secondes, 7,894 secondes, 5,523 seconde et 4,266 secondes. D'après la Fig. 7, lorsque le sondé prend en moyenne 8,853 secondes de réflexion avant de répondre PEUT-ÊTRE, cela ne peut qu'appuyer l'ambiguïté de la phrase. On notera également que la proposition P9.2 « Tous les musées de la ville ferment le ou les même jours. » de la phrase 9 « Tous les musées de la ville ferment certains jours de la semaine. » atteint un temps de réponse moyen de 21.043 secondes et qu'il s'agit d'une proposition majoritairement notée PEUT-ÊTRE. Cela dénote que la proposition P9.2 soulevait elle-même une ambiguïté en plus de celle apportée par la phrase exemple 9. Nous en arrivons donc à la conclusion que cette proposition a mal été choisie et nous la corrigerons dans la prochaine version du questionnaire.

	P1			P2	2	F	23		P4			P5		Р	6		P7	,	Р	8		P9)		P	10		P :	11]	P12	2
1	2	3	1	2	3	1	2	1	2	3	1	2	3	1	2	1	2	3	1	2	1	2	3	1	2	3	4	1	2	1	2	3
33	23	24	2 8	25	97	27	26	49	45	49	20	10	ാവ	27	22	34	21		29	97	22	27	25	39	23	21	26		40	44	19	2 0
99		22	22	24	٦ <i>١</i>	13		40	40	49	ฮฮ	40	29		20		18		29	21	21	21	18		19	16	21		40	44	43	19

	OUI majoritaire	Proposition vraie selon la majorité des sondés
	NON majoritaire	Proposition fausse ou incohérente selon la majorité des sondés
	PEUT-ÊTRE majoritaire	Proposition ambiguë mais le sondé en a conscience
	OUI et PEUT-ÊTRE majoritaires	Proposition ambiguë mais plutôt vraie
	NON et PEUT-ÊTRE majoritaires	Proposition ambiguë mais plutôt fausse
	OUI et NON majoritaires	Proposition ambiguë et interprétée différemment selon les sondés

Fig. 10 – Interprétation et simplification de la Fig. 9 avec la mise en avant de la/des note(s) majoritaire(s) pour chaque proposition

4.3.4 Catégorie n°4 - « OUI » et « PEUT-ÊTRE » majoritaires

Les propositions répondant à ce critère sont considérées comme ambiguës en pleine conscience par certains sondés (ceux qui répondent PEUT-ÊTRE) et sont vraies pour d'autres (ceux qui répondent OUI). Certaines personnes y voient donc différents sens possibles alors que d'autres sont convaincues que c'est vrai et n'y voient pas une potentielle ambiguïté. Parmi ces propositions, on peut citer la P6.1 « Certaines entreprises vont faire faillite. » et la P6.2 « Toutes les entreprises vont faire faillite. » de la phrase 6 « Les entreprises vont faire faillite. ».

4.3.5 Catégorie n°5 - « NON » et « PEUT-ÊTRE » majoritaires

A l'inverse, certaines propositions comme la P10.2 « Tous les étudiants cherchent un travail différent. » de la phrase 10 « Des étudiants cherchent du travail. » ont un nombre de NON et de PEUT-ÊTRE élevé. Certains sondés voient l'ambiguïté de « Des » alors que d'autres sont convaincus que « Des » veut dire « Tous » dans ce cas là.

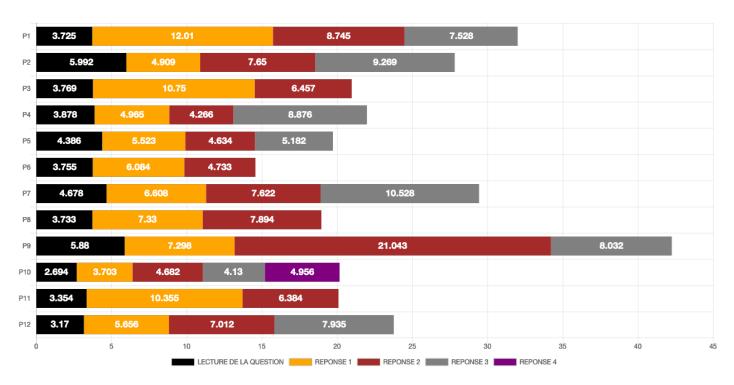


Fig. 11 – Temps de réflexion moyen par phrase (en secondes)

