



HAL
open science

Vers un outil sémantique d'autocodage qualitatif pour l'évaluation de l'acceptabilité des innovations

Doriane Simonnet

► **To cite this version:**

Doriane Simonnet. Vers un outil sémantique d'autocodage qualitatif pour l'évaluation de l'acceptabilité des innovations. Linguistique. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALL002 . tel-03689721

HAL Id: tel-03689721

<https://theses.hal.science/tel-03689721v1>

Submitted on 7 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Sciences du langage Spécialité Informatique et sciences du langage

Arrêté ministériel : 25 mai 2016

Présentée par

Doriane SIMONNET

Thèse dirigée par **Georges ANTONIADIS**, Université Grenoble Alpes

préparée au sein du **Laboratoire Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles**
dans **l'École Doctorale Langues, Littératures et Sciences Humaines**

Vers un outil sémantique d'autocodage qualitatif pour l'évaluation de l'acceptabilité des innovations

Towards a qualitative autocoding semantic tool for the evaluation of the acceptability of innovations

Thèse soutenue publiquement le **3 mars 2022**,
devant le jury composé de :

Monsieur Georges ANTONIADIS

PROFESSEUR DES UNIVERSITES EMERITE, Université Grenoble Alpes, Directeur de thèse

Monsieur Alain POLGUÈRE

PROFESSEUR DES UNIVERSITES, Université de Lorraine, Rapporteur

Monsieur Guy LAPALME

PROFESSEUR EMERITE, Université de Montréal, Rapporteur

Monsieur François PORTET

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes, Examineur

Madame Tita KYRIACOPOULOU

PROFESSEURE DES UNIVERSITES, Université Paris 12, Présidente

Monsieur Jean CAELEN

DOCTEUR EN SCIENCES, INSTITUT CARNOT , Examineur



RÉSUMÉ

Les techniques du Traitement Automatique des Langues (TAL) et les méthodes d'analyse qualitative des données textuelles entretiennent une certaine affinité épistémologique. Malgré cela, l'analyse qualitative ne profite pas pleinement des apports potentiels du TAL. En particulier, les travaux visant à une réelle automatisation du codage qualitatif des données restent somme toute assez rares. Cette thèse se donne pour ambition d'investiguer le potentiel de différentes techniques du TAL dans cet objectif et pour une tâche qui nécessite un certain degré d'expertise humaine. Elle vise à la création d'un outil utilisable dans un contexte industriel et pour une méthode d'analyse spécifique qui permet d'évaluer l'acceptabilité des innovations. Cette méthode mobilise une grille de 20 codes qui présentent une complexité sémantique plus élevée que ceux traditionnellement utilisés en analyse qualitative outillée.

Nous explorons les moyens de parvenir à effectuer cette tâche à travers une approche ascendante puis une approche descendante. Pour la première, nous réalisons une exploration lexicométrique sur un corpus de données d'études anciennes afin de définir le profil lexical des données attendues pour chaque code. Puis, nous traitons le problème qui nous est posé comme une tâche de classification en testant des classifieurs statistiques de plusieurs types. Nous investiguons également les possibilités offertes par la projection d'une ressource syntaxico-sémantique sur le corpus.

Nous suivons ensuite une approche descendante moins conventionnelle. Pour celle-ci, nous réalisons une modélisation experte du paradigme de l'entretien qualitatif d'évaluation de l'acceptabilité sous forme d'ontologie. Cette modélisation est assortie à un lexique construit de manière *ad hoc*. Nous proposons ainsi une architecture originale d'outil d'analyse sémantique dans laquelle les triplets de l'ontologie servent de support à l'interprétation et dont des sous-ensembles constituent des règles de classification. Nous obtenons un outil d'analyse hyperspécialisé dont les performances dépassent celles obtenues par le *machine learning* sur notre corpus d'entraînement. Cet outil est porté jusqu'à l'opérationnalisation, par son intégration dans une plateforme d'autocodage, en vue de la mise en place d'un processus d'apprentissage continu.

Mots-clefs : TAL, analyse sémantique, autocodage qualitatif, acceptabilité, ontologies

ABSTRACT

The techniques of Natural Language Processing (NLP) and the methods of qualitative analysis of textual data share epistemological similarities. However, qualitative analysis does not fully benefit from the potential of NLP. In particular, works aiming at an actual automation of qualitative data coding remain quite rare. This thesis aims at investigating the potential of different NLP techniques for this purpose and for a task that requires a certain degree of human expertise. It aims at creating a tool that can be used in an industrial context and for a specific analysis method that allows to evaluate the acceptability of innovations. This method uses a codebook composed of 20 codes that have a higher semantic complexity than those traditionally used in computer-aided qualitative analysis.

We explore ways to achieve this task through a bottom-up and then a top-down approach. For the former, we perform a lexicometric exploration on a corpus of old study data to define the lexical profile of the expected data for each code. Then, we treat the problem as a classification task by testing statistical classifiers of several types. We also investigate the possibilities offered by the projection of a syntaxico-semantic resource on the corpus.

We then follow a less conventional top-down approach. For this one, we carry out an expert-informed modelling, in the form of an ontology, of the acceptability evaluation qualitative interview paradigm. This modelling is matched with a lexicon built in an *ad hoc* way. We thus propose an original architecture for a semantic analysis tool in which the ontology's triples are used as a support for textual interpretation and which subsets constitute classification rules. We achieve building a hyperspecialized analysis tool which performances exceed those obtained by machine learning on our training corpus. This tool is carried to the point of operationalization, by being integrated into an autocoding platform, for the purpose of setting up a continuous learning process.

Key words: NLP, semantic analysis, qualitative autocoding, acceptability, ontologies

TABLE DES MATIÈRES

RÉSUMÉ	I
ABSTRACT	II
TABLE DES MATIÈRES	III
TABLES DES FIGURES, TABLEAUX ET ENCADRÉS	VI
Table des figures	vi
Table des tableaux.....	vi
Table des encadrés.....	viii
Table des figures en annexe.....	viii
Table des tableaux en annexe	ix
Table des encadrés en annexe	x
LEXIQUE	XI
NOTATIONS	XIII
INTRODUCTION	1
1. Contexte et demande de l'entreprise.....	1
2. Champs de recherche et positionnement	2
3. Approche et plan	4
PARTIE I – BESOINS INDUSTRIELS ET OUTILS EXISTANTS	6
CHAPITRE I : BESOINS APPLICATIFS	7
1. Innovation et acceptabilité : enjeux et définitions	8
2. Caractéristiques de la méthode CAUTIC®.....	11
3. Les Critères CAUTIC®.....	19
4. Nature des données.....	24
5. Conclusions : résumé et points saillants de la demande.....	31
CHAPITRE II : L'ANALYSE QUALITATIVE OUTILLÉE	34
6. Les analyses de données qualitatives et leurs besoins.....	34
7. Solutions pour outiller l'analyse qualitative.....	44
CHAPITRE III : POSITIONNEMENT ET APPROCHE	57
8. Originalité de la recherche	57
9. Approches ascendantes vs approches descendantes	81
PARTIE II – APPROCHE ASCENDANTE	85
CHAPITRE IV : CORPUS D'APPRENTISSAGE	86
1. Agrégation et structuration des données	86
2. Filtrage expert.....	92
CHAPITRE V : EXPÉRIMENTATIONS BASÉES SUR DES APPROCHES	
STATISTIQUES	106
3. Exploration textométrique	107
4. Les méthodes de classification de texte	121
5. Tests de classifieurs	129

CHAPITRE VI : EXPÉRIMENTATION SUR LE RECOURS À UNE RESSOURCE SYNTAXICO-SÉMANTIQUE	140
6. Les ressources sémantiques pour le français.....	141
7. Présentation de Verb ̂ net	144
8. Expérimentation.....	147
PARTIE III : APPROCHE DESCENDANTE	158
CHAPITRE VII : MODÉLISATION EXPERTE DES CRITÈRES CAUTIC®	159
1. Les ontologies comme outils d'analyse sémantique.....	160
2. Processus de modélisation experte.....	165
CHAPITRE VIII : PROJECTIONS DU MODÈLE EXPERT	188
3. Projection des classes sémantiques.....	188
4. Expérimentations sur des systèmes par patrons	193
5. Expérimentations sur des systèmes de règles basées sur les triplets de l'ontologie	200
CHAPITRE IX : VERS UN SYSTÈME EN APPRENTISSAGE CONTINU	212
6. Augmentation du lexique.....	212
7. Expérimentation en conditions réelles.....	219
8. Bilan et pistes d'amélioration.....	239
CONCLUSIONS ET PERSPECTIVES	242
1. Résumé et réponses aux questionnements.....	242
2. Perspectives	245
BIBLIOGRAPHIE	250
ANNEXES	265
ANNEXE 1 : INTERFACES ET ARCHITECTURES	266
1. Interface et architecture de <i>Cody</i> (Rietz, 2021)	266
2. Architecture du système proposé.....	268
3. Interface de la plateforme d'analyse de l'entreprise Ixiade	269
ANNEXE 2 : CORPUS	271
4. Volumétrie du corpus non filtré	271
5. Éléments chiffrés et illustratifs du processus de filtrage du corpus	272
ANNEXE 3 : ANALYSE LEXICOMÉTRIQUE	286
6. Exemple de données extraites	286
7. Analyse par catégories morphosyntaxiques.....	291
ANNEXE 4 : TESTS DE CLASSIFIEURS	292
8. Résultats pour les approches par sacs de mots.....	292
9. Résultats pour les approches par TF-IDF	295
10. Résultats pour les approches par <i>embedding</i>	298
11. Performances en fonction des méthodes de vectorisation et des algorithmes	300
12. Mise en parallèle des observations textométriques et des résultats des classifieurs.....	301
ANNEXE 5 : PROJECTION DE VERB̂NET	303
13. Corpus annoté.....	303
ANNEXE 6 : MODÉLISATION EXPERTE	312

14. Analyse de la collecte de verbatim types	312
15. Ontologie.....	317
16. Règles de classification.....	323
17. Augmentation du lexique	331
18. Expérimentation en conditions réelles.....	336

TABLES DES FIGURES, TABLEAUX ET ENCADRÉS

Table des figures

FIGURE 1 : A gauche : modèle UTAUT (figure de Venkatesh et al. (2003)) – A droite : modèle CAT (figure de Kulviwat et al. (2007))	10
FIGURE 2 : Déroulement de la méthode CAUTIC®.....	14
FIGURE 3 : Représentation schématique de la méthode d'analyse CAUTIC®	17
FIGURE 4 : Framework de l'analyse textuelle selon Lacity & Janson (1994, p.140)	35
FIGURE 5 : Méthodes pour l'analyse des données qualitatives – Figure de Voynnet Fourboul & Point (2001, p. 1448)	37
FIGURE 6 : Etapes des analyses de contenu – figure de Dany (2016), à partir des travaux de Bardin (1977), Bauer (2012), Robert & Bouillaguet (2007) et Schreier (2014)	38
FIGURE 7 : Fichiers des données d'études dans les archives d'Ixiade.....	88
FIGURE 8 : Membres et rôles thématiques de la classe beg-58.2 de Verbænet.....	145
FIGURE 9 : Exemple de frame de la classe beg.....	146
FIGURE 10 : Exemples de classes gravitant autour de la classe User.....	173
FIGURE 11 : Exemple de modélisation	175
Figure 12 : Lexique pour les catégories lexico-sémantiques associées à la classe EaseOfUsePerception.....	183
FIGURE 13 : Lexique pour les catégories lexico-sémantiques associées à la classe AdequacyPerception	183
FIGURE 14 : Classes de l'ontologie	186
FIGURE 15 : Fonctionnement du système proposé	204
FIGURE 16 : Exemple de tableau conceptuel	208
FIGURE 17 : Corrections effectuées par Critères.....	228

Table des tableaux

TABLEAU 1 : Les quatre Niveaux des Critères CAUTIC®.....	19
TABLEAU 2 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 1– « Assimilation aux savoir-faire »	20
TABLEAU 3 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 2 – « Association aux pratiques »	21
TABLEAU 4 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 3 – « Appropriation à l'identité socio-culturelle ».....	22
TABLEAU 5 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 4 – « Adaptation à l'environnement »	23
TABLEAU 6 : « Verbosité et représentation de l'interprétation » dans les logiciels d'aides à l'analyse qualitative - Tableau de Lejeune (2017).....	47
TABLEAU 7 : Techniques du TAL pour l'analyse des données qualitatives – Tableau de Kaufmann et al (2020, p. 754).....	50

Tableau 8 : Ambitions, données, techniques et résultats de quatre travaux d'autocodage (Crowston et al., 2012; Kaufmann et al., 2020; Marathe & Toyama, 2018; Rietz, 2021)	61
TABLEAU 9 : Retours d'expériences et recommandations pour l'adéquation entre pratiques du codage en Sciences Sociales et outils	76
TABLEAU 10 : Résultats du filtrage du corpus par accord inter-juge	95
TABLEAU 11 : Répartition des verbatim par études et domaines, et par type analyse/illustratif dans le corpus final filtré.....	97
TABLEAU 12 : Répartition des verbatim par critère et niveau dans le corpus final filtré.....	97
TABLEAU 13 : Moyenne d'âge et répartition des participants en genres, statuts professionnels/particuliers dans le corpus final filtré.....	97
TABLEAU 14 : Score Alpha de Krippendorff pour l'accord inter-juge obtenu sur l'annotation des verbatim issus des archives dans le processus de filtrage du corpus	100
TABLEAU 15 : Synthèse de l'analyse textométrique pour les Critères du Niveau 1.....	112
TABLEAU 16 : Synthèse de l'analyse textométrique pour les Critères du Niveau 2.....	113
TABLEAU 17 : Synthèse de l'analyse textométrique pour les Critères du Niveau 3.....	114
TABLEAU 18 : Synthèse de l'analyse textométrique pour les Critères du Niveau 4.....	115
TABLEAU 19 : Relations lexicales entre les intitulés des Critères et les lexies du corpus.....	117
TABLEAU 20 : Avantages et inconvénients de différentes approches de vectorisation – Reproduction partielle et traduction du tableau de Kowsari et al. (2019, pp. 48-49).....	125
TABLEAU 21 : Combinatoire de prétraitements testés pour la classification de texte	131
TABLEAU 22 : Meilleurs scores par corpus obtenus par ML et configurations correspondantes.....	133
Afin d'observer plus en détails la distribution des résultats entre les classes, nous nous intéressons aux matrices de confusion obtenues par les configurations ayant les meilleurs f-scores, que nous présentons dans les TABLEAUX 23 et 24. Pour ce qui concerne la tâche monolabel, nous constatons principalement une tendance à la sur-classification dans les Critères 1.1, 1.4, 2.3, 2.4 et 3.1, ainsi qu'à la classe Inclassifiable pour le corpus qui la contient. Ces faibles précisions sont corrélées à une surreprésentation de ces Critères dans les données d'apprentissage. En outre, nous observons des résultats	
TABLEAU 23 : Matrices de confusion pour les meilleurs scores en monolabel	134
TABLEAU 24 : Matrices de confusion pour le meilleur score en multilabel.....	135
TABLEAU 25 : Couverture de l'annotation avec Verbønet.....	153
TABLEAU 26 : Exemple de règle de classification produite sur la base d'un verbatim stéréotypique.....	176

TABLEAU 27 : Résultats des tests de paramètres algorithmiques pour le système de règles basé sur les triplets de l'ontologie.....	203
TABLEAU 28 : Evaluation du potentiel de couverture des trois ressources lexicales envisagées pour augmenter le lexique.....	217
TABLEAU 29 : Evaluation du potentiel de précision des trois ressources lexicales envisagées pour augmenter le lexique.....	217
TABLEAU 30 : Volumétrie de l'augmentation du lexique.....	218
TABLEAU 31 : Répartition textes/analystes pour l'expérimentation en conditions réelles.....	224
TABLEAU 32 : Résultats pour l'expérience en conditions réelles.....	225

Table des encadrés

ENCADRÉ 1 : Critères du Niveau 1 – « Assimilation aux savoir-faire ».....	20
ENCADRÉ 2 : Critères du Niveau 2 – « Association avec les pratiques ».....	21
ENCADRÉ 3 : Critères du Niveau 3 – « Appropriation à l'identité socio-professionnelle ».....	22
ENCADRÉ 4 : Critères du Niveau 4 – « Adaptation à l'environnement ».....	23
ENCADRÉ 5 : Exemple de règle utilisée dans le système de Crowston et al. (2012).....	62
ENCADRÉ 6 : Extrait du corpus de verbatim d'études passées.....	90
ENCADRÉ 7 : Exemple de verbatim étiqueté sémantiquement.....	191

