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

	and the same of th								
63	4 4 4 00	1	11111	90	00	00 00 00	6	10	11
	15.		:::			::			
	14		:::						
	7		:::			::			
	3		: : :			• • •			
	0		: : :			::			
	()		:::			::			
	, 0								
	20		:::			::			
	E		90						
	0		nnée			::			
			op			::			
	ail		cture et développement Hébergement du site web et dialogue avec la base de données Conception de la base de données Conception du formulaire			::			
	trav		la b			::	_		
	of a		avec			: :	stre		
	, 418		ens : :			::	me		
	76	lisé	lialo nées				d se		
	me	réa	don don				con		
	tiels	vail	web we de		(0)	litat	86		
	pro vers	tra	ppe site		laire	résu Itats	ır le		
	Dni Exis	np	rvelo t du de la du fo	89	rman	des	pou		
on	ion ion du	ion	t de	9800	n fo	e l'é	ves		
ucti	icat icat	ript	cture et développement Hébergement du site web et dialog Conception de la base de données Conception du formulaire	ns F	n d	ultats de l'étude Présentation des résultats Analyse des résultats	ecti	эхе	
Introduction	Présentation du problème Quantificateur Universell Quantificateur Existentiels Anni Sun Gestion du temps et répartition du travail	Description du travail réalisé	Structure et développement 3.1 Hébergement du site web e 3.2 Conception de la base de d 3.3 Conception du formulaire	Questions posées	Diffusion du formulaire	Résultats de l'étude 6.1 Présentation des r 6.2 Analyse des résult	Perspectives pour le second semestre	Annexe	Références
Int	F 6 6		Stru 3.1 3.2	Que	Diff	Rés 6.1	Pe	A	fére
	11 2 11	2	60	4	10	9	>	5	Ré

Introduction Première partie

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'ex-pression de la quantification » proposé par M. Christian Retord II enseignant-chercheur à la Faculté des Sciences et responsable de l'équipe TEXTEZ du LIRMM que nous remercions pour son aide et sa présence lors de nos réunions quinzomadaires

rentes personnes dans l'expression de la quantification. Notre formulaire est composé de 12 phrases (que l'on appellera phrases exemples) plus <mark>ou moins ambiguës a</mark>vec chacunes 2 à 4 propositions de phrases. Chacune de ces propositions de phrases doit être évaluée par l'utilh<u>s</u>eteur en fonction de si elle lui parait avoir le même sens Durant ce semestre, nous avons développé un formulaire web qui permet d'étudier les préférences de difféque la phrase exemple. L'utilisateur a trois choix de notations: « oui », « peut-être » et « non ». Le temps de réponse aux questions est chronométré de façon à pouvoir évaluer l'éventuelle hésitation de l'utilisateur lorsqu'il répond. Les ordres des phrases et des propositions de phrases sont aléatoirisés afin d'évaluer si cela joue un rôle dans le choix de<mark>r</mark>éponse. A la fin du questionnaire, on lui demande de renseigner si la français est sa langue maternelle, son statut et son secteur d'activité. 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 modèlé ^a Bootstrap. La programmation côté serveur est réalisée en PHP avec une base de données MySQL. La programmation côté client est réalisée

is expensed our certain(S) quelque(S) oles

Présentation du problème

donne & les delinisat 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 ambiguité. 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 5 771 - Ambiguïtés des terminaisons

represent tales

- Ambiguïtés de la grammaire et de la sémantique

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

Quantificateur Universel

- La quantification universelle peut signifier intuitivement * attribuer la propriété Q à toutes les entités d'un ensemble P ». 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
- Selon Brissonet Lasersohn, il n'est pas analysé comme un véritable quantificateur, mais plutôt comme la propriété.
- Le quantificateur universel <CHAQUE>: Contrairement à tout, ce quantificateur demande que chaque entité du domaine ait été observé isolement, il a une sémantique distributive et est malvenu dans des phrases génériques. Son emploi est plutôt descriptif P(x)->Q(x)

un «réducteur de marge» .Son emploi repose sur l'observation de l'existence d'une règle