4.3.6 Catégorie nº6 - « OUI » et « NON » majoritaires

Sur les 33 propositions du questionnaire, une seule répond à ce critère. Il s'agit de la P1.2 « Un gardien différent doit surveiller la salle qui lui a été assignée. » de la phrase 1 « Un gardien surveille chaque salle du musée. ». Cette proposition est interprétée différemment selon les sondés. Ceux qui répondent OUI considèrent que chaque gardien surveille une salle différente et que celle-ci lui a été assignée. Ceux qui répondent NON considèrent que surveiller « chaque salle » ne restreint pas la surveillance qu'à une seule. Il n'y a pas de bonne ou de mauvaise réponse ici et comme dans tout le questionnaire. Dans ce cas là cependant, il y a une grosse différence d'interprétation sans que le sondé en ai conscience. Le temps de réflexion associé à cette proposition est de 8,745 secondes, ce qui est dans la moyenne des phrases à laquelle les sondés répondent majoritairement PEUT-ÊTRE. Avant de répondre à cette question la personne hésite longtemps sans pour autant répondre PEUT-ÊTRE. Cette phrase est donc très ambiguë mais le sondé donne plutôt un avis tranché à la fin.

4.4 Interprétation de « les » et « des »

Quelques données à propos de l'interprétation des mots « les » et « des » par les sondés sont rassemblées dans la Fig. 12. On constate que « les » est interprété comme « tous/toutes » ou « certains/certaines » alors que « des » est seulement interprété comme « certains/certaines ». En effet, « L'ensemble F contient des éléments de l'ensemble E. » sous-entend que l'on ne fait pas référence à tous les éléments de l'ensemble E, mais qu'il peut y avoir au moins un élément de E qui n'est pas dans F.

Mot	Interprétation	Note majoritaire	Phrases
	Chaque		P5.1
Les	Tous/Toutes		P5.3, P6.2
	Certains/Certaines		P6.1
Des	Tous/Toutes		P10.1, P10.2
Des	Certains/Certaines		P10.3, P10.4

Fig. 12 – Interprétation de « les » et « des » par les sondés

Il est difficile de tirer d'autres conclusions sur l'interprétation d'un mot en particulier avec les phrases que nous avons choisies. En effet, le fait de placer plusieurs quantificateurs dans la même phrase rend l'analyse des notations difficiles car on ne sait pas quel(s) quantificateur(s) a/ont fait pencher la balance lorsque le sondé a noté la phrase en question.

Cinquième partie

Perspectives pour le second semestre

1 Diversité des phrases

Afin d'améliorer le questionnaire, nous allons construire automatiquement des phrases ambiguës pour apporter des variations dans le questionnaire. L'idée est de garder quasiment les mêmes 12 squelettes de phrases que l'on a étudié ce semestre et de varier les noms, les verbes, les compléments, les temps des phrases et les quantificateurs. Un des points critique est l'accord des mots de la phrase en fonction des temps et du contexte. Voici un exemple avec la phrase n°7 « Un cercle touche chacun des carrés. ». On peut faire varier ce qui a été évoqué pour générer des phrases avec le même squelette (voir FIG. 13).

Préposition 1	Nom 1	Verbe	Quantificateur	Préposition 2	Nom 2		
Un	cercle	touche	chacun	des	carrés		
Un	Un cercle		tous	les	carrés		
Un	cercle	touche	$un\ seul$	des	carrés		
Un	cercle	touchera	chacun	des	carrés		
Des	cercles	touchent	chacun	des	carrés		
Un	triangle	touche	chacun	des	carrés		
Un	chasseur	tire sur	chacun	des	canards		

Fig. 13 – Exemple de variations avec la phrase n°7 « Un cercle touche chacun des carrés »

Afin de réaliser cette génération automatique de phrases, nous comptons rassembler les mots par champ lexical (voir Fig. 14). On peut par exemple former les phrases: « Un footballeur fait une passe à chacun de ses coéquipiers. », « Les ailiers font des passes à leurs attaquants. », « La patronne salue tous ses collègues. » ou « Les ouvriers saluent chacun le président. ».