Table des figures en annexe

FIGURE EN ANNEXE 1 : Architecture de Cody.....	266
FIGURE EN ANNEXE 2 : Vue d'ensemble de l'interface de Cody.....	266
FIGURE EN ANNEXE 3 : Focalisation sur les modules de codage de Cody.....	267
FIGURE EN ANNEXE 4 : Architecture proposée.....	268
FIGURE EN ANNEXE 5 : Interface plateforme d'analyse Ixiade – Codage sur texte entier.....	269
FIGURE EN ANNEXE 6 : Interface plateforme d'analyse Ixiade – Verbatim filtrés.....	270
FIGURE EN ANNEXE 7 : Décisions prises à la suite du filtrage par accord inter-juge par rapport au nombre de mots par verbatim.....	272
FIGURE EN ANNEXE 8 : Répartition des syntagmes les plus discriminants par catégories morphosyntaxiques – Nombre d'occurrences.....	291
FIGURE EN ANNEXE 9 : Répartition des syntagmes les plus discriminants par catégories morphosyntaxiques – Nombre de lexies.....	291
FIGURE EN ANNEXE 10 : F-scores obtenus en fonction de la distribution des classes dans le corpus et de la variabilité lexicale estimée par classe – Critères seuls.....	301
FIGURE EN ANNEXE 11 : F-scores obtenus en fonction de la distribution des classes dans le corpus et de la variabilité lexicale estimée par classe – Critères + Inclassifiable.....	301
FIGURE EN ANNEXE 12 : F-scores obtenus en fonction de la distribution des classes dans le corpus et de la variabilité lexicale estimée par classe – Meilleure configuration W2V.....	302
FIGURE EN ANNEXE 13 : Exemple d'étiquetage morphosyntaxique du corpus pour l'expérience Verbønet.....	304
FIGURE EN ANNEXE 14 : Extrait du corpus étiqueté avec Verbønet.....	306

FIGURE EN ANNEXE 15 : Répartition des classes de Verbønet dans les différents Niveaux	307
FIGURE EN ANNEXE 16 : Répartition des classes de Verbønet dans les différents Critères du Niveau 1	308
FIGURE EN ANNEXE 17 : Répartition des classes de Verbønet dans les différents Critères du Niveau 2	309
FIGURE EN ANNEXE 18 : Répartition des classes de Verbønet dans les différents Critères du Niveau 3	310
FIGURE EN ANNEXE 19 : Répartition des classes de Verbønet dans les différents Critères du Niveau 4	311
FIGURE EN ANNEXE 20 : Exemples de classes gravitant autour de la classe Concept	317
FIGURE EN ANNEXE 21 : Concept testé dans le cadre de l'étude de l'expérimentation en conditions réelles	336
FIGURE EN ANNEXE 22 : Corrections par Critère par textes	340

Table des tableaux en annexe

TABLEAU EN ANNEXE 1 : Répartition des verbatim du corpus non filtré par domaines	271
TABLEAU EN ANNEXE 2 : Répartition des verbatim du corpus non filtré par Niveaux et/ou Critères attribués	271
TABLEAU EN ANNEXE 3 : Participants des études du corpus non filtré : ratios hommes/femmes, professionnels/particuliers et moyenne d'âge	272
TABLEAU EN ANNEXE 4 : Matrice de confusion entre les classifications des verbatim en Niveaux dans les archives et celles obtenues par évaluation inter-juge	272
TABLEAU EN ANNEXE 5 : Matrice de confusion entre les classifications des verbatim en Critères dans les archives et celles obtenues par évaluation inter-juge	273
TABLEAU EN ANNEXE 6 : Matrice de confusion de l'accord inter-juge sur les Niveaux dans le processus de filtrage du corpus	274
TABLEAU EN ANNEXE 7 : Matrice de confusion de l'accord inter-juge sur les Critères dans le processus de filtrage du corpus	274
TABLEAU EN ANNEXE 8 : Exemples commentés de verbatim qui suscitent un fort accord inter-juge dans le processus de filtrage	277
TABLEAU EN ANNEXE 9 : Exemples commentés de verbatim étant unanimement considérés comme inclassifiables pendant le processus de filtrage	280
TABLEAU EN ANNEXE 10 : Exemples commentés de verbatim qui provoquent un désaccord important pendant le processus de filtrage	285
TABLEAU EN ANNEXE 11 : Exemple de n-grams extraits par analyse textométrique	290
TABLEAU EN ANNEXE 12 : Résultats pour l'approche par sacs de mots	294
TABLEAU EN ANNEXE 13 : Résultats pour l'approche TF-IDF	297
TABLEAU EN ANNEXE 14 : Résultats pour l'approche par embedding	299
TABLEAU EN ANNEXE 15 : Performances par approches de vectorisation	300
TABLEAU EN ANNEXE 16 : Performances par algorithmes	300
TABLEAU EN ANNEXE 17 : Analyse des verbatim types	316

TABLEAU EN ANNEXE 18 : Exemples de règles construites sur des verbatim stéréotypiques.....	326
TABLEAU EN ANNEXE 19 : Exemples de règles construites sur des verbatim réels	330
TABLEAU EN ANNEXE 20 : UL extraites à partir de WoNef.....	332
TABLEAU EN ANNEXE 21 : UL extraites à partir de WOLF	333
TABLEAU EN ANNEXE 22 : UL extraites à partir de JeuxDeMots	335
TABLEAU EN ANNEXE 23 : Questions posées aux participants lors de l'étude test.....	337
TABLEAU EN ANNEXE 24 : Exemples de classifications réussies.....	342
TABLEAU EN ANNEXE 25 : Exemples de classifications erronées	344

Table des encadrés en annexe

ENCADRÉ EN ANNEXE 1 : Exemple de création d'une règle à partir d'un verbatim	319
ENCADRÉ EN ANNEXE 2 : Exemples de sous-classes de ReactionToConcept.....	320
ENCADRÉ EN ANNEXE 3 : Exemples de classes relatives aux caractéristiques du Concept	320
ENCADRÉ EN ANNEXE 4 : Exemples de classes pouvant être ou non relatives au Concept.....	321
ENCADRÉ EN ANNEXE 5 : Exemples de classes relatives à l'utilisateur	321
ENCADRÉ EN ANNEXE 6 : Exemples de classes générales non intrinsèquement liées à l'utilisateur ou au Concept.....	322
ENCADRÉ EN ANNEXE 7 : Exemples de classes liées à la présentation du Concept.....	322
ENCADRÉ EN ANNEXE 8 : Extraits de posts collectés pour l'étude test.....	339

LEXIQUE

Ce lexique précise les acceptions utilisées pour certains termes tout au long de cette thèse ainsi que la signification de certains acronymes.

ADQ : Analyse des Données Qualitatives.

ADT : Analyse des Données Textuelles.

BERT : *Bidirectional Encoder Representations from Transformers*.

BOW : *Bag-of-words*.

CAQDAS : *Computer-Aided Qualitative Data Analysis Software*.

CAUTIC® : Conception Assistée par l'Usage pour les Technologies, l'Innovation et le Changement, nom de la méthode d'analyse que nos travaux visent à automatiser.

CIFRE : Convention Industrielle de Formation par la Recherche.

Concept : Avec une majuscule, le mot *Concept(s)* renverra aux concepts innovants (produits ou services) qui sont développés par les clients d'Ixiade et qui sont testés lors des études.

Critère : Avec une majuscule, le mot *Critère(s)* renverra aux 20 Critères de la méthode CAUTIC® qui constituent les classes de la tâche de classification au cœur de nos travaux.

DELAC : Dictionnaire Electronique des Mots Composés ([Silberztein, 1990](#)).

KNN : *k-nearest neighbors*.

KSE : *Knowledge Sharing Effort*.

LIDILEM : Laboratoire de Linguistique et Didactique des Langues Étrangères et Maternelles.

LVF : Les Verbes Français ([Dubois & Dubois-Charlier, 1997](#)).

ML : *Machine Learning*.

Niveau : Avec une majuscule, le mot *Niveau(x)* renverra aux quatre Niveaux de la méthode CAUTIC® qui structurent les classes (Critères) de la tâche de classification au cœur de nos travaux.

OCDE : Organisation de Coopération et de Développement Economiques.

R&D : Recherche et Développement.

SVM : *Support Machine Vector*.

TAL : Traitement Automatique des Langues.

TF-IDF : *Term-Frequency – Inverse Document Frequency*.

UGA : Université Grenoble Alpes.

UL : unités lexicales.

UX : *User Experience* (Expérience Utilisateur).

verbatim : Extrait de texte allant du syntagme au petit paragraphe, issu du matériau collecté dans les études et considéré comme unité à classifier.

W2V : *Word2Vec*.

yoomaneo : Plateforme communautaire en ligne développée par Ixiade et sur laquelle les données textuelles des études sont collectées.

NOTATIONS

Notations générales :

CONCEPT (petites majuscules) : notation des concepts.

/n-grams/ : notation des n-grams extraits et analysés dans l'exploration textométrique présentée dans le Chapitre V.

OntologyClass : notation des classes d'ontologies.

OntologyClass_ontologyProperty_OntologyClass : notation d'un triplet d'ontologies.

ontologyProperty : notation des propriétés d'ontologies.

'signifié' : notation des signifiés.

« *verbatim* » (italique entre guillemets) : verbatim (i.e., segment codé ou à coder dans le texte selon la grille d'analyse) de notre corpus ou issus d'autres études.

Navigation dans le document au format PDF :

Élément en bleu : élément cliquable (référence bibliographique ou renvoi vers un objet ou un point spécifique du document).

↑_{px} : renvoie au clic, à partir de la bibliographie ou d'un objet en annexe, vers la ou les référence(s) dans le corps de texte.

INTRODUCTION

L'analyse textuelle de contenu qualitatif est une pratique très commune dans le domaine des Sciences Humaines. Le Traitement Automatique des Langues (TAL) peut apporter une aide substantielle pour la réalisation de cet exercice en proposant toute une variété de techniques et outils qui permettent, en partie, de l'automatiser. Malgré des travaux allant dans ce sens, toutes les possibilités d'aide sont loin d'être à ce jour explorées. En outre, les analystes soulignent souvent que les outils qui leur sont proposés n'entrent pas en adéquation avec leurs pratiques. Cette thèse entend apporter des pistes de solution à cette problématique.

Les travaux que nous présentons ont été réalisés dans le cadre d'une Convention Industrielle de Formation par la Recherche (CIFRE) à l'initiative de l'entreprise Ixiade. Ils visent à investiguer les possibilités offertes par le TAL pour la (semi-)automatisation du déploiement d'une méthode d'analyse qualitative particulière. Dans cette introduction, nous présentons le contexte de la thèse et la demande de l'entreprise qui l'a initiée, ainsi que les champs de recherche qu'elle se propose d'investiguer et son positionnement dans ces derniers. Nous introduisons ensuite le plan qui structure le document.

1. Contexte et demande de l'entreprise

La présente thèse a été réalisée dans le cadre d'une collaboration entre le Laboratoire de Linguistique et Didactique des Langues Étrangères et Maternelles (LIDILEM) de l'Université Grenoble Alpes (UGA) et l'entreprise Ixiade. Cette dernière est une agence d'accompagnement à l'innovation dont une des spécialisations majeures est la réalisation d'études d'acceptabilité pour ses clients. *L'acceptabilité* désigne la propension qu'ont les potentiels utilisateurs de services et produits innovants à les « accepter », c'est-à-dire les intégrer dans leurs usages. Elle entre dans les domaines plus globaux de la sociologie des usages et de *l'Expérience Utilisateur* (désignée communément par *UX* pour *User Experience*). Ixiade propose à ses clients d'évaluer leurs innovations dans le cadre d'études qualitatives. Ces dernières consistent en la réalisation d'entretiens, de *focus groups*¹ ou d'échanges en ligne auprès d'utilisateurs potentiels à qui l'innovation

¹ Mode de collecte impliquant un groupe de discussion de 5 à 10 personnes.

est présentée et qui sont invités à exprimer leurs opinions à son propos. Les contenus ainsi collectés sont analysés avec une méthode baptisée « Conception Assistée par les Usages dans les Technologies de l'Information et de la Communication » (CAUTIC®), qui a été développée par le sociologue des usages Philippe Mallein et est sous licence CNRS. Elle est conçue pour fournir une évaluation complète de l'acceptabilité selon plusieurs dimensions. Elle permet ainsi à Ixiade de fournir à ses clients une estimation du potentiel de leur innovation à être adoptée par les utilisateurs, ainsi qu'un ensemble de recommandations pour améliorer ce potentiel.

La recherche que nous présentons est motivée par la volonté de l'entreprise d'automatiser une partie de l'analyse CAUTIC®. En effet, cette étape constitue une activité particulièrement chronophage pour les analystes. D'autre part, Ixiade souhaite orienter ses pratiques d'études vers des modes de collecte digitaux, en favorisant le recours à une plateforme dont elle est propriétaire. Cette orientation implique la probabilité d'une augmentation significative de la quantité de contenu à traiter ; cela pourrait susciter, à terme, une impossibilité d'analyse manuelle. Confrontée au constat qu'aucun outil existant n'était en mesure de répondre à ses besoins au vu de la spécificité de sa méthode, l'entreprise a donc souhaité initier des travaux de recherche en TAL afin d'investiguer les possibilités qu'offre cette discipline pour parvenir à une solution adaptée.

Ixiade souhaite donc disposer d'un outil pour réduire le temps passé à analyser ses contenus d'étude. Cela représente pour elle un enjeu important ; la réduction de ressources affectées à cette tâche lui permettrait de proposer à ses clients des prestations moins coûteuses et plus riches en termes de volumes de données traitées.

2. Champs de recherche et positionnement

Si la demande qui est formulée par l'entreprise est motivée par des enjeux opérationnels et économiques, la question qu'elle adresse à la recherche à cette occasion soulève de nombreux défis.

La pratique de l'analyse qualitative outillée n'est pas nouvelle ; dès les années 1990 ont été développés des logiciels communément désignés par l'acronyme *CAQDAS* pour *Computer-Aided Qualitative Data Analysis Software*. Ils visaient davantage à offrir une facilitation d'ordre pratique aux analystes qu'à proposer une prise en charge partielle de leur tâche par des processus d'automatisation. Il s'agissait en effet, pour la plus grande

partie d'entre eux, de permettre une facilitation logicielle du processus appelé *codage*. Ce dernier est une pratique au cœur de l'analyse qualitative qui consiste à attribuer à des segments de texte d'intérêt des *codes* appartenant à la grille d'analyse qui guide l'étude.

Aujourd'hui, le TAL est en mesure de produire des outils d'un grand intérêt pour assister les analyses qualitatives, pour le compte de la recherche comme de l'industrie, avec des techniques telles que l'analyse des opinions ou encore le *topic modelling*. Il est cependant surprenant que des travaux visant, à proprement parler, à la production de systèmes d'autocodage adaptés aux besoins variés des chercheurs et chargés d'études soient encore très rares ; seuls quelques travaux relativement récents s'intéressent à cette problématique malgré les nombreux questionnements qu'elle soulève et les opportunités opérationnelles dont elle est porteuse (N.-C. Chen et al., 2018; Crowston et al., 2012; Kaufmann et al., 2020; Marathe & Toyama, 2018). C'est donc dans ce champ peu investigué de l'autocodage qualitatif que s'inscrivent les présents travaux.

Une des difficultés majeures de ce type d'outils est que les besoins des chercheurs sont très diversifiés et que chaque étude menée comporte son lot de spécificités, ce qui se traduit par des grilles de codes extrêmement diverses et changeantes. La demande qui nous est adressée concerne en revanche l'automatisation d'une méthode qui sert à étudier un sujet bien particulier – l'acceptabilité des innovations – et qui utilise des codes prédéterminés qui ne changent pas d'une étude à l'autre. Ces codes comportent en revanche un degré de complexité sémantique certain. Les travaux présentés dans cette thèse proposent donc d'étudier les possibilités offertes par le TAL pour l'autocodage limité à une méthode bien spécifique mais relativement complexe à appliquer, même pour un analyste humain.

Nous avons donc une opportunité particulière d'étudier dans quelle mesure, la problématique de la variabilité des codes écartée, le TAL peut être utilisé pour construire des solutions d'autocodage qui entrent en réelle adéquation avec les besoins des analystes. Bien souvent en effet, il appartient à ces derniers de composer avec les outils dont ils disposent malgré des écarts certains entre les dimensions généralistes de ces outils et leur besoin de produire une analyse fine de phénomènes complexes. En nous appuyant sur une grille de codes stable mais dont la maîtrise – bien humaine – nécessite une appropriation des phénomènes qui sont liés à l'acceptabilité, nous avons également une deuxième opportunité ; il va s'agir de comprendre comment et dans quelle mesure le

TAL est capable de reproduire une analyse initialement réalisée par des experts qui mobilisent à cet effet leurs connaissances et prennent donc en considération nombre d'éléments extralinguistiques.

D'autre part, si les possibilités de conception d'outils d'autocodage qualitatif sont peu étudiées, les différentes techniques de TAL qui sont susceptibles d'être mobilisées pour cela sont quant à elles nombreuses et largement documentées dans la littérature. Cet écart entre, d'un côté, des systèmes de TAL variés et de plus en plus performants et, de l'autre, une quasi-absence de solutions d'autocodage satisfaisantes, montre bien que la mise à disposition des premiers pour les secondes reste un défi complexe à relever.