2 Quantificateur Existentiels

Ale policie d'au moins une chose. En mathématiques : « ∃x ∈ R. * a lit * Il mointe a mathématiques : « ∃x ∈ E » se lit « Π existe au moins un x élément de l'ensemble E ».

parmis les quantificateurs existentiels on a:

tonnez de

sure un

- Certain: Certains est traité comme indéfini au sens strict dans de nombreuses approches, alors que plusieurs auteurs l'excluent de cette classe. Il semblerait que certains a en général une prédilection marquée pour les portées large et intermédiaire
- ou mans an - Un: Dans la langue française le sens du quantificateur un n'est jamais compris dans les phrases, sa compréhension est toujours associé au chiffre un. Son interprétation réelle équivaut à « au moins » dans les phrases dans lesquelles il est utilisé.

5. Graphème: lettre ou groupe de lettres transcrivant un phonème (élément sonore du langage parlé).

1. Site Web de M. Christian RETORÉ: https://www.irram.fr/ retore/ 2. TEXTE: Exploration et exploitation de données textuelles 3. LIRMM: Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier

4. Modèle: de l'anglais template

pas porvible: Sayet pues 1 get landvane dans

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'.

Faut noter qu'en français comme en logique, l'utilisation de tout éguivaut à celui de chaque mais la récipréque de l'utilisation de tout éguivaut a celui de chaque mais la récipréque des fencionne pas car dans la plupart des énoncés avec 'chaque', 'tout' induit une relation causale peu évidente ou même d'ailleurs quasi impossible.

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:

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

 $\forall i, \exists b, ramener(b,i)$

Les formules (I) et (2) sont sémantiquement différentes. L'énoncé est plus ambiguë quand le sens est obscur pour celui qu'il l'interpréte. Ce qui est tout à fait naturel.

Troisième partie

Gestion du temps et répartition du travail

pent fun ex

Quatrième partie

Description du travail réalisé

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

3 Structure et développement

3.1 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.

3.2 Conception de la base de données

La base de donnée est composée d'un unique table UTILISATEUR 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 réponses à chacune des 33 propositions. Le temps de réponse à chacune des 12 phrases.

- Le temps de lecture de chacune des 12 phrases.

- Le temps de réponse moyen aux propositions de chacune des 12 phrases.
- 3.3 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 milien d'étude et de jeur statut. Ces trois questions simples ont êté convenues car elles ne rentrent pas dans l'intimité des sondés et qu'elles permettent de les classe dans des catégories qui indiquent potentiellement une manière de raisonne

4 Questions posées

- Diffusion du formulaire 2
- Résultats de l'étude 9
- 6.1 Présentation des résultats
- Analyse des résultats

Cinquième partie

Perspectives pour le second semestre

Afin d'améliorer le questionnaire, nous allons construire automatiquement des phrases ambigués pour apporter des variations dans le questionnaire. L'idée est de garder 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 phrases « Un cercle touche chacun des carrés. ». On peut faire varier ce qui a été évoqué pour générer des phrases avec le même squelette.

Nom 2	carrés	carrés	carrés	carrés	carrés	carrés	canards	canards
Préposition 2	des	les	des	des	des	des	des	des
Quantificateur	chacun	tous	nn senl	chacun	chacun	chacun ,	chacun	chacun
Verbe	touche	touche	touche	touchera	touchent	touche	touche	tire sur
Nom 1	cercle	cercle	cercle	cercle	cercles	triangle	chasseur	chasseur
Préposition 1	Un	Un	Un	Un	Des	Un	Un	Un

日

Sixième partie Annexe 10

Références

[1] Donald E. Knuth. The Art of Computer Programming. Addison-Wesley, 2019.

metor ici

les arti ples ever leus.

que vous ever leus.

il pur librant mieux

il pur l'Itulis soient

ches sans le texte

11