Groupe 1 (touche)	Groupe 2 (tire sur)	Groupe 3 (fait une passe à)	Groupe 4 (salue)
cercle	chasseur	footballeur	collègue
triangle	canard	handballeur	patron
carré	sanglier	basketteur	patronne
losange	lapin	attaquant	ouvrier
rectangle	chevreuil	ailier	président
droite	lièvre	coéquipier	présidente
segment	faisan	gardien	secrétaire

Fig. 14 – Exemple de groupes de mots rassemblés par champ lexical et en lien avec la Fig. 13

2 Autres améliorations

Nous n'avons pas eu le temps de faire l'analyse en fonction de si le français est la langue maternelle du sondé. Il faut également pouvoir diffuser le questionnaire plus largement et à des personnes ayant des secteurs d'activité ou d'étude plus variés. En récoltant ces données, on pourrait alors potentiellement mettre en relation les préférences des locuteurs avec leur secteur d'activité ou d'étude. Enfin, concernant le questionnaire actuel, le renseignement de l'adresse mail n'est pas opérationnel dû à des problèmes de PHP qui n'ont pas pu être résolus. Tout cela peut donc être amélioré dans la prochaine version du questionnaire au second semestre.

Sixième partie

Annexe

1 Compte rendu des réunions

1.1 Lundi 14/10/19

Mettre les phrases manuellement (faire la génération automatique plus tard, afin de garder assez de temps pour la diffusion et l'analyse statistique).

Questionnaire:

- La langue française est-elle votre langue maternelle?
- Quel est votre âge?
- Quel est votre domaine d'étude (et/ou secteur d'activité)?

Fin décembre:

- Avoir des pages de questions faites à la main
- Avoir une base de données
- Avoir un timer (chronométrer le temps de réponse)

Après Noël: Améliorer

- Avoir un vrai accès au site (si on ne l'avait pas avant)
- Génération de phrases/questions

Site et Questionnaire:

- Faire la demande pour avoir un serveur accessible depuis l'extérieur
- Des phrases avec 2 quantificateurs:
 - ?? (Pour tout)
 - Exemples: Tous, Chaque, Les, Tous les, ...
 - Existentiel (Il existe)
 - Exemples: Un, Certains, Quelques, ...
 - Et permuter les quantificateurs de places
- Il faut des phrases ambiguës

Exemples de phrases:

- Les enfants prendront une pizza
- Est-ce qu'elle a des enfants? Oui, un. (On peut vouloir dire 'Pour tout', 'Il existe', '1')

Voir l'article d'Arthur Capetier sur ses expériences audios.

1.2 Lundi 04/11/19

Changement sur les phrase

- Mettre la 11ème phrase avec Tous les touristes
- Mettre la 10ème phrase *Un travail*, la phrase n'est pas très ambiguë.

Écrire des phrases

- Faire varier Tous les avec chaque ect...
- exemple: Les mêmes phrase: *Un étudiant doit avoir sa carte étudiant semble* être équivalent à *Tous les etudiants*

Utiliser des quantificateurs comme UN, AU MOINS... Et des quantificateurs universels TOUS LES, UN, LE.

1.3 Lundi 18/11/19

Avancement du site Web

- Mettre la première question (sur les vidéos surveillences) en dernier. En effet cela semble être une question qui n'a pas beaucoup d'ambiguïté en fonction de l'âge.
- Ajouter les sections Droit, Santé, Science Humaine, Science, Lettre, Science Eco.
- Stocker l'ordre des réponses et des questions des utilisateurs et les utiliser pour l'étude statistique.
- Mettre le timer pour chronométrer les questions.
- Génération attention des mots
- Attention sur la diffusion des résultats, ne pas parler du résultat des autres mais les résultats par rapport aux autres.

Rédaction du compte rendu

- Commencer la bibliographie
- Répartition des tâches, chacun doit prendre un partie et rédiger un plan détaillé.
- Déroulé du projet, organisation, mettre les statistiques à la fin du projet sur les données saisies du premier site
- Ajout éventuel d'un diagramme de Gantt.

1.4 Lundi 02/12/19

Regarder si le choix des mots pour symboliser un quantificateur influence le sens que comprend le locuteur.

1.5 Jeudi 18/12/19

Pour le rapport:

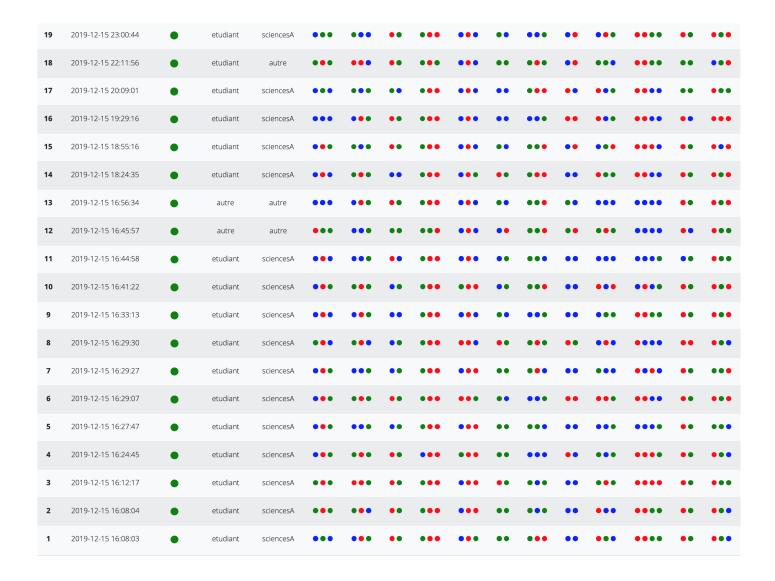
- Mettre un diagramme de Gantt
- Parler des articles que nous avons lu et analysé

2 Table Utilisateur - Schéma de la base de données

▼	# o1 : int(4)	# o11_1 : int(4)	# t9_3 : decimal(8,3)
<pre>@ id : int(11)</pre>	# o2 : int(4)	# o11_2 : int(4)	# t10_1 : decimal(8,3)
dateEnvoi : datetime datetime	# o3 : int(4)	# o12_1 : int(4)	# t10_2 : decimal(8,3)
langueM : varchar(40)	# o4 : int(4)	# o12_2 : int(4)	# t10_3 : decimal(8,3)
statut : varchar(40)	# o5 : int(4)	# o12_3 : int(4)	# t10_4 : decimal(8,3)
secteur : varchar(40)	# o6 : int(4)	# t1 : decimal(8,3)	# t11_1 : decimal(8,3)
email : varchar(40)	# o7 : int(4)	# t2 : decimal(8,3)	# t11_2 : decimal(8,3)
1_1 : varchar(1)	# o8 : int(4)	# t3 : decimal(8,3)	# t12_1 : decimal(8,3)
1_2 : varchar(1)	# o9 : int(4)	# t4 : decimal(8,3)	# t12_2 : decimal(8,3)
1_3 : varchar(1)	# o10 : int(4)	# t5 : decimal(8,3)	# t12_3 : decimal(8,3)
2_1 : varchar(1)	# o11 : int(4)	# t6 : decimal(8,3)	# tq1 : decimal(8,3)
2_2 : varchar(1)	# o12 : int(4)	# t7 : decimal(8,3)	# tq2 : decimal(8,3)
2_3 : varchar(1)	# o1_1 : int(4)	# t8 : decimal(8,3)	# tq3 : decimal(8,3)
3_1 : varchar(1)	# o1_2 : int(4)	# t9 : decimal(8,3)	# tq4 : decimal(8,3)
3_2 : varchar(1)	# o1_3 : int(4)	# t10 : decimal(8,3)	# tq5 : decimal(8,3)
4_1 : varchar(1)	# o2_1 : int(4)	# t11 : decimal(8,3)	# tq6 : decimal(8,3)
4_2 : varchar(1)	# o2_2 : int(4)	# t12 : decimal(8,3)	# tq7 : decimal(8,3)
4_3 : varchar(1)	# o2_3 : int(4)	# t1_1 : decimal(8,3)	# tq8 : decimal(8,3)
5_1 : varchar(1)	# o3_1 : int(4)	# t1_2 : decimal(8,3)	# tq9 : decimal(8,3)
5_2 : varchar(1)	# o3_2 : int(4)	# t1_3 : decimal(8,3)	# tq10 : decimal(8,3)
5_3 : varchar(1)	# o4_1 : int(4)	# t2_1 : decimal(8,3)	# tq11 : decimal(8,3)
6_1 : varchar(1)	# o4_2 : int(4)	# t2_2 : decimal(8,3)	# tq12 : decimal(8,3)
6_2 : varchar(1)	# o4_3 : int(4)	# t2_3 : decimal(8,3)	# tr1 : decimal(8,3)
7_1 : varchar(1)	# o5_1 : int(4)	# t3_1 : decimal(8,3)	# tr2 : decimal(8,3)
7_2 : varchar(1)	# o5_2 : int(4)	# t3_2 : decimal(8,3)	# tr3 : decimal(8,3)
7_3 : varchar(1)	# o5_3 : int(4)	# t4_1 : decimal(8,3)	# tr4 : decimal(8,3)
8_1 : varchar(1)	# o6_1 : int(4)	# t4_2 : decimal(8,3)	# tr5 : decimal(8,3)
8_2 : varchar(1)	# o6_2 : int(4)	# t4_3 : decimal(8,3)	# tr6 : decimal(8,3)
<pre>9_1 : varchar(1)</pre>	# o7_1 : int(4)	# t5_1 : decimal(8,3)	# tr7 : decimal(8,3)
<pre>9_2 : varchar(1)</pre>	# o7_2 : int(4)	# t5_2 : decimal(8,3)	# tr8 : decimal(8,3)
9_3 : varchar(1)	# o7_3 : int(4)	# t5_3 : decimal(8,3)	# tr9 : decimal(8,3)
10_1 : varchar(1)	# o8_1 : int(4)	# t6_1 : decimal(8,3)	# tr10 : decimal(8,3)
10_2 : varchar(1)	# o8_2 : int(4)	# t6_2 : decimal(8,3)	# tr11 : decimal(8,3)
10_3 : varchar(1)	# o9_1 : int(4)	# t7_1 : decimal(8,3)	# tr12 : decimal(8,3)
<pre>10_4 : varchar(1)</pre>	# o9_2 : int(4)	# t7_2 : decimal(8,3)	
<pre>11_1 : varchar(1)</pre>	# o9_3 : int(4)	# t7_3 : decimal(8,3)	
<pre>11_2 : varchar(1)</pre>	# o10_1 : int(4)	# t8_1 : decimal(8,3)	
<pre>12_1 : varchar(1)</pre>	# o10_2 : int(4)	# t8_2 : decimal(8,3)	
<pre>12_2 : varchar(1)</pre>	# o10_3 : int(4)	# t9_1 : decimal(8,3)	
12_3 : varchar(1)	# o10_4 : int(4)	# t9_2 : decimal(8,3)	