Notre travail consistera donc à étudier, parmi les alternatives techniques variées du TAL, lesquelles sont les plus adaptées à l'objectif fixé. Le domaine qui est concerné est l'analyse sémantique. En outre, les grands principes de la tâche impliquant l'attribution de codes prédéterminés à des segments de texte, nous traitons une problématique qui peut être définie comme une tâche d'annotation ou de classification. Nous explorerons en outre des techniques et approches de textométrie, de classification par *Machine Learning* (ML), d'annotation syntaxico-sémantique par ressources externes, ainsi que de modélisation et analyse textuelle par ontologies.

3. Approche et plan

L'approche que nous proposons consiste tout d'abord à mettre en exergue les spécificités de l'analyse qualitative en termes de défis adressés au TAL et à bien identifier les particularités de la méthode CAUTIC® qui viennent s'y superposer. Cela sera l'objet de la Partie I de cette thèse. Sur cette base, nous menons ensuite une investigation sur le potentiel qu'ont différentes techniques du TAL pour résoudre le problème que nous avons précédemment défini. Cette investigation nous mène à considérer deux grandes approches, l'une ascendante et l'autre descendante. Ces deux approches structurent la suite de la thèse, en étant respectivement présentées dans les Parties II et III.

Plus précisément, cette thèse est divisée en trois parties qui comportent chacune trois chapitres. La **Partie I** contient l'état de l'art ainsi que l'explicitation du positionnement de la thèse et de sa problématique. Dans le **Chapitre I** nous proposons un exposé détaillé des besoins applicatifs qui motivent la demande de l'entreprise. Nous présentons pour cela une introduction au domaine de l'acceptabilité et à la méthode

CAUTIC® en particulier, puis faisons un point sur la nature des données à traiter avant de résumer la demande de l'entreprise en mettant en évidence ses points saillants. Afin d'ancrer la demande de l'entreprise ainsi précisée dans le champ de recherche dans lequel elle s'inscrit, le **Chapitre II** présente un état de l'art de l'analyse qualitative outillée, et le **Chapitre III** précise le positionnement des travaux, formalise la problématique et introduit les grandes approches suivies dans la suite.

La **Partie II** présente les travaux menés selon une approche ascendante. Le **Chapitre IV** décrit la manière dont nous avons constitué un corpus d'apprentissage et en précise le contenu. Le **Chapitre V** expose les expérimentations menées dans le cadre d'une approche statistique, à travers une exploration textométrique des données et des tests de classifieurs. Enfin, le **Chapitre VI** présente une nouvelle expérimentation menée quant à elle sur la base du recours à une ressource syntaxico-sémantique pour le français, Verbønet (Danlos et al., 2016).

La **Partie III** présente une proposition d'approche descendante et ses trois chapitres en détaillent les différentes étapes. Le **Chapitre VII** introduit la phase de modélisation experte de la méthode CAUTIC® que nous avons réalisée afin de créer un modèle d'analyse basé sur une ontologie. Le **Chapitre VIII** expose plusieurs projections de ce modèle sur les données. Enfin, le **Chapitre IX** explique comment nous l'opérationnalisons et l'évaluons, à travers une phase d'augmentation de base lexicale puis une expérimentation en conditions réelles d'analyse visant à préparer son utilisation dans une perspective d'apprentissage continu.

PARTIE I – BESOINS

INDUSTRIELS ET OUTILS EXISTANTS

CHAPITRE I : BESOINS APPLICATIFS

Les travaux de recherche que nous présentons dans cette thèse visent à apporter une réponse à une problématique industrielle spécifique de la société Ixiade ; il s'agit d'aller vers une semi-automatisation d'une tâche d'extraction et de classification de texte dans le contexte d'une méthode d'analyse qualitative intitulée *méthode CAUTIC*[®] (Conception Assistée par l'Usage pour les Technologies, l'Innovation et le Changement). Aussi, il nous paraît judicieux de commencer par présenter le contexte dans lequel s'inscrit cette méthode, ses grands principes et son fonctionnement, afin de bien définir les besoins applicatifs du système que nous entendons construire.

Ces questions seront abordées grâce au double éclairage théorique et pratique dont nous avons pu bénéficier durant les années passées au sein de l'entreprise. Il nous a en effet paru important d'acquérir une connaissance pointue des métiers que nos algorithmes doivent assister et imiter. Nous avons donc, parallèlement à ces travaux, exercé le métier de chargée de projets pour Ixiade pendant quatre ans. A ce titre, nous avons travaillé sur plusieurs projets clients, de manière ponctuelle ou bien complète en intervenant à chacune des différentes phases d'un même projet. Cette ambivalence dans les rôles que nous avons personnellement assumés (analyste vs taliste, experte vs agnostique) nécessitera, dans le cadre du présent exercice de thèse, de clarifier notre positionnement. Cette question sera traitée par la suite. Pour l'heure, nous nous contentons de préciser que les informations et réflexions qui vont alimenter ce chapitre sont basées à la fois sur des travaux scientifiques, des échanges réguliers avec les membres d'Ixiade dans une démarche de transmission de la connaissance métier, et de notre propre pratique de ce dernier.

Afin de présenter la demande d'Ixiade, donc, nous commencerons par une introduction au domaine de l'innovation en nous attardant sur la notion d'acceptabilité. Puis, nous nous concentrerons sur la méthode qui est au cœur de notre thèse, CAUTIC[®], en en faisant une présentation générale suivie d'une présentation des critères de classification qu'elle utilise. Nous aborderons ensuite la question de la nature des données que cette méthode sert à analyser et qui constitueront nos corpus. Enfin, nous conclurons en résumant la demande de l'entreprise en mettant en avant ses points saillants.

1. Innovation et acceptabilité : enjeux et définitions

Le contexte applicatif général de nos travaux est celui de l'accompagnement aux projets d'*innovation*. Il convient alors dans un premier temps de se demander « Qu'est-ce que l'innovation ? ». En effet, si l'objet de nos travaux n'est pas de nous employer à en délimiter précisément les contours – des spécialistes de différents domaines s'y attèlent avec un niveau d'expertise bien supérieur au nôtre – il nous semble important à ce stade de donner quelques indications afin de définir le périmètre de notre objet d'étude, ainsi que d'aborder la complexité du sujet auquel la méthode CAUTIC®, au cœur de nos travaux, s'attaque. Nous présenterons donc brièvement quelques aspects de l'innovation afin de la définir dans ses grandes lignes.

1.1. Le domaine de l'innovation

Dans ces quelques lignes, nous aborderons l'innovation sous trois de ses facettes : l'innovation comme enjeu économique et stratégique, comme procédé multiforme, et comme prise de risque.

1.1.1. L'innovation comme enjeu économique et stratégique

L'innovation est aujourd'hui un enjeu économique et stratégique majeur pour les entreprises et les états. Elle est en effet considérée comme élément clef pour assurer la croissance et la compétitivité des pays industrialisés, mais également comme principal outil de réponse à des enjeux mondiaux majeurs actuels et futurs (environnement, grandes transformations sociales, etc.). Elle est, à ce titre, fortement encouragée par les pouvoirs publics, sous forme d'aides directes ou indirectes. A l'échelle internationale, selon l'Organisation de Coopération et de Développement Economiques (OCDE), les dépenses intérieures brutes liées à la R&D représentent 2,476% du PIB mondial².

1.1.2. L'innovation comme procédé multiforme

L'innovation est considérée à bien des égards comme un objet complexe, notamment de par le fait qu'elle peut prendre des formes très variées. Outre le fait qu'elle concerne tous les domaines d'activité, il existe plusieurs types d'innovations en termes d'objet final ciblé ; le *Manuel d'Oslo* (OCDE/Eurostat, 2005, p. 47) – guide publié par

² <https://data.oecd.org/fr/rd/depenses-interieures-brutes-de-r-d.htm> (consulté le 20/12/21).

l'OCDE – distingue les innovations de produits (biens et services), de procédés, de marketing et d'organisation.

Un autre axe de catégorisation des innovations est celui du degré de nouveauté créée. On oppose en effet généralement, pour le résumer assez schématiquement, l'innovation incrémentale à l'innovation de rupture. La première désigne un processus d'amélioration de l'existant dans un marché et un environnement concurrentiel donné, tandis que la deuxième consiste en la création de quelque chose de singulièrement nouveau, qui va notamment avoir pour impact de changer la nature même de l'avantage concurrentiel (Villon de Benveniste, 2013).

Enfin, quelle que soit leur nature, les projets d'innovation sont généralement définis et impactés par de nombreux paramètres qui leur confèrent toute leur singularité et leur complexité : la multiplicité des acteurs concernés, les facteurs humains qui entrent en jeu, les dynamiques économiques et sociales, et bien d'autres éléments encore. Aussi, l'étude de l'innovation est un domaine largement pluridisciplinaire qui mobilise des individus spécialistes de domaines techniques mais aussi économiques et des Sciences Humaines.

1.1.3. L'innovation comme prise de risque

Un aspect non négligeable de l'innovation en tant que démarche industrielle et commerciale est l'importante prise de risque qu'elle implique. En effet, on estime qu'en moyenne 90 à 95% des projets d'innovation se soldent par des échecs (Andréani, 2004). Si ce chiffre est discuté et peut être nuancé selon les secteurs concernés (Le Nagard-Assayag et al., 2015), et si les conséquences négatives sont plus ou moins importantes pour les entreprises selon la phase du projet durant laquelle l'échec est constaté, il n'en reste pas moins qu'innover est une démarche certes nécessaire pour les industriels qui souhaitent rester compétitifs mais aussi particulièrement périlleuse, quels que soient la taille et le degré de maturité de l'entreprise (Kleinknecht & Van Der Panne, 2012).

Dès lors, le recours à des méthodes et outils destinés à tester plusieurs aspects relatifs au produit ou service innovant prend tout son sens dans la nécessité de minimiser, autant que possible, la prise de risque. Ces aspects sont multiples et leur analyse relève

souvent de champs d'étude spécifique : stratégie, désirabilité, ergonomie, utilisabilité³, etc. Nous nous attarderons au point suivant sur l'acceptabilité, qui est l'aspect que la méthode CAUTIC® permet d'analyser et d'évaluer.

1.2. L'acceptabilité

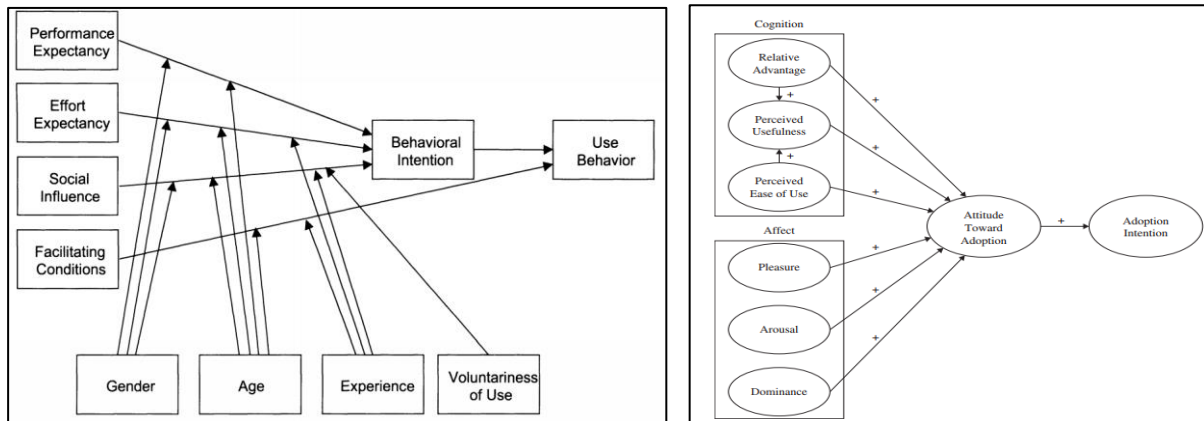


FIGURE 1 : A gauche : modèle UTAUT (figure de Venkatesh et al. (2003)) – A droite : modèle CAT (figure de Kulviwat et al. (2007))

L'acceptabilité désigne, dans les grandes lignes, la capacité de l'innovation à être *a priori* acceptée par les utilisateurs ; elle est considérée comme un facteur clef dans la décision des consommateurs d'adopter ou non l'innovation et joue donc un rôle crucial dans le succès de cette dernière. Plusieurs modélisations ont été proposées pour la représenter. Le premier modèle à avoir servi de référence est le *Technology Acceptance Model* (TAM) (Davis, 1989; Davis et al., 1989) qui considère que les deux facteurs d'acceptabilité principaux sont l'utilisabilité et l'utilité perçues. Des modèles plus récents, comme l'*Unified Theory of Acceptance and Use of Technology* (UTAUT) (Venkatesh et al., 2003) et le *Consumer Acceptance of Technology* (CAT) (Kulviwat et al., 2007) – illustrés dans la FIGURE 1 – prennent en compte des paramètres cognitifs, émotionnels et sociaux.

Ce bref tour d'horizon du domaine de l'innovation ayant permis de décrire le cadre dans lequel s'inscrit la méthode CAUTIC®, nous proposons dans la section suivante de la présenter.

³ L'utilisabilité renvoie, selon l'ISO 9241-11 (1998), au « degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié ».

2. Caractéristiques de la méthode CAUTIC®

La méthode CAUTIC® a été créée par le sociologue des usages Philippe Mallein au sein de la Maison des Sciences de l'Homme (MSH) de Grenoble, ce qui lui a valu d'être distingué du prix Cristal du Centre National de la Recherche Scientifique (CNRS) en 1999. Elle est issue du programme de recherche éponyme mené entre 1996 et 1999 par des chercheurs de plusieurs disciplines (économistes, psychologues, sociologues, linguistes, ingénieurs, etc.) et soutenu par plusieurs acteurs privés dont Schneider Electric, France Télécom ou encore EDF. Elle a été portée par l'association Dispositifs d'Interaction, Cognition et Analyse sociologique de l'Usage (DICAU), dissoute depuis 2002.

Sa mise en pratique dans un contexte commercial a été réalisée pour la première fois par la société AdValor qui a été créée à la suite des différents ateliers réalisés dans ce contexte et qui a été en activité jusqu'en 2010. L'utilisation de la méthode CAUTIC® est aujourd'hui soumise à l'obtention d'une licence d'exploitation accordée par le CNRS et l'UGA⁴. La société Ixiade, au sein de laquelle notre travail est conduit, est titulaire de cette licence et applique cette méthode depuis plus d'une quinzaine d'années.

La section qui va suivre est une synthèse construite à partir de plusieurs sources et expériences personnelles : Pizelle et al. (2014, pp. 39-73)⁵, Anjembe (2014, pp. 64-72), une formation suivie personnellement au sein d'Ixiade en décembre 2017 et dispensée par Pascal Pizelle, et enfin notre propre pratique de la méthode en contexte de projets clients réels. Nous y proposons une description des objectifs de la méthode, des grands principes théoriques qui la régissent et de son déroulement.

2.1. Objectifs poursuivis

CAUTIC® est une méthode d'étude qualitative qui permet d'évaluer le degré d'acceptabilité d'un produit ou d'un service, c'est-à-dire de déterminer *si, dans quelle mesure, comment et pourquoi* les utilisateurs sont susceptibles d'accepter l'innovation. On considère en effet que les qualités techniques ou esthétiques de l'innovation sont, certes, importantes, mais qu'elles ne suffisent pas, et qu'il est absolument nécessaire d'étudier sa capacité à être intégrée dans les usages, au quotidien. Cette notion d'*usages*⁶ est en effet

⁴ Anciennement, par l'Université Pierre Mendès France (UPMF), qui fait aujourd'hui partie de l'UGA.

⁵ Ouvrage édité par Editions d'Innovation qui appartient à Ixiade.

⁶ Voir dans point suivant (2.2.2, p. 12).

centrale dans la méthode CAUTIC[®], au même titre que celle de *sens* porté par l'innovation ; c'est à travers les *significations d'usage*, que l'on recueille lors du déploiement de la méthode, que l'on peut saisir le *sens de l'usage* de l'innovation (Anjembe, 2014, p. 65). En d'autres termes, il s'agit d'investiguer le degré d'adéquation entre l'individu et ce qui est proposé par le concept innovant, afin de comprendre si l'innovation *fait sens* pour lui (Pizelle et al., 2014, pp. 17-18). *In fine*, l'objectif de l'utilisation de CAUTIC[®] est ainsi d'aider l'entreprise ou l'individu porteur du projet innovant en lui fournissant une évaluation du potentiel de son innovation en termes d'acceptabilité sociale, ainsi que de lui indiquer des voies d'amélioration possibles.

Précisons également que la méthode CAUTIC[®] est conçue pour pouvoir traiter tout type de changement : innovation de produits et services, changement dans une organisation, dans des procédés, etc. La méthode peut en outre être mise en œuvre à plusieurs stades du processus d'innovation (conception, développement, réalisation et développement) ; à chacune de ces étapes, son but est d'indiquer si le projet doit être poursuivi ou s'il doit être réorienté. L'acceptabilité est en effet considérée comme une « preuve de validité » (Pizelle et al., 2014, p. 39) de la conception et un indicateur du cadre dans lequel l'innovation va se développer.

2.2. Principes théoriques

La méthode CAUTIC[®] s'appuie sur plusieurs principes structurants, dont les plus importants sont l'importance du contexte, la notion d'*usage* et celle de *significations d'usage*.

2.2.1. Importance du contexte

Un des fils conducteurs et postulat de la méthode est le fait qu'une innovation arrive toujours dans un contexte dans lequel elle devra s'insérer ; elle « n'arrive pas dans un « désert social » (elle n'émerge pas *ex nihilo*) mais [...] va rencontrer, via son usage par des utilisateurs, toute une série de pratiques, de représentations et de normes déjà en place ou existantes » (Pizelle et al., 2014, p. 40). L'évaluation de l'innovation doit donc nécessairement rendre compte de ce contexte et s'y ancrer.

2.2.2. Notion d'usages

La méthode CAUTIC[®] s'inscrit dans le cadre d'une approche dite *Usage*. Cette dimension centrale implique que la démarche consiste à se concentrer moins sur la

compréhension de l'utilisateur que sur celle, comme son nom l'indique, des *usages*. Nous ne développerons pas en détail ici cette notion dont la description et l'analyse constituent un champ de recherche pluridisciplinaire et fait l'objet d'une littérature abondante qui propose des définitions variées, et parfois divergentes, du terme⁷. Nous retiendrons seulement qu'elle n'est pas synonyme de *pratiques*, ni d'*utilisation* ou encore d'*habitudes*. Dans le cadre de la méthode sur laquelle nous nous concentrons, elle renvoie plus précisément à la « manière dont un individu agit et s'imagine agir avec un objet ; manière dont un individu utilise et envisage l'utilisation d'un objet (produit ou service) ; manière, enfin, dont un individu identifie des caractéristiques intéressantes ou transforme un objet et ses représentations pour l'adapter à ses pratiques » (Pizelle et al., 2014, p.42). Elle comporte donc intrinsèquement les idées de médiation, d'hybridation, de négociation, de convergences entre l'utilisateur et les dispositifs techniques qu'il utilise.

2.2.3. Significations d'usage

Ce que la méthode CAUTIC® cherche à repérer et analyser sont plus précisément les *significations d'usage*, c'est-à-dire ce qui *fait sens* dans l'innovation pour les utilisateurs, et ce, sur deux registres : celui des pratiques et celui des représentations. On entend par *compréhension des pratiques* l'analyse des interactions de l'utilisateur avec les objets de son environnement et avec les autres individus. Les *représentations*, quant à elles, sont la façon dont la personne va projeter et intégrer ou non l'innovation au sein de son système de valeurs, croyances et opinions. En d'autres termes, les *significations d'usage* renvoient au sens et à la valeur qu'un utilisateur attribue à l'innovation qui lui est proposée. Dans une logique de recherche d'améliorations, on va ainsi tenter de comprendre ce qui va créer de la valeur en repérant le sens que les utilisateurs donnent à l'innovation, i.e. quelles *significations d'usage* ils projettent sur cette innovation.

Après cette description des grands principes de la méthode CAUTIC®, nous présentons, de manière plus concrète, le déroulement des études qui la mettent en œuvre.

⁷ Voir par exemple Hussenot (2006) pour un résumé des différentes acceptions et de leur évolution.

2.3. Déroulement d'une étude CAUTIC®

Comme illustré dans la FIGURE 2, le déroulement d'une étude qui déploie la méthode CAUTIC® suit les étapes traditionnelles des études qualitatives sur lesquelles viennent se superposer les spécificités de cette méthode. Nous explicitons et détaillons à présent ce déroulement.

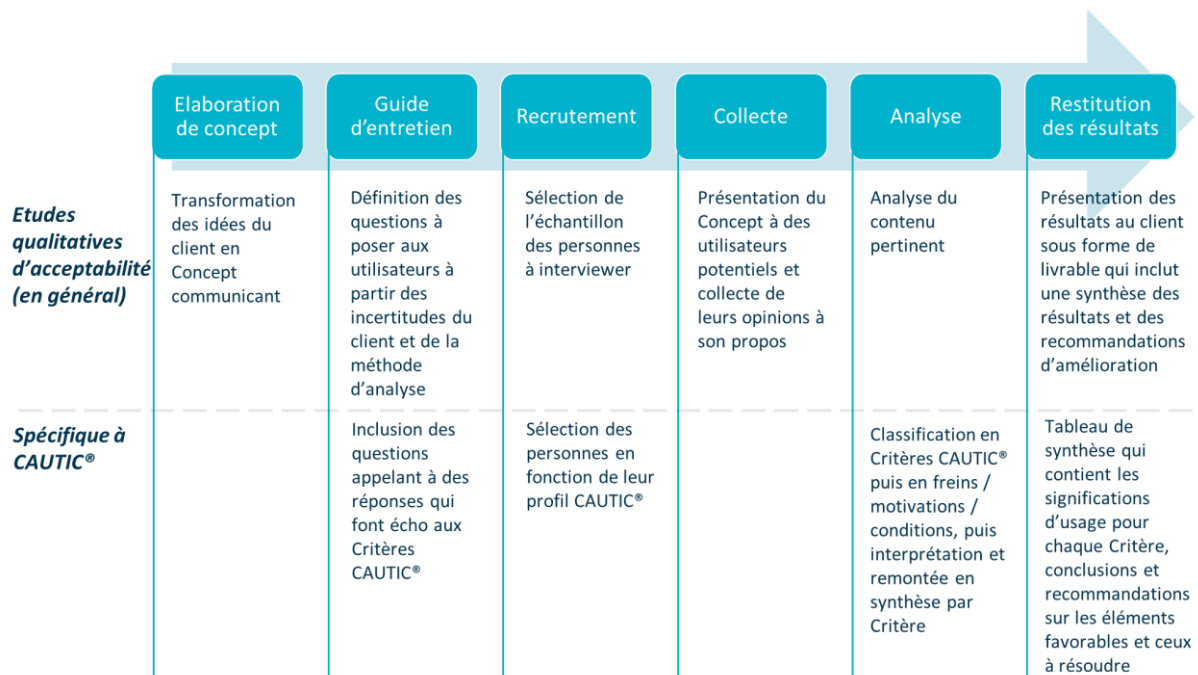


FIGURE 2 : Déroulement de la méthode CAUTIC®

2.3.1. Elaboration d'un Concept communicant

La mise en pratique de la méthode CAUTIC® nécessite, en amont, l'élaboration d'un Concept « orienté usage », c'est-à-dire la production d'une définition du Concept innovant qui incarnera l'objet testé au cours de l'étude. Comme indiqué dans le *Lexique* de cette thèse (p. xi), dorénavant, lorsque nous utiliserons le mot *Concept* avec une majuscule, nous ferons référence à l'objet (produit ou service) développé par le client, testé par Ixiade, et formalisé selon les modalités que nous expliquons dans le présent point ; cette convention typographique permettra de le distinguer des autres acceptions possibles de *concept*.

Cette étape d'élaboration de Concept est réalisée avec le client et est une démarche qui vise à transformer ses idées en une présentation synthétique (diapositives ou vidéo) qui doit permettre aux participants à l'étude de le comprendre et de se forger une opinion à son propos de manière efficace et rapide ; on vise en général un maximum de 10 minutes

de présentation. Cette définition du Concept, et sa présentation, sont structurées autour de sept questions :

- 1) « Dans quel contexte le Concept s'inscrit-il ? »
- 2) « Qu'est-ce que c'est ? »
- 3) « A qui s'adresse-t-il ? »
- 4) « A quoi sert-il ? »
- 5) « Comment s'utilise-t-il ? »
- 6) « Comment fonctionne-t-il ? »
- 7) « Quels sont les bénéfices qu'il apporte ? »

L'ensemble de la présentation doit être la plus neutre possible, c'est-à-dire qu'elle ne doit pas chercher à promouvoir le Concept mais à le décrire de manière factuelle. A ce titre, un vocabulaire à connotation neutre doit être employé ; par exemple, à la question 2, on ne dira pas que le Concept est *une application révolutionnaire*, car la perception ou non de la nouveauté par le participant est un des éléments que l'on cherche à évaluer. Dans cette même logique, les réponses à la question 7 portant sur les bénéfices apportés par le Concept sont formalisées avec le client mais non présentées au participant car on recherche précisément à comprendre s'il perçoit ces bénéfices sans qu'on ne les lui présente. Ainsi, si on estime qu'un bénéfice réside dans le fait que le Concept permet un gain de temps, on se contentera de présenter, à la question 5 (« Comment ça s'utilise ? »), un scénario dans lequel un utilisateur met une certaine durée de temps à accomplir une tâche avec le Concept ; si, spontanément, le participant dit que cela lui permettrait d'être plus rapide, alors la perception du bénéfice est validée.

2.3.2. Guide d'entretien

Le guide d'entretien est construit pour intégrer des questions qui visent à obtenir des réponses aux différents aspects qu'étudie la méthode CAUTIC®. Elles sont mêlées à d'autres, plus spécifiques au projet, et portent sur certaines incertitudes précises des clients. Les questions sont formulées de manière à être très ouvertes et tendent à amener le participant à répondre aux différents points qu'étudie la méthode CAUTIC®, sans jamais – ou alors, très rarement – les verbaliser directement ; il ne faut pas voir les Critères⁸

⁸ Comme indiqué dans le Lexique (p. xiii), le mot Critère(s), avec une majuscule, désignera tout au long de cette thèse les 20 classes utilisées par la méthode CAUTIC®, pour distinguer cette acception des autres possibles.

CAUTIC®, qui seront présentés plus loin, comme une grille de questions qui sont posées telles quelles aux participants, mais plutôt comme une *check-list* des points qu'on aimerait qu'il aborde.

2.3.3. Recrutement

La phase de recrutement vise, comme son nom l'indique, à trouver des participants, tout en s'assurant qu'ils répondent à certains critères spécifiques au projet (par exemple, appartenir à une certaine tranche d'âge, travailler dans une entreprise d'une certaine taille, etc.).

2.3.4. Collecte

La collecte peut s'effectuer selon trois modalités : entretiens en face-à-face, en *focus groups* ou par enquêtes en ligne. Les deux premiers sont les modes de collecte historiques mais restent largement utilisés. Les entretiens individuels en face-à-face durent généralement 1h, tandis que les *focus groups* réunissent entre 7 et 10 personnes sur une durée de trois à quatre heures. On estime qu'il faut 12 personnes interviewées pour récolter 75% des informations complètes, et une étude compte généralement 15 à 20 personnes, même si ce nombre peut varier de manière significative selon les projets.

Le troisième mode de collecte est réalisé, chez Ixiade, par le biais d'une plateforme communautaire propriétaire, *yoomaneo*⁹. Lancée en 2019 – soit un peu plus d'un an après le début de nos travaux de thèse – cette plateforme comporte deux volets principaux : un volet discussions et un volet études. Les utilisateurs de *yoomaneo* sont invités à s'inscrire à plusieurs communautés selon leurs centres d'intérêt et / ou domaines de spécialité (santé, énergie, sport, etc.). Ils peuvent ensuite discuter autour d'innovations, changements ou tendances dans ces thèmes, et sont invités à participer à des études. La collecte est alors étalée sur plusieurs jours avec des questions ouvertes sur des plages de quelques heures ou quelques jours. Lorsque les participants répondent à une question, ils ont accès aux réponses des autres membres et peuvent y réagir en les commentant et en leur attribuant des points de pertinence¹⁰.

Quel que soit le mode de collecte utilisé, il s'agit d'une démarche semi-directive, c'est-à-dire que, si les questions ouvertes du guide d'entretien structurent la

⁹ <https://www.yoomaneo.com/fr>.

¹⁰ Sorte d'équivalent du *like* sur les réseaux sociaux.

conversation, l'expression spontanée est largement privilégiée (il est coutume de dire, par hyperbole, que l'entretien idéal est celui où l'interviewer n'a pas parlé).

2.3.5. Analyse

L'analyse des données ainsi collectées – illustrée dans la FIGURE 3 – consiste, en premier lieu, en une mise à plat (i.e., organisation du matériau discursif) ; l'analyste parcourt le texte, extrait les contenus pertinents et les classe sur deux axes. Ces « contenus pertinents » sont des unités de texte allant de l'ordre du syntagme au petit paragraphe et qui sont appelées *verbatim*. C'est cette acception du mot que nous utiliserons tout au long de cette thèse et, étant donné que nous serons amenée à en présenter de nombreux exemples, nous ferons pour cela usage d'une double indication typographique en utilisant des guillemets et l'italique afin de les différencier clairement des autres citations et exemples de nature différente (cela vaut également pour les verbatim d'autres travaux que nous citerons).

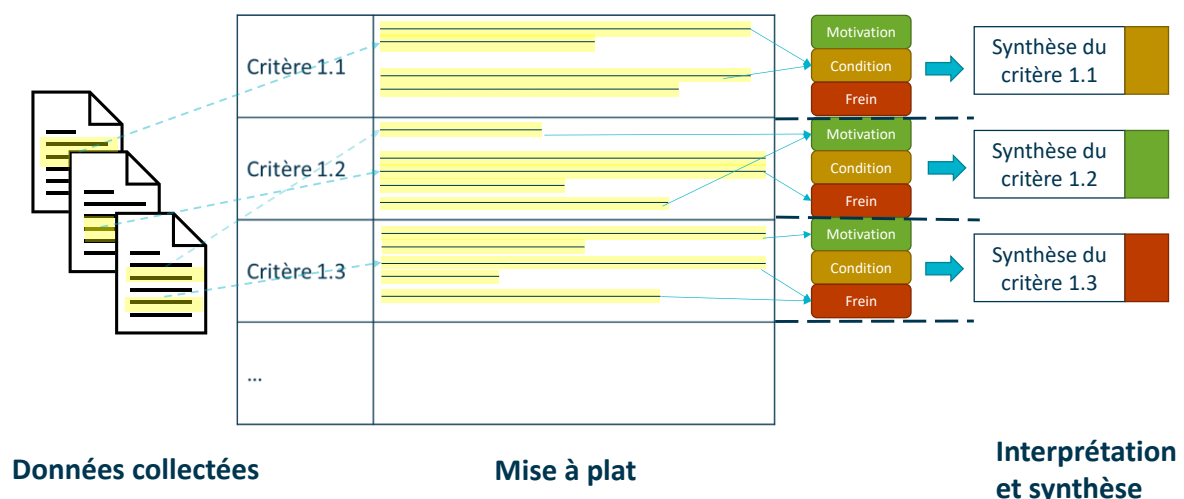


FIGURE 3 : Représentation schématique de la méthode d'analyse CAUTIC®

Le premier axe de classification est celui qui nous intéresse dans le cadre de cette thèse. Il s'agit d'une grille de Critères organisés en quatre Niveaux¹¹. Ces Critères étant centraux dans notre travail, nous prendrons le temps de les détailler dans le point 3 de ce chapitre. D'autre part, précisons dès à présent qu'un même verbatim peut être classifié dans plusieurs Critères. Le deuxième axe de classification consiste à caractériser le

¹¹ Comme indiqué dans le Lexique (p. xiii), le mot *Niveau(x)*, avec une majuscule, désignera tout au long de cette thèse ces quatre Niveaux qui structurent les Critères de la méthode CAUTIC®, afin de distinguer cette acception des autres possibles pour ce lexème.

verbatim selon qu'il exprime un frein, une motivation ou une condition à l'égard du Concept. La semi-automatisation de cette tâche est l'objet d'une seconde thèse CIFRE conduite au sein d'Ixiade, avec le Laboratoire d'Informatique de Grenoble (LIG), par Emmanuelle Kelodjoue.

En termes de modalité pratique d'extraction, les analystes peuvent avoir recours à deux techniques. La première consiste simplement à copier-coller les verbatim dans les lignes d'un tableur et à leur attribuer les classes. La deuxième consiste à utiliser un outil d'extraction spécifique qui permet de sélectionner les verbatim dans le texte depuis une interface Web et de leur attribuer un Critère en même temps, puis d'exporter toutes les extractions dans un tableur une fois cette tâche terminée. Cette possibilité était offerte sur l'outil *Dixit*, développé par Ixiade, qui n'est aujourd'hui plus disponible. Un deuxième outil du même type a été développé pendant que nous réalisons nos travaux de thèse¹².

Une fois cette mise à plat réalisée sur la totalité des données collectées, l'analyste obtient pour chaque Critère une liste de verbatim étiquetés selon leur orientation motivationnelle. L'analyste entreprend alors une phase d'interprétation du contenu. Il réalise ensuite une synthèse par Critère qui consiste à dégager et mettre en exergue les points soulevés par les participants les plus fréquents et / ou pertinents, puis à attribuer à ce Critère un statut de validation ; le Critère sera alors considéré comme *Validé* si les motivations dominent, *Validé sous condition* si ce sont les conditions, et *Non validé* si ce sont les freins.