3 Données récoltées

#	Date d'envoi	Français	Statut	Secteur	P1	P2	Р3	P4	P5	P6	P7	P8	P9	P10	P11	P12
52	2019-12-23 18:38:04	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
51	2019-12-20 00:17:51	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
50	2019-12-19 17:36:59	•	etudiant	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
49	2019-12-19 17:36:32	•	enseignant	economie	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
48	2019-12-19 11:32:00	•	etudiant	economie	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
47	2019-12-18 22:16:16	•	etudiant	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
46	2019-12-18 19:18:33	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
45	2019-12-18 18:46:22	•	autre	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
44	2019-12-18 18:44:55	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
43	2019-12-18 17:51:16	•	autre	economie	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
42	2019-12-18 17:49:45	•	autre	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
41	2019-12-18 17:14:43	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
40	2019-12-18 16:24:27	•	autre	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
39	2019-12-18 16:21:40	•	etudiant	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
38	2019-12-18 16:18:41	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
37	2019-12-18 16:10:08	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
36	2019-12-18 16:06:40	•	autre	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
35	2019-12-18 15:52:27	•	etudiant	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
34	2019-12-18 15:48:09	•	autre	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
33	2019-12-18 15:30:07	•	autre	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
32	2019-12-18 15:29:50	•	etudiant	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
31	2019-12-18 15:10:29	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
30	2019-12-18 15:05:01	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
29	2019-12-18 15:03:00	•	autre	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
28	2019-12-18 14:57:58	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
27	2019-12-18 14:43:23	•	etudiant	sciencesH	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
26	2019-12-17 22:22:18	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
25	2019-12-17 13:42:21	•	autre	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
24	2019-12-17 13:09:47	•	enseignant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
23	2019-12-16 22:45:10	•	etudiant	autre	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
22	2019-12-16 12:42:01	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
21	2019-12-16 09:30:24	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••
20	2019-12-16 08:18:44	•	etudiant	sciencesA	•••	•••	••	•••	•••	••	•••	••	•••	••••	••	•••



Références

- [1] BRISSON C. Plural, « all », and the nonuniformity of collective predication. *Linguistics and Philosophy*, pages 129–184, 2013.
- [2] MARI A. et RETORE C. Conditions d'assertion de « chaque » et de « tout » et règles de déduction du quantificateur universel. *Travaux de linguistique*, pages 89–106, 2016.
- [3] DOWTY G. Distributive predicate and all. Collective predicate, 1987.
- [4] KLEIBER G. La quantification universelle en trio: « tous les », « chaque » et « tout ». Studii de lingvistica, page 139–157, 2011.