2.3.6. Livrable

L'analyse réalisée, les résultats sont restitués au client sous forme de document PowerPoint / PDF. Cette présentation contient un descriptif détaillé qui inclut les points saillants par Critère, un tableau de synthèse des statuts de validation par Critère, ainsi que des recommandations pour la suite du projet.

Les grandes lignes de la méthode CAUTIC® ayant été introduites, nous détaillons à présent les Critères de classification qu'elle utilise et qui sont au cœur de nos travaux.

¹² Nous en ferons une présentation dans le dernier chapitre (Chapitre IX, point 7.1, p. 219).

3. Les Critères CAUTIC®

Afin de présenter les Critères CAUTIC®, nous commencerons par introduire les quatre Niveaux sur lesquels ils se répartissent, puis nous en donnons la liste complète.

3.1. Niveaux de Critères

Les Critères sont répartis en quatre Niveaux principaux qui permettent de répondre à autant de questions centrales dans le processus d'acceptation d'une innovation. Ces quatre axes, résumés dans le TABLEAU 1, correspondent chacun à :

- Un des verbes clefs (*Assimiler, Associer, Approprier* et *Adapter*) qui dénotent les processus nécessaires au fait que les significations d'usages soient favorables au succès de l'innovation.
- Un des Niveaux qui participent à la création des significations d'usages : savoir-faire, pratiques, identité et environnement.

L'ordre de ces Niveaux n'est pas anodin ; la méthode est en effet construite selon une logique de cercles concentriques qui partent du niveau cognitif de l'individu (Niveau 1), puis s'étendent à ses pratiques (Niveau 2), puis à son identité / ses valeurs (Niveau 3) et enfin à son environnement (Niveau 4). On part ainsi de l'individu seul pour aller vers l'environnement, et ce de manière assez progressive au cours de l'entretien en posant des questions qui suivent l'ordre des Niveaux.

Verbe clef	Niveau traité	Question à laquelle on veut répondre
<i>Assimiler</i>	Niveau des savoir-faire	L'utilisateur pense-t-il avoir suffisamment de connaissances techniques pour bien utiliser l'innovation ?
<i>Associer</i>	Niveau des pratiques	L'utilisateur estime-t-il que ses pratiques existantes sont compatibles avec les changements amenés par l'innovation ?
<i>Approprier</i>	Niveau de l' identité	L'utilisateur estime-t-il que l'innovation est en accord avec son identité privée et/ou professionnelle et ses valeurs ?
<i>Adapter</i>	Niveau de l' environnement	L'utilisateur estime-t-il que l'innovation est en phase avec son environnement privé, professionnel et/ou industriel ?

TABLEAU 1 : Les quatre Niveaux des Critères CAUTIC®

3.2. Liste des Critères

Nous détaillons à présent la liste des 20 Critères également répartis sur les quatre Niveaux et qui constitueront les classes de notre système. Pour ce faire, nous présentons pour chaque Niveau la liste des Critères dans des encadrés, puis nous donnons, dans un tableau, pour chaque Critère, un exemple de verbatim qui peut être classifié dans ce Critère. Nous attirons l'attention du lecteur sur le fait que ces exemples sont fournis à titre indicatif, pour améliorer la compréhension, mais qu'en aucun cas ils ne sont exhaustivement emblématiques de ce qui peut être classifié dans chaque Critère ; une description détaillée de ceci sera présentée dans la suite de cette thèse. Notons également que ces exemples sont des verbatim ayant réellement été prononcés par des participants à l'occasion d'études passées conduites par Ixiade, et que, afin de préserver la confidentialité des données, nous remplacerons certains éléments par des classes générales indiquées entre crochets (par exemple [nom de produit concurrent]).

3.2.1. Critères du Niveau 1 – « Assimilation aux savoir-faire »

1.1 – « L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ? »

1.2 – « Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ? »

1.3 – « L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ? »

1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? »

1.5 – « L'usage du concept est-il envisageable comme un système d'usage courant ? »

ENCADRÉ 1 : Critères du Niveau 1 – « Assimilation aux savoir-faire »

1.1	« Il y a un côté sympa, un côté fun qui fait qu'on a envie de l'avoir. On a envie de dire : « Vous pouvez le mettre en place s'il vous plaît ». J'aime bien. »
1.2	« Donc il faut qu'ils soient sur le même réseau Wifi ? »
1.3	« c'est ce que l'on fait déjà sur les groupes électrogènes, ventilateurs, compresseurs, etc. »
1.4	« La fonction différentielle intégrée est aussi intéressante. »
1.5	« Moi, je m'en servais pour [tâche métier de la participante] mais l'utiliser au moment de [autre tâche métier de la participante], je ne pourrais pas. »

TABLEAU 2 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 1 – « Assimilation aux savoir-faire »

Les Critères du Niveau 1 – « Assimilation aux savoir-faire » sont listés dans l'ENCADRÉ 1, et les exemples de verbatim sont donnés dans le TABLEAU 2. Ces Critères servent à évaluer le degré de compréhension du Concept et l'intérêt spontané manifestés par le participant ; ces deux points constituent le sujet du Critère 1.1 – « L'utilisateur comprend-il le concept et le trouve-t-il intéressant ? » qui est en fait un Critère de synthèse des autres du même Niveau et dans lequel on va classer des verbatim relativement génériques.

3.2.2. Critères du Niveau 2 – « Association aux pratiques »

2.1 – « Le concept est-il simple à utiliser ? »

2.2 – « Le concept compléterait-il sans les concurrencer les habitudes existantes ? »

2.3 – « La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédible les pratiques nouvelles proposées ? »

2.4 – « Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ? »

2.5 – « L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ? »

ENCADRÉ 2 : Critères du Niveau 2 – « Association avec les pratiques »

2.1	« Tel qu'il est présenté là, il a l'air facile d'utilisation autant pour la mise en œuvre que pour avoir les résultats. »
2.2	« On est plus aguerris techniquement grâce à ce type d'outils, c'est un moyen d'affiner la connaissance sur le sujet. »
2.3	« Actuellement il faut une vis mais là, les clips plastiques ça va péter. »
2.4	« Pour moi on peut moduler notre installation électrique tout seul, donc ça c'est un problème résolu. »
2.5	« Je pense que j'aurais un bel outil parce que c'est relativement puissant mais que je n'aurais ni le temps, ni les moyens d'exploiter. Je pense que c'est quelque chose qui tomberait en désuétude et c'est dommage parce que je pense qu'il y a des choses intéressantes à faire mais dans mon activité quotidienne... »

TABLEAU 3 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 2 – « Association aux pratiques »

Dans les Critères du Niveau 2 – « Association aux pratiques » (ENCADRÉ 2 et TABLEAU 3) la notion d'hybridation des pratiques est omniprésente, principalement dans les Critères 2.2 et 2.3 qui font se compléter et se télescoper pratiques anciennes et pratiques proposées par le Concept. Les autres Critères se concentrent sur des aspects

relatifs à l'utilisation *per se* du Concept, à travers des dimensions telles que la facilité d'utilisation ou la possible résolution de problèmes grâce au Concept.

3.2.3. Critères du Niveau 3 – « Appropriation à l'identité socio-professionnelle »

Les Critères du Niveau 3 – « Appropriation à l'identité socio-professionnelle » (ENCADRÉ 3 et TABLEAU 4) concernent tout ce qui a trait aux valeurs, au rôle et à l'imaginaire du participant. La notion d'« imaginaire d'appropriation » dans le Critère 3.4. renvoie tout simplement au fait que le participant va verbaliser une association entre le Concept et quelque chose qui n'a aucun rapport avec ce dernier ; le jeu du portrait chinois (« si le Concept était un animal ? », « si c'était un modèle de voiture ? », etc.) a longtemps été utilisé pour explorer ce point lors de la phase de collecte, mais cette pratique a été laissée de côté au fil de années au profit de la recherche de comparaisons plus spontanées. D'autre part, les « extensions d'usage » du Critère 3.5 correspondent à toutes les utilisations que le participant envisage pour le Concept et qui n'étaient pas prévues par les concepteurs.

3.1 – « L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ? »

3.2 – « Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ? »

3.3 – « Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ? »

3.4 – « Le concept suscite-t-il un imaginaire d'appropriation ? »

3.5 – « Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ? »

ENCADRÉ 3 : Critères du Niveau 3 – « Appropriation à l'identité socio-professionnelle »

3.1	« C'est pour des gens ouverts aux technologies. »
3.2	« Je dirais que c'est quelque chose de très valorisant. »
3.3	« Donc il faut qu'il y ait une personne avec un IPAD accroché au cou et qui regarde en permanence... c'est anti-social ça. »
3.4	« Je pense à une oie parce que dès que quelque chose se passe mal dans son environnement, elle se met à gueuler. » [en référence à une fonction d'alarme du Concept]
3.5	« J'ai l'impression que ça peut être utilisable en dehors de [champ d'application prévu pour le Concept] : on fait beaucoup de [autre champ d'application], je pense que ça pourrait être adapté, avec des modifications. »

TABLEAU 4 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 3 – « Appropriation à l'identité socio-culturelle »

3.2.4. Critères du Niveau 4 – « Adaptation à l’environnement »

Les Critères du Niveau 4 – « Adaptation à l’environnement » (ENCADRÉ 4 et TABLEAU 5) s’articulent autour de deux déclinaisons possibles : le contexte professionnel de l’utilisateur et le contexte privé. On s’intéresse, dans ce Niveau, à tout ce qui touche au tissu de relations sociales / professionnelles de l’utilisateur et à l’impact que pourrait avoir le Concept sur celui-ci. Les Critères 4.1 et 4.5 se concentrent sur une possible commercialisation en questionnant, respectivement, sa plausibilité par rapport aux notions d’air du temps et de contexte concurrentiel, et son prix possible.

4.1 – « L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ? »

4.2 – « Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ? »

4.3 – « Le concept est-il adapté à la place qu'occupent les professionnels dans la filière économique / aux relations familiales et à leur évolution ? »

4.4 – « Le concept est-il adapté aux types d'organisation et à leur capacité d'évolution / aux manières de vivre de l'utilisateur et à leur évolution ? »

4.5 – « Quel prix l'utilisateur consent-il à payer pour le concept ? »

ENCADRÉ 4 : Critères du Niveau 4 – « Adaptation à l’environnement »

4.1	« Donc ce projet, il est... comment dire... il est tout à fait en adéquation avec l'évolution de [domaine visé par le Concept] qu'on va avoir dans les années qui viennent, c'est clair. »
4.2	« Si j'envoie les résultats uniquement aux [partenaires] qui ont le logiciel, ça réduit mes possibilités. C'est comme si je me pacais avec eux. »
4.3	« Donc c'est pour ça que les choses qui sont assez novatrices sont intéressantes pour nous. Parce que voilà, on prouve qu'on est toujours à l'écoute du marché, qu'on est capable de... qu'on est capable de s'intégrer dans ce marché, d'installer des produits qui sont viables et qui font des gains. »
4.4	« La pérennité du produit, autant sur du classique on peut se débrouiller mais là si ça passe de mode, le constructeur décide d'abandonner ce produit par contre là on a des soucis... »
4.5	« Honnêtement je n'ai aucune référence en fait par rapport à ça parce que je n'ai jamais téléchargé d'appli payante donc en fait je ne sais pas combien ça vaut. »

TABLEAU 5 : Exemples de verbatim pouvant être classifiés dans les différents Critères du Niveau 4 – « Adaptation à l’environnement »

Dans cette section, nous avons brièvement exposé au lecteur les Critères de la méthode CAUTIC®. Nous abordons à présent, de manière tout aussi macro, la nature des données récoltées lors des études et qui constitueront les corpus que notre système devra traiter.

4. Nature des données

Les données à traiter sont principalement de deux natures différentes : des transcriptions d'oral sur logiciel de traitement de texte pour les entretiens et *focus groups*, et du contenu saisi en ligne par des utilisateurs et exporté dans un tableur pour les études menées sur *yoomaneo*.

4.1. Aspects formels

Les données orales transcrites sont des documents au format *.docx* dans lesquels la couleur des différentes parties permet de distinguer les locuteurs ; généralement, du noir pour l'interviewer et du bleu pour l'interviewé. Lorsque plusieurs participants sont impliqués – c'est-à-dire dans le cas des *focus groups* ou les rares cas où un entretien rassemble deux participants – les locuteurs sont précisés en début de ligne par un code ou un prénom suivi de deux points (*Justine :* ou *P02 :*). D'autre part, les parties inaudibles qui n'ont pas pu être transcrites sont indiquées par un code de format (*? timecode ?*), généralement mis en évidence par une couleur différente. Il est bien évident que ce format de document est adapté à une lecture humaine et non à un traitement informatique.

Les données issues de *yoomaneo* sont exportées depuis la plateforme dans un format tableur *.xlsx*. Chaque ligne correspond à un *post* (réponse à une question de l'étude ou commentaire sur une des réponses données par un autre participant) et indique, en plus du contenu, le locuteur, un code de *post*, si besoin, le code de la réponse auquel le *post* répond, les points de pertinence attribués au *post* par les autres utilisateurs, ainsi que les émoticônes utilisées (il est possible sur *yoomaneo* d'associer des émoticônes aux *posts*). Ce format est bien entendu plus adapté à un traitement informatique, et est d'ailleurs considéré comme peu pratique par les analystes pour une extraction humaine des verbatim.

4.2. Aspects linguistiques

D'un point de vue linguistique, si d'autres observations seront réalisées par la suite, nous pouvons d'ores et déjà noter quelques caractéristiques spécifiques des données qui constitueront nos corpus. Il faut pour cela distinguer les données orales transcrites des données récoltées en ligne sur *yoomaneo*. Cette distinction est d'autant plus importante que les premières seront nos données d'apprentissage tandis que les secondes seront nos données cibles, et que leurs différences doivent donc être clairement identifiées. Ainsi, nous décrirons dans un premier temps les modalités de transcription utilisées pour les textes oraux transcrits, puis nous aborderons les pratiques langagières observées dans les données de *yoomaneo*. Ensuite, nous discuterons des caractéristiques sociolinguistiques puis psycho-linguistiques des données – quel que soit le média – ainsi que leur exploitation dans le cadre de la méthode CAUTIC®.

4.2.1. Modalités de transcription des textes oraux transcrits

Les transcriptions d'oral – réalisées en interne de l'entreprise ou par des prestataires – sont à mi-chemin entre de la transcription synthétique et du mot à mot tel que pratiqué dans les travaux de recherche linguistique portant sur l'oral. On retrouve en effet des caractéristiques typiques de transcription d'oral conversationnel : indications de disfluences, connecteurs phatiques (*oui, d'accord*), marqueurs de structuration (Gülich, 1999) typiques de l'oral (*ben, quoi, ah*), syntaxe moins rigide qu'à l'écrit, etc. Notons cependant que toutes les disfluences ne sont pas transcrites. Certaines sont en effet omises si elles nuisent trop à la compréhension, le principe général étant de conserver la marque d'une hésitation sans pour autant restituer l'intégralité de sa manifestation orale (un *euh... ben... je... euh...* pourra être simplement restitué par un *euh...*); ces transcriptions ont été réalisées dans l'optique d'être vectrices des informations utiles pour de l'analyse qualitative et non pour une analyse linguistique. Or, le fait qu'une hésitation soit présente intéresse l'analyste en études qualitatives, mais c'est là la seule information dont il a besoin, et sa lecture ne doit pas être entravée par un nombre trop important de disfluences.

Si toute démarche de transcription comporte une certaine part d'interprétation de la part du transcripateur, cette part est donc assez notable dans le cas de nos données, notamment de par le fait que le texte soit ponctué (Blanche-Benveniste & Jeanjean, 1987). En outre, si des marques d'oralité sont présentes dans les textes, d'autres aspects sont

complètement gommés par la transcription : cette dernière ne fait pas usage de « trucages orthographiques » (*Ibid.*) (*p'tite, chais pas*, etc.) et a même tendance à une surnormalisation de certains mots ; dans de nombreuses transcriptions, notamment, *ça* est transcrit par *cela*.

4.2.2. Usages linguistiques des données saisies en ligne

Les données récoltées sur *yoomaneo* sont bien évidemment d'une autre nature que celles que nous venons de décrire. Il s'agit tout d'abord, cette fois, de données primaires, directement saisies en ligne par les participants, en opposition aux données transcrites qui sont *de facto* des données secondaires. De plus, l'écart entre ces données et de l'écrit standard est également d'une autre nature ; il s'agit principalement de fautes de frappe, d'accord, et, surtout, de ponctuation / d'usage de signes typographiques, avec notamment une utilisation des apostrophes et espaces qui divergent des standards (**je saurais pas trop, c est bizarre*). Ces variations sont assez typiques des contenus postés en ligne, mais précisons que nous sommes dans un type d'écrit qui ressemble à ce que l'on peut trouver sur des forums « sérieux » plus que sur les réseaux sociaux grand public : l'utilisation de *hashtags* n'est pas pratiquée, les « extensions graphiques » (Dejong & Mercier (2002); cités par Hadi-Denoueix (2014, p. 189)) à visée expressive (*troooooooooo bien !*) sont extrêmement rares, les néographismes typiques du langage sms (*tkl, je c pas*, etc.) ne sont pas non plus utilisés, etc. Il faut cependant noter que *yoomaneo* est une plateforme relativement nouvelle et que les usages peuvent être amenés à évoluer.

4.2.3. Aspects sociolinguistiques et leur traitement

Les données traitées étant produites par de nombreux locuteurs et via des médias différents, il est bien évident que de nombreuses variations y sont présentes à des degrés plus ou moins importants. Cependant, il faut noter que les manifestations de ces variations sont réduites par les conventions de transcription et la nature de l'exercice de l'entretien.

En effet, comme nous l'avons évoqué, dans le cas des transcriptions d'oral, le transcripteur a souvent tendance à normaliser les paroles transcrites ; non seulement les aspects phonostylistiques ne sont aucunement restitués par la transcription, mais la syntaxe est presque systématiquement normalisée – sans laisser de trace de la normalisation – si la phrase prononcée est agrammaticale. Il est également possible que le lexique soit parfois standardisé. Sur ce dernier point, nous avons déjà évoqué l'exemple

du remplacement de *ça* par *cela*, qui est une pratique qui a tendance à disparaître aujourd'hui, mais nous pouvons également trouver sporadiquement quelques exemples de substitutions lexicales vers un lexique plus standard. Cela dit, ces substitutions restent plutôt rares, et l'on tend plutôt à vouloir restituer le discours tel qu'il a été énoncé dans le souci d'en préserver l'authenticité. Ainsi, les variations diatopiques et diastratiques¹³ sont complètement neutralisées par la transcription pour ce qui concerne la prononciation, et peuvent apparaître, dans une certaine mesure, à travers le lexique. On peut, en effet, pour ce qui est de l'ordre diatopique, trouver quelques régionalismes. Quant au diastratique, la normalisation des transcriptions ainsi que le cadre assez codifié des échanges confèrent aux textes un registre de langue assez neutre ; nous nous situons toujours dans un registre plutôt courant ponctué, d'un côté, de quelques tournures familières – voire quelques jurons – et, de l'autre, de l'utilisation d'une langue potentiellement plus soutenue mais qui relève surtout de l'emploi de mots appartenant à une langue de spécialité. D'autre part, cette observation est bien valable quel que soit le mode de collecte des données ; la variation diamésique au sein du corpus n'implique pas nécessairement de gros écarts sur le plan diaphasique (le degré d'autosurveillance des locuteurs est assez similaire quel que soit le média). Les exemples 1) et 2) ci-dessous illustrent bien cela. Tous deux concernent des problématiques de continuité de service / anticipation des pannes. Le premier est issu d'une transcription d'un entretien oral dans lequel le participant est interrogé en sa qualité de professionnel en industrie. Le deuxième est un *post yoomaneo* pour un projet adressé au grand public.

- 1) « *On a jamais euh... on a jamais de grosses pannes, franchement on a une industrie qui est relativement simple, on a jamais de pannes qui font qu'on a arrêté euh... la dernière fois qu'on a arrêté c'était l'année dernière, c'était parce qu'il y avait eu une tempête et qu'il n'y avait plus d'électricité nulle part donc euh à vrai dire ça me sert à pas grand-chose euh... le système dans ce cas-là donc euh... C'est très très sophistiqué et je pense que c'est*

¹³ Nous entendons ici seulement les variations relatives aux classes sociales ; comme le souligne à juste titre [Touveron \(2012\)](#), la notion de variation diastratique usuellement utilisée pose problème en ce qu'elle englobe deux phénomènes différents : la distinction entre les « couches sociales » – qui vont impliquer les notions de langue populaire, soutenue, etc. – et les groupes professionnels, qui vont quant à eux impliquer les langues spécialisées. A plus forte raison, dans le cadre de nos travaux, nous ne pouvons traiter ces deux notions de manière indifférenciée, en ce que les vocabulaires relatifs aux langues spécialisées auront une importance certaine, tandis que les niveaux de langue, nous allons le montrer, ne seront en aucune mesure déterminants.

adapté à des sociétés qui ont un impératif de production, de ne pas arrêter les chaînes de production sous peine de perdre de la marchandise, etc. Nous on est loin de tout ça. »

- 2) « *L'utilisateur maîtrise les pannes et perturbations éventuels de [nom du Concept] et du fournisseur internet. C'est un service aide au diagnostic très valorisant pour l'utilisateur qui ne subit plus la panne sans comprendre et sans pouvoir agir, mais en agissant efficacement. »*

Le vocabulaire est donc assez similaire quel que soit le mode de collecte. Il tourne bien sûr autour des notions de pratiques, d'habitudes, etc., mais l'on retrouve également les champs lexicaux du monde de l'entreprise, de l'innovation, du changement. Enfin, on trouve évidemment la présence forte d'un vocabulaire métier et technique dans la majorité des textes.

Au-delà du constat de l'existence et de la force ou faiblesse des variations dans les données, il convient de se demander *si* et *comment* les paramètres sociolinguistiques sont pris en compte voire exploités par l'analyste. La réponse à cette question est que ces paramètres ne sont pas utilisés au moment de la mise à plat. En effet, pour étudier les potentielles différences entre plusieurs groupes de répondants, on utilisera tout simplement une segmentation de ces derniers selon des paramètres pertinents pour l'étude (âge, catégorie socioprofessionnelle, métier, pays, etc.), et l'on étudiera les différences entre les réponses des différents groupes, mais dans une phase suivante à celle de la mise à plat ; au cours de celle-ci, les matériaux sont traités sans tenir compte des caractéristiques sociologiques des individus interrogés. Bien entendu, parfois, un composant de l'identité du répondant peut aider à interpréter un verbatim, mais c'est justement au niveau de l'interprétation et non de la mise à plat¹⁴ que cette connaissance entre en jeu ; si, par exemple, un répondant appartenant au segment *séniors* d'une étude dit à propos du Concept « *non mais c'est pour les jeunes !* », la réussite de la mise à plat nécessite seulement d'identifier le fait que, dans ce verbatim, le participant est en train d'identifier un groupe de personnes comme étant la cible du Concept (et doit donc être classifié dans le Critère 3.1 – « *L'utilisateur comprend-il que le concept le concerne et sait-*

¹⁴ Cf. FIGURE 3 (p. 17).

il à qui d'autre il peut s'adresser ? »). Mais, le fait de savoir si ce groupe de personnes est celui dont il fait partie ou non n'est utile que pour la classification en freins / motivations / conditions et pour de l'interprétation. Bien entendu, lors d'une réalisation d'analyse par un cerveau humain au cours de laquelle l'analyste sait qu'il est en train de travailler sur des données collectées auprès de personnes assez âgées, toutes ces opérations (identification de l'appartenance au Critère 3.1, identification du fait que ce soit un frein, et interprétation comme un blocage générationnel) sont réalisées de manière quasi simultanée à la lecture seule du texte. Il n'en reste pas moins que dans l'optique d'une segmentation rigide de ces opérations, imposée par leur implémentation informatique, la mise à plat seule – objet de nos travaux – n'a pas besoin d'informations relatives au profil de la personne – renseignées directement ou déduites à partir d'indices dans le texte – pour pouvoir fonctionner.

Il nous est donc possible d'en conclure que, si des paramètres sociolinguistiques ne sont pas négligeables dans la caractérisation de nos données, et si ces paramètres peuvent servir dans d'autres phases de l'analyse que celle qui nous intéresse dans le cadre de nos travaux, ils n'en constituent aucunement un enjeu central.

4.2.4. Aspects psycho-linguistiques et leur traitement

Des indices permettant de prendre en compte les aspects psycho-cognitifs des réponses des participants sont présents dans les textes mais ne peuvent être – et ne sont – que peu exploités.

Les transcriptions de disfluences dont nous avons déjà parlé peuvent traduire des hésitations qui elles-mêmes indiquent une gêne ou une incompréhension. Cependant, comme nous l'avons évoqué, elles ne sont pas restituées fidèlement et ne peuvent donc être considérées comme un indice à prendre en compte de manière systématique. La compréhension du Concept par le participant est le sujet de tous les Critères du Niveau 1, ce qui signifie que de nombreux autres indices, purement textuels, suffisent à analyser cet aspect, et que la présence d'une hésitation n'a pas un poids particulièrement significatif.

Les Critères du Niveau 3, quant à eux, traitent des valeurs et du rapport à l'identité des participants. On y trouve donc parfois des réactions assez fortes d'un point de vue émotionnel. Cependant, là encore, la force de l'émotion ressentie est déterminée en phase d'interprétation et non pendant la mise à plat. Il est pourtant nécessaire de pouvoir détecter l'expression, par exemple, d'une crainte, même si celle-ci n'est pas verbalisée à

travers le champ lexical de la peur ; le verbatim « *avec ce système-là, il y aura encore plus d'ondes de partout dans la maison* » traduit bel et bien une crainte non explicitement exprimée, et devra être classifié dans le Critère 3.3 – « *Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?* » qui englobe, entre autres, tout ce qui a trait aux peurs relatives à la sécurité. Mais, ce qui permet d'identifier le fait que ce verbatim traduise une peur n'est pas tant lié à une quelconque analyse de la psychologie de l'individu qu'à une connaissance préalable des réactions possibles face aux changements, des attentes et craintes communément partagées par certaines personnes, ainsi qu'à des indices purement linguistiques ; le fait que la mention des *ondes* est souvent liée à une crainte de leurs potentiels impacts sur la santé fait partie de la connaissance métier de l'analyste, et *encore plus* et *de partout* mettent une emphase quasi hyperbolique sur une omniprésence de ces ondes, ce qui tend à confirmer un sentiment de submersion.

Sur cet aspect de l'exploitation de l'émotion exprimée dans l'analyse, on peut se questionner quant à l'utilisation des émoticônes dans les textes collectés en ligne. Sur ce point, précisons que des émoticônes de réaction sont proposées aux participants dans un champ séparé de celui de la réponse par texte ; il s'agit d'un mode collecte adapté à une autre méthode utilisée par l'entreprise, et ces informations sont donc appelées à être traitées séparément de la méthode CAUTIC®. Bien entendu, il est possible que des participants utilisent des *smileys*¹⁵ dans leurs *posts*, mais cela reste à ce jour relativement marginal : l'observation d'un corpus de trois études réalisées sur *yoomaneo* montre que seulement 13% des participants ont utilisé des émoticônes dans leurs *posts*, soit 61 émoticônes pour un corpus d'environ 59 000 mots répartis sur 2 638 *posts* (et ces chiffres descendent à 7% d'utilisateurs pour huit émoticônes seulement si l'on enlève les données d'un questionnaire adressé à des participants de moins de 35 ans seulement). En outre, une part très importante de ces émoticônes est utilisée comme marques de sympathie / de complicité dans l'interaction entre participants et non comme des réactions émotionnelles au Concept¹⁶ et n'apportent donc pas d'information utile pour l'analyse.

Pour ce qui est des données orales transcrites, bien entendu, la perte d'informations liées à la prosodie est inéluctable. Si, en théorie, le matériau écrit est

¹⁵ Nous utilisons indifféremment *smiley* et *émoticônes*, aucune distinction entre les deux termes ne semblant s'imposer (Halté, 2020).

¹⁶ Par exemple : « *Oui c'est sûr, on a tous plus ou moins ce problème d'engorgement du réseau :)*... » ou « *Je suis du même avis que vous :)* ».

amplement suffisant pour les analystes, ces derniers admettent volontiers que le fait d'avoir eux-mêmes conduit la collecte ou bien d'avoir écouté l'enregistrement – ce qui est parfois fait, mais bien loin d'être systématique – peut apporter des indices précieux ; un *c'est super !* peut être teinté d'enthousiasme, mais aussi parfois d'une certaine forme d'apathie, voire de sarcasme, que le contexte peut parfois ne pas suffire à détecter clairement. Cela étant dit, dans le cadre de nos travaux, nous devons nous baser uniquement sur ces données secondaires ; si quelques indications de type [*soupir*] ou [*rire*] peuvent avoir été ajoutées ici et là par le transcripteur pour indiquer un certain ton ou une marque d'ironie, cette pratique n'est pas systématique et ces occurrences sont en nombre trop restreint pour pouvoir en tenir compte. Il nous faudra donc omettre, dans une large mesure, cette nature secondaire de nos données et les traiter comme si elles étaient primaires.

Enfin, comme nous l'avons indiqué à propos des paramètres sociolinguistiques, les informations qui relèvent d'aspects psycho-cognitifs et émotionnels dans les données écrites sont davantage utiles à la classification en freins / motivations / conditions et à l'interprétation qu'à la tâche qui nous incombe ; pour la classification selon les Critères CAUTIC®, il s'agit de repérer ce sur quoi portent les opinions exprimées, et non leur teneur ni leur polarité. En somme, bien que la problématique abordée par CAUTIC® soit profondément, et par nature, d'ordre psycho-sociologique, et que les données langagières sur lesquelles nous travaillons soient caractérisées par de nombreux aspects de ce même ordre qu'il serait tout à fait intéressant d'étudier, la tâche que nous devons réaliser est bel et bien d'ordre linguistique (et en particulier, nous y reviendrons, sémantique).

Après ce point sur les données qui achève notre présentation de la méthode CAUTIC®, nous en venons à présent aux conclusions quant à la demande d'Ixiade à laquelle notre thèse répond.

5. Conclusions : résumé et points saillants de la demande

L'objet précis de nos travaux est l'automatisation de la tâche d'extraction et de classification des verbatim dans les 20 Critères CAUTIC®. Pour le dire simplement, le système que nous devons construire doit remplacer, autant que faire se peut, l'analyste dans cette tâche, en repérant dans les textes les verbatim pertinents et en les classifiant correctement dans un des Critères. Nous aurons donc un système qui devra prendre en

entrée les contenus textuels intégraux collectés lors des études, et donner en sortie des verbatim extraits de ces textes et associés à un Critère.

Pour cela, nous disposons de données d'apprentissage issues d'études passées et qui consistent en des transcriptions d'oral. Les données cibles – celles que notre système devra être en mesure de traiter – seront principalement des données collectées en ligne (bien qu'il soit souhaitable que notre système puisse également traiter des données orales transcrites). A ce titre donc, les différences entre des deux types de données doivent être prises en compte, et leur impact anticipé.

Une autre caractéristique qui tend à rendre les données hétérogènes est la multiplicité des domaines – et donc des langages – de spécialité qu'ils englobent. Une part importante des études qui sont réalisées par Ixiade s'inscrivent dans le domaine de l'électricité, mais cela ne constitue en aucun cas une généralité, et présente en outre le risque de biaiser l'apprentissage par une sur-représentation de ce vocabulaire. Le système que nous devons construire doit s'affranchir au maximum de la prise en compte du domaine du Concept testé, si bien pour rester en cohérence avec les principes de la méthode CAUTIC® qui se veut d'application transdisciplinaire que pour répondre aux besoins commerciaux de l'entreprise.

Quant à la tâche que notre algorithme devra réaliser, nous y reviendrons plus en détail dans le chapitre suivant, mais nous pouvons d'ores et déjà souligner plusieurs points. Le premier est qu'il ne s'agit pas uniquement de classification de données présélectionnées, mais qu'il faut dans un premier temps être en mesure d'extraire dans le texte ce qui constitue ou non un verbatim avant de le classifier. D'autre part, nous avons une vingtaine de classes, avec la possibilité qu'un verbatim appartienne à plusieurs d'entre elles ; ce nombre important et la prise en compte de cette possibilité viennent complexifier le problème de manière non négligeable. Enfin, et surtout, cette classification n'est pas thématique, pas déterminée par un quelconque principe de polarité comme c'est le cas en analyse des sentiments, mais nécessite bel et bien un certain degré d'interprétation du texte, ce qui constitue le principal point de difficulté.

Enfin, à ces exigences, il faut ajouter celles relatives à la compatibilité pratique entre l'outil et le métier d'analyste. En effet, il va nous falloir proposer un système qui soit relativement rapide d'exécution, ait assez de flexibilité et, surtout, dont l'utilisation ne représente pas une contrainte pour l'analyste ; nous – l'entreprise et nous-même –

envisageons l'outil comme une aide et non un remplacement de l'analyste, dont la compétence et l'intelligence bien humaine demeurent les valeurs ajoutées principales de la prestation d'étude vendue au client. A ce titre, notre outil devra permettre de prodiguer un certain degré d'explication des choix réalisés par l'algorithme afin qu'ils puissent être discutés et réajustés si besoin ; il ne doit pas contraindre les choix de classification en rendant difficile le retour au matériau en cas de désaccord entre l'homme et la machine.

Dans ce chapitre, nous avons explicité les détails de la demande de l'entreprise qui définit les objectifs opérationnels des travaux présentés dans cette thèse. Nous proposons au chapitre suivant de présenter le domaine de l'analyse qualitative outillée dans lequel ils s'inscrivent.

CHAPITRE II : L'ANALYSE QUALITATIVE OUTILLÉE

Après avoir défini précisément la mission confiée par l'entreprise dans le cadre de nos travaux, il convient à présent de prendre un peu de hauteur et de faire un tour d'horizon des réponses que les outils informatiques, et le TAL en particulier, apportent aux besoins des analystes dans le cadre des analyses qualitatives. Nous ne restreindrons pas notre champ d'investigation aux tâches d'annotation et de classification au service de besoins en marketing. Au contraire, nous élargissons ici notre revue de la littérature aux recherches en Sciences Humaines en général et aux outils disponibles pour différentes tâches susceptibles d'être réalisées par des analystes. Ce choix est motivé par le fait que le travail effectué par Ixiade avec la méthode CAUTIC® n'est pas éloigné des pratiques d'analyse qualitative en général : il s'agit de ce que l'on appelle du *codage* de données textuelles, i.e. identifier des passages de texte et leur assigner des catégories. Aussi, restreindre nos recherches aux outils destinés spécifiquement à l'industrie ou encore à l'analyse d'entretiens dans un champ particulier nous semble peu pertinent, en ce que ce chapitre se donne pour ambition d'illustrer ce dont les analystes disposent en termes d'outils de manière générale pour leurs recherches. Nous mettrons cependant l'accent sur l'étape de l'analyse et non sur les autres problématiques de traitement des données afin de ne pas nous éloigner de notre sujet (par exemple, nous ne développerons pas les questions de l'échantillonnage ou de l'agrégation des données).

L'objectif de ce chapitre étant de comprendre comment l'informatique peut aider l'analyste, nous le divisons en deux parties, en présentant en premier lieu ce que sont les analyses qualitatives et quels sont leurs besoins, puis nous proposons un inventaire des solutions existantes pour y répondre.

6. Les analyses de données qualitatives et leurs besoins

L'analyse qualitative est un procédé très largement employé et commenté notamment en sociologie, marketing et sciences de gestion. Dans ce point, nous présentons tout d'abord la diversité des approches qui se cachent derrière ce terme. Puis, nous nous concentrons plus précisément sur la pratique du codage et ses difficultés. Enfin, nous commentons la dimension interprétative et subjective au cœur de l'analyse qualitative.

6.1. La diversité de l'analyse qualitative

Les différents types de recherches basées sur des données qualitatives sont abondamment documentées dans la littérature, sans qu'un réel consensus quant à leur taxonomie ne semble réellement atteint, notamment dû à l'évolutivité constante des pratiques dans le domaine (Voynnet-Fourboul, 2012). Nous ne tenterons pas ici de dresser une typologie précise, mais plutôt de fournir au lecteur un aperçu de la diversité rencontrée dans ce domaine.

6.1.1. Les différentes approches de l'Analyse des Données Textuelle (ADT)

Si le terme *études qualitatives* peut désigner des approches qui utilisent parfois des données extratextuelles (observations de situations, prise en compte du non-verbal, de codes sémiotiques, etc.), on peut tout de même admettre une relation quasi synonymique entre l'*analyse des données qualitative* (ADQ) et l'*analyse des données textuelles* (ADT). Ces deux termes désignent en effet l'ensemble des « approches, méthodes et outils informatiques qui visent à découvrir l'information contenue dans un corpus textuel » (Boughzala et al., 2014, p. 91). Cette quasi-synonymie renferme cependant un paradoxe notable ; dans la très classique opposition entre approches qualitatives et quantitatives, une partie de l'ADT repose sur une inclinaison très franche à la quantification – notamment de par l'utilisation fréquente d'outils informatiques d'aide à l'analyse que nous évoquerons par la suite – et vient, sinon insérer une part de quantitatif dans le qualitatif, *a minima* reproduire cette opposition (Rihoux, 2017).

Text analysis approaches	Assumptions			Validity checks	Examples
	Research method	Nature of text	Role of researcher		
Positivist	Identification of nonrandom variation	Objective	Outsider	Quantitative	<ul style="list-style-type: none"> • Content analysis • Verbal protocol analysis • Script analysis
Linguistic	Study language-reality relationship	Emergent	Outsider	Primarily qualitative	<ul style="list-style-type: none"> • Speech act analysis • Discourse analysis
Interpretivist	Analyze the cultural influences of the writer or speaker and the interpreter	Subjective	Insider	Qualitative	<ul style="list-style-type: none"> • Hermeneutics • Intentional analysis

FIGURE 4 : Framework de l'analyse textuelle selon Lacity & Janson (1994, p.140)

L'ADT englobe plusieurs approches pour lesquelles les chercheurs proposent des typologies différentes qui illustrent bien la diversité des postulats et méthodologies qu'elle peut englober. [Lacity & Janson \(1994\)](#), par exemple, basent leur classification selon les postulats sous-jacents adoptés à l'égard des données textuelles, et identifient alors trois classes illustrées dans la FIGURE 4¹⁷. [Fallery & Rodhain \(2007\)](#), quant à eux, distinguent quatre types d'analyses : l'*analyse lexicale* qui cherche à répondre à la question « De quoi parle-t-on ? », l'*analyse linguistique* pour « Comment parle-t-on ? », l'*analyse thématique* pour « Comment interpréter un contenu ? » et enfin la *cartographie cognitive* pour « Comment structurer une pensée ? ». Selon les auteurs, la première, l'analyse lexicale, consiste dans les grandes lignes à inventorier le lexique du corpus par calcul statistique des fréquences et regroupements en classes. L'analyse linguistique étudie quant à elle les connecteurs, la progression thématique et les marqueurs de formes, et prend comme unité de base la proposition grammaticale (sujet, verbe, prédicat). L'analyse thématique consiste à coder des extraits de texte avec des catégories thématiques découvertes au fur et à mesure de la lecture. Enfin, la cartographie cognitive représente les idées et leurs interconnexions par des graphes. Les auteurs précisent que les deux premières placent l'analyse statistique *ex-ante*, tandis qu'elle est *ex-post* dans les deux dernières. [Paillé & Mucchielli \(2016\)](#) parlent de *stratégies* qui précèdent les méthodes et identifient :

- L'analyse en mode écriture : l'analyste procède par réécritures, reformulations du matériau.
- L'analyse par questionnement analytique : on tente de répondre à des questions grâce au corpus.
- L'analyse phénoménologique : lectures des données d'entretiens qui se concentrent sur les expériences et événements en réduisant au minimum les interprétations *a priori*, réduction du contenu sous forme d'énoncés et productions de récits phénoménologiques à partir de ces dernières.
- L'analyse thématique : repérage et regroupement des thèmes du corpus.

¹⁷ Les postulats sont expliqués ainsi par l'auteur : 1) dans l'approche positiviste, le langage correspond à une réalité objective, 2) dans l'approche linguistique, le langage est un acte qui façonne la réalité et la fait émerger, 3) dans l'approche interprétativiste, le sens véhiculé par le langage est subjectif (i.e., le locuteur, l'interlocuteur et l'observateur conféreront des sens différents). Ainsi, le chercheur adopte des postures d'*outsiders* dans le cadre des deux premières approches, et d'*insiders* dans le cas de l'interprétativisme.

- L'analyse à l'aide des catégories conceptualisantes : recours à des catégories qui ne sont pas thématiques mais des descriptions de phénomènes¹⁸.

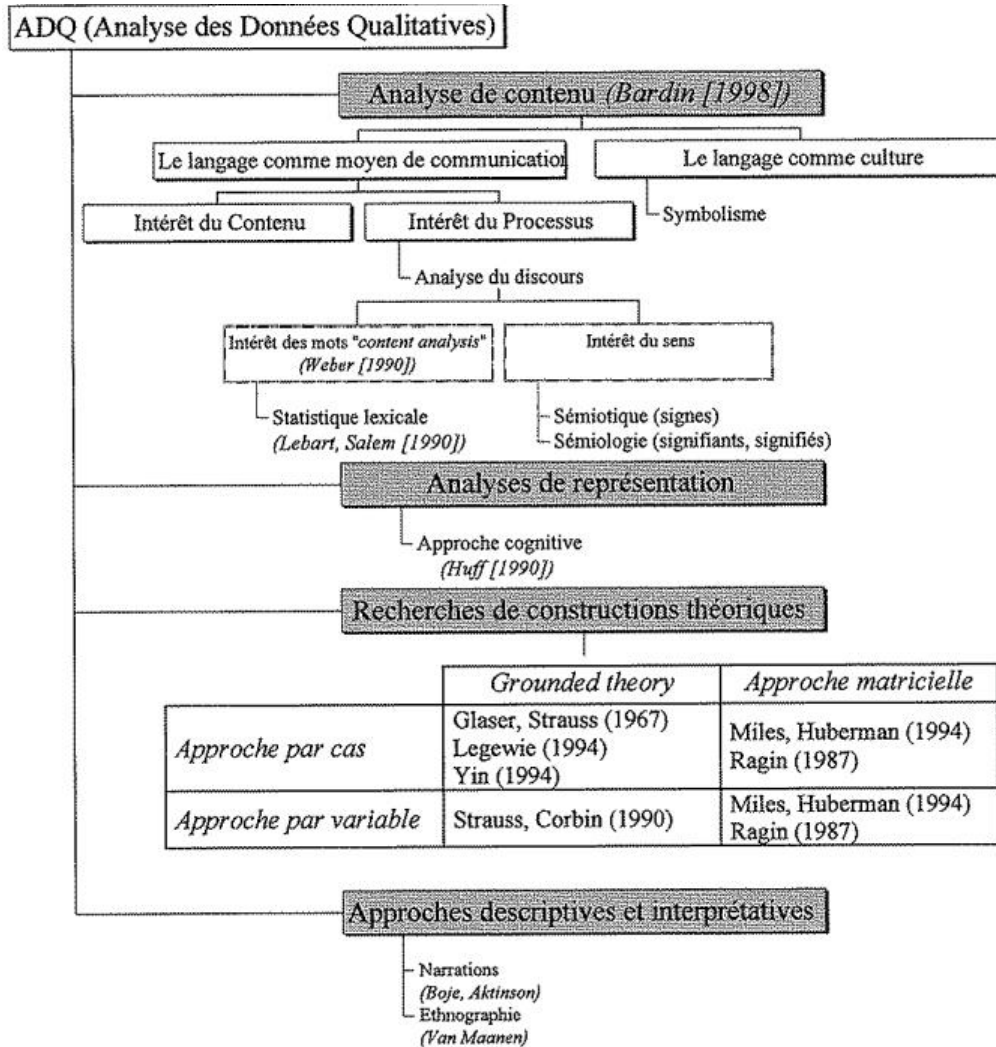


FIGURE 5 : Méthodes pour l'analyse des données qualitatives – Figure de Voynnet Fourboul & Point (2001, p. 1448)

¹⁸ « Par exemple, j'analyse des transcriptions verbatim de patients atteints de cancer en phase terminale qui me livrent leur expérience de la souffrance en lien avec le système de santé (Daneault et alii, 2010 ; Daneault et alii, 2001b). Ces patients me décrivent l'environnement des soins médicaux comme un lieu de combat : leur propre combat contre la souffrance générée par le manque d'écoute du système, mais aussi le combat des médecins contre la maladie et celui de leurs proches en lien avec l'issue fatale de leur maladie. Dans la marge des transcriptions, je génère une première catégorie d'analyse : *Combat généralisé*. Puis une deuxième va s'imposer progressivement : *l'Arène des soins*. » (Paillé & Mucchielli, 2016, Chapitre 12). Les références données dans cette citation sont : Daneault, S., Dion, D., Sicotte, C., Yelle, L., Mongeau, S., Lussier, V., Coulombe, M., & Paillé, P. (2010). Hope and noncurative chemotherapies : Which affects the other? *Journal of Clinical Oncology*, 28(13), 2310-2313. Daneault, S., Mongeau, S., Lussier, V., Hudon, E., Paillé, P., Dion, D., & Yelle, L. (2001). The burden of suffering: Should palliative care be the sole berrer? *7th Congress of the European Association for Palliative Care*, Palerme, Italy

Enfin, Voynnet Fourboul & Point (2001) proposent une représentation schématique – peu commentée dans le texte mais qui a le mérite d'être assez claire – des méthodes d'analyse qualitative¹⁹ que nous montrons dans la FIGURE 5. Les auteurs distinguent donc quatre types principaux : les analyses de contenu, les analyses de représentation, les recherches de construction théorique et les approches descriptives et interprétatives.

6.1.2. L'analyse de contenu

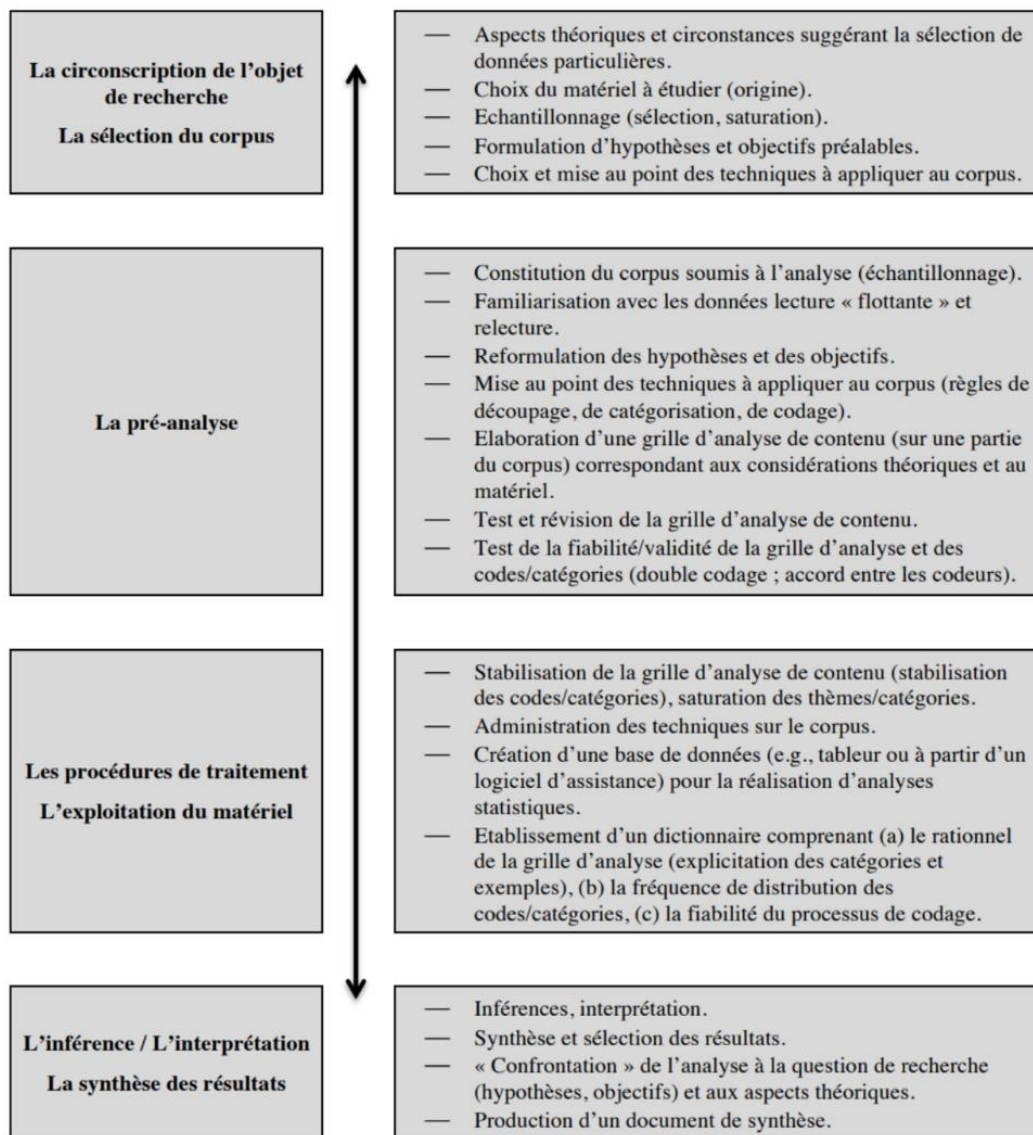


FIGURE 6 : Etapes des analyses de contenu – figure de Dany (2016), à partir des travaux de Bardin (1977), Bauer (2012), Robert & Bouillaguet (2007) et Schreier (2014)

¹⁹ Cette figure est présentée par les auteurs pour illustrer les approches appropriées pour l'analyse qualitative en sciences de gestion mais est tout à fait valable pour d'autres domaines.

La notion d'*analyse de contenu* est souvent citée dans la littérature, et il existe un certain flou, tant terminologique qu'épistémologique, autour de cette expression ; Voynnet-Fourboul (2012) entreprend de faire le point sur ce qu'est l'analyse de données qualitatives et entend mettre en avant l'opposition entre l'analyse de contenu et ce qui relève de la construction théorique (qui s'appuie sur la *Théorie Ancrée* (Glaser & Strauss, 1967)). L'auteure rappelle à cette occasion que l'analyse de contenu, en français, désigne un ensemble de méthodes assez hétérogènes, tandis que *content analysis*, en anglais, correspond souvent à ce qui est appelé en français *analyse lexicologique* (par conséquent, nous utiliserons le terme anglais, dans la suite, pour le différencier de l'analyse de contenu lorsque nécessaire). En français, donc, l'analyse de contenu doit être comprise comme « un ensemble de techniques d'analyse des communications visant, par des procédures systématiques et objectives de description du contenu des énoncés, à obtenir des indicateurs (quantitatifs ou non) permettant l'inférence de connaissances relatives aux conditions de production/réception (variables inférées) de ces énoncés » (Bardin, 1977, p. 43)²⁰. Ces méthodes sont souvent articulées autour de quatre étapes, présentées dans la FIGURE 6, qui sont : la sélection de l'objet de recherche et du corpus, la pré-analyse, le traitement / l'exploitation du matériel, et enfin l'inférence / interprétation / synthèse du résultat.

6.1.3. Mixité des approches

Comme nous l'avons indiqué, les différentes typologies de types d'analyse que l'on peut trouver dans la littérature permettent de mettre en évidence des grandes catégories d'approches et de techniques mais, dans les faits, les pratiques ne sont ni rigides ni figées, et sont en constante évolution ; depuis quelques années, les méthodes de type analyse des sentiments ou analyses des opinions sont également utilisées.

Certains auteurs militent, en outre, pour des méthodes d'analyses mixtes. Ceci n'est d'ailleurs pas nouveau et, dans les années 1970 déjà, Denzin (1978) plaidait en faveur d'une *triangulation* des méthodes (i.e. la combinaison de plusieurs méthodes pour un même objectif d'étude)²¹ (Flick, 2017). Plus récemment, on peut par exemple citer Moscarola & Boughzala (2016) qui expriment la nécessité d'avoir recours à un mélange

²⁰ Très fréquemment citée, par exemple par Dany (2016) et Negura (2006).

²¹ Ce concept, assez large, englobe également la pratique de différents modes de collectes (pas seulement d'analyse), et a beaucoup été associé à la coexistence de méthodes qualitatives et quantitatives, bien qu'il ne désigne pas originellement uniquement cela.

de méthodes traditionnelles (propriétés lexicales) à des approches sémantiques *ad hoc* et à de l'analyse des sentiments. Desmarais & Moscarola (2004) proposent quant à eux d'effectuer un contrôle par analyse lexicale des extraits de discours codés par une analyse de contenu. Enfin, l'étude de cas de Langridge & Ahern (2003) porte sur l'étude des préoccupations des différentes parties-prenantes dans le contexte de la spécialisation avancée en soins infirmiers ; les auteures déclarent qu'une première analyse phénoménologique s'est révélée insuffisante pour répondre aux questions de recherche et a dû être complétée par une analyse du discours.

6.2. Le codage en question

Bien qu'au cœur de l'analyse qualitative, le processus de codage est relativement peu commenté dans la littérature et il existe une certaine confusion quant à la stabilité du sens attribué par les chercheurs à ce terme (Elliott, 2018). Nous proposons les deux définitions suivantes qui nous semblent, dans leur complémentarité, bien refléter l'essence de la pratique :

« Le codage est un processus d'analyse des données textuelles qualitatives qui consiste à les décomposer pour voir ce qu'elles révèlent avant de les reconstituer d'une manière qui fasse sens. » (Creswell (2015, p. 156), cité par Elliott (2018))²²

« Dans le cadre d'une enquête qualitative, un code est le plus souvent un mot ou un syntagme qui attribue symboliquement un attribut synthétique, saillant, captant l'essence et/ou évocateur pour une partie des données linguistiques ou visuelles. [...] La portion de données à coder [...] peut être d'une ampleur variable, allant d'un seul mot à une phrase complète ou encore à une page entière de texte ou un flux d'images animées. » (Saldaña, 2009, p. 3)²³

²² Nous traduisons : "Coding is the process of analyzing qualitative text data by taking them apart to see what they yield before putting the data back together in a meaningful way."

²³ Nous traduisons : "A code in qualitative inquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data. [...] The portion of data to be coded [...] can range in magnitude from a single word to a full sentence to an entire page of text to a stream of moving images."

Le codage a généralement pour lignes directrices les notions d'identification, d'extraction et de décontextualisation de l'unité traitée (i.e., on cherche à faire en sorte que, sortie de son contexte, l'unité soit sémantiquement indépendante). Le second principe commun dans le codage est l'attribution d'une catégorie à cette unité. Au-delà de ces grands principes, cette pratique peut revêtir des aspects divers sur plusieurs points, qui vont impliquer, dans une logique d'informatisation, des procédés différents.

6.2.1. Longueur des segments

Le codage peut être réalisé sur des segments plus ou moins longs, allant du simple syntagme au texte entier, sans que des recommandations claires issues de la littérature ne viennent en aide à l'analyste qui se demande quelle est l'unité d'analyse la plus appropriée (Campbell et al., 2013) ; dans l'optique d'une automatisation, la tâche informatique n'est pas la même. Le codage de textes entiers renvoie au domaine de la classification ou catégorisation de documents, tandis que celui de segments plus courts (paragraphe, phrases, syntagmes), est effectué par des systèmes appartenant à d'autres domaines du TAL que nous développerons par la suite. Quoi qu'il en soit, si cette question de longueur de segments analysés ne représente pas un obstacle majeur pour l'analyste humain – qui identifiera et étiquettera plus ou moins de la même manière un petit syntagme et un paragraphe de quelques phrases en sachant naturellement où doit commencer et terminer l'extrait exploité – il est évident que cette question devient beaucoup moins triviale lorsqu'il s'agit d'automatiser la tâche, et ce, pour deux raisons. La première est qu'un système automatique requiert souvent une segmentation *a priori*, mais ce problème peut être contourné. La deuxième, plus gênante, concerne l'évaluation des performances. La question de l'*unitization* (Campbell et al., 2013; Krippendorff, 1995) concerne les problèmes de longueur et de découpage des segments considérés comme porteurs d'un sens particulier qui peut différer d'un codeur à l'autre, voire créer des chevauchements. Ce problème vient compliquer l'évaluation de l'accord inter-juge, et par conséquent, l'évaluation du système, dans la mesure où l'on peut se demander à partir de quelle longueur et selon quels critères une portion de texte identifiée par le système est identique ou différente d'une annotation humaine qui aurait, par exemple, inclus cette portion dans un segment beaucoup plus long ou au contraire n'en aurait extrait qu'une partie.

6.2.2. Codage inductif vs codage déductif

La liste des étiquettes utilisées dans le codage peut être établie selon deux modalités : soit les catégories sont définies au fur et à mesure de la découverte des textes, soit une grille de catégories est définie au préalable. On parlera alors, dans le premier cas, de *codage inductif* (émergent) et, dans le deuxième cas, de *codage déductif (a priori)* (Elliott, 2018). Là encore, ces pratiques se traduisent en tâches informatiques de natures différentes ; le codage déductif correspondra à des approches supervisées, tandis que le codage inductif pourra correspondre à des approches non-supervisées, bien que ce parallèle ne soit pas parfait (Wiedemann, 2016, pp. 31;36).

6.2.3. Sémantique des codes

Si la pratique du codage peut se résumer de manière simple à attribuer une étiquette à un segment de texte, cette étiquette peut bien évidemment être de plusieurs types en fonction de la nature et/ou du niveau d'abstraction de ce qu'elle désigne. Ici encore, un réel consensus n'est pas atteint, mais des réflexions et typologies sont dessinées.

Paillé & Mucchielli (2016) proposent plusieurs types d'étiquettes : la *rubrique* lorsque l'on annote le sujet sans renseigner ce qui est dit à son propos, le *thème* qui renseigne sur le sujet et sur la teneur de ce qui est exprimé, l'*énoncé* qui est une brève synthèse de ce qui est annoté, le *code* qui indique la rubrique, le thème ou l'énoncé mais dans un format abrégé et peu compréhensible, et enfin la *catégorie conceptualisante* – issue de la tradition de la Théorie Ancrée – qui désigne un phénomène. Les auteurs illustrent cette typologie avec l'exemple (pp. 19-22) d'un extrait dans lequel une femme s'exprimerait à propos de la conciliation plus ou moins difficile de sa vie de famille, professionnelle, amoureuse et sociale ; on pourrait attribuer à cet extrait des rubriques comme *Vie familiale* avec des sous-rubriques comme *Difficultés attachées à la vie familiale*, des thèmes comme *Soins des enfants* et des sous-thèmes comme *Soins difficiles des enfants*, des énoncés comme « La participante dit qu'elle n'arrive plus à joindre les deux bouts concernant le travail qu'exige le soin de ses enfants », des codes comme *Diff. V. F.*, et enfin des catégories conceptualisantes comme *Surcharge parentale*.

D'autre part, Elliott (2018) rappelle que lorsque la pratique du codage consiste à identifier dans les données les textes qui sont des réponses à des points de recherche (l'auteure explique que l'on peut, par exemple, chercher tous les éléments répondant à la

première question posée aux interviewés, tout au long de l'interview et non seulement dans une logique chronologique, ce qui est exactement ce que nous cherchons à faire dans nos travaux), certains auteurs préfèrent parler d'*indexing* (Richards & Richards, 1994) ou de *labelling* (Rapley, 2011).

Enfin, précisons que l'on oppose généralement les codes *in vivo*, dont la dénomination est directement construite avec les termes présents dans le texte, aux codes *descriptifs*, qui vont consister en la désignation de concepts de plus haut niveau (Marathe & Toyama, 2018; Rietz & Maedche, 2021; Saldaña, 2016, pp. 102-110). Par exemple la phrase « *Les transports en commun, par ici, sont un vrai cauchemar.* » pourrait donner lieu à un code *in-vivo* de type *transports cauchemar* alors qu'un code descriptif pourrait être *mobilité*.

6.3. Dimension interprétative et subjective

La confrontation avec les données réelles, qui constitue l'essence même de l'analyse qualitative, implique nécessairement des écarts importants entre la méthode telle que théorisée et la pratique ; Marshall (2002), à propos du codage, oppose les *explications des manuels* et la *réalité désordonnée*²⁴ des données. Nous avons introduit dans le précédent chapitre la nécessité d'une part d'interprétation dans le processus de codage de la méthode CAUTIC®. Cette nécessité, loin d'être une spécificité de cette méthode, est inhérente à la démarche qualitative dans laquelle cette capacité d'interprétation du chercheur est souvent centrale. Comme nous l'avons déjà abordé, la recherche qualitative a longtemps été l'objet de critiques qui visaient son manque d'objectivité et donc de validité. La dimension interprétative est en effet porteuse des difficultés inhérentes que sont le manque de reproductibilité, les biais de subjectivité, la difficulté à généraliser les résultats, et le manque de transparence et de traçabilité du codage (Kaufmann, 2014, p.17)²⁵.

Une certaine systématisation des pratiques et un positionnement plus clair ont permis à l'analyse qualitative, au fil du temps, d'affirmer sa légitimité. Il n'en reste pas moins qu'on ne peut s'attendre à une totale rigidité dans la manipulation et l'interprétation des données qui serait garante d'une absence totale de subjectivité de la

²⁴ Nous traduisons *textbooks accounts* et *messy reality*.

²⁵ Résumant les travaux de De Ruyter & Scholl (1998), Martin & Stenner (2004), Mays & Pope (1995) et Neale et al. (2005).

part de l'analyste ; la nature même de l'exercice implique un certain degré d'interprétation des textes. En outre, les évaluations inter-juges des tâches de codage sont – en grande partie pour des questions de gestion des ressources – loin d'être systématisées, en particulier dans les études dont le matériau consiste en des entretiens (Campbell et al., 2013; MacPhail et al., 2015). Elles sont surtout documentées dans le champ spécifique du *content analysis* (O'Connor & Joffe, 2020) et sont même considérées par certains comme antithétiques avec la dimension interprétative de l'analyse qualitative (*Ibid.*).

Après ce tour d'horizon visant à définir l'analyse qualitative des données textuelles et ses grandes problématiques, nous présentons à présent les outils existants pour répondre à ses besoins.

7. Solutions pour outiller l'analyse qualitative

Nous abordons à présent la pratique outillée de l'analyse qualitative à travers les solutions qu'offrent l'informatique et le TAL en particulier. Nous ferons pour cela dans un premier temps un bref point sur le TAL dans les sciences sociales et l'industrie, puis nous proposerons un état des lieux des solutions utilisées, en présentant distinctement les outils désignés comme CAQDAS (*Computer-Aided Qualitative Data Analysis Software*) et les approches davantage automatisées.

7.1. Le TAL dans les Sciences Sociales et l'industrie

Le TAL et le *text mining* sont aujourd'hui en plein essor dans la recherche et dans l'industrie. Leur potentiel est perçu depuis longtemps : il y a une vingtaine d'années, Tan (1999) – parmi d'autres – estimait que le *text mining* avait un potentiel commercial plus fort que celui du *data mining* en général, dans la mesure où 80% des informations d'une entreprise correspond à du contenu textuel. Aujourd'hui, les technologies du TAL sont implantées dans une vaste diversité de domaines industriels et de services (médecine, finance, etc.) et représentent un marché mondial de 11,6 milliards de dollars²⁶.

Concernant les Sciences Sociales, les CAQDAS et les outils du Web sémantique et du TAL ont permis d'étendre les méthodologies de l'analyse qualitative et, dans une

²⁶ <https://www.globenewswire.com/en/news-release/2021/03/02/2185147/28124/en/The-Worldwide-Natural-Language-Processing-Industry-is-Expected-to-Reach-35-1-Billion-by-2026.html> (consulté le 22/07/2021).

certaines mesures, de l'automatiser grâce à des outils quantitatifs. Cette démocratisation de l'usage d'outils est relativement récente et s'est imposée graduellement depuis deux décennies, et ce, malgré le fait que ces outils existent depuis longtemps. Par le passé, les chercheurs travaillant sur des analyses qualitatives y voyaient une menace de détournement de leurs études vers des méthodes quantitatives. D'autre part, la légitimité scientifique des analyses qualitatives étant déjà souvent questionnée dans la communauté de par leur part de subjectivité, l'utilisation d'outils informatiques à cet égard a eu également du mal à être reconnue ; à la fin des années 90, [Jenny \(1997\)](#) rapporte²⁷ que ses collègues étrangers – non français – « se plaignent d'une insuffisante reconnaissance par les autorités académiques, voire du mépris implicitement affiché à l'encontre de toute activité logicielle, réduite au rang d'assistance technique subalterne et a priori suspecte » (p. 65). L'auteur sous-entend d'ailleurs que la situation en France n'est pas très éloignée. En outre, le recours à ces outils n'était pas perçu comme neutre en termes d'impact sur la méthodologie, la réflexivité du chercheur et les résultats obtenus ([Marshall, 2002](#); [Woods et al., 2016](#))²⁸. Si leur usage est maintenant accepté, cela est non seulement dû au volume important de texte aujourd'hui disponible à analyser, mais également à la réduction toujours plus importante de l'écart épistémologique entre la manière dont les chercheurs en analyse qualitative perçoivent l'objet de leur recherche et ce que les algorithmes sont capables d'identifier ; [Wiedemann \(2013\)](#) analyse cette évolution comme ayant pour facteurs clefs la possibilité d'extraction de sens et l'inclusion du contexte dans les modèles d'analyses linguistiques computationnelles.

En outre, isolément du contexte qui nous intéresse, les tâches que doit accomplir le TAL (repérer, extraire, classifier) pour parvenir à reproduire le codage sont largement étudiées et commentées en tant que telles. Cependant, il semble que, vu par le prisme du chercheur en études qualitatives, les travaux qui affrontent les questions de l'utilisation du TAL pour l'analyse qualitative en termes d'apports méthodologiques se font rares, malgré quelques résultats encourageants ([Crowston et al., 2012](#); [Guetterman et al., 2018](#)).

²⁷ Sur la base des travaux de [Mangabeira \(1996\)](#) et [Miles & Weitzman \(1996\)](#).

²⁸ Non que cet impact soit nécessairement négatif.

7.2. Approches et outils

Les outils informatiques d'analyse de texte, analyse de contenu, analyse qualitative, ont été répertoriés à maintes reprises et inclus, à chaque fois, dans des typologies différentes. [Wiedemann \(2013\)](#) propose une typologie (centrée sur les outils dédiés au *content analysis*) qui se donne comme objectif de marquer le progrès des logiciels dont les fonctionnalités sont passées de la possibilité d'observation surfacique de documents au niveau des termes à celle d'effectuer des extractions de sens à partir de structures sémantiques complexes. [Lejeune \(2008\)](#) fait quant à lui reposer sa typologie d'outils selon la manière dont les catégories de l'analyse sont construites ; ceux qui les génèrent automatiquement, ceux qui demandent à l'utilisateur de les construire et ceux qui font appel à des registres²⁹. Dans une autre publication plus récente, il oppose, sur un premier axe, les logiciels qui permettent seulement de réorganiser du contenu à ceux qui ont une couche interprétative et, sur un deuxième axe, ceux qu'il qualifie de « bavards » (i.e. proposant des résultats au lancement, sans intervention de l'utilisateur) et les « muets » (i.e. attendant que l'utilisateur les interroge) ([Lejeune, 2017](#)). Il argumente le fait que les croisements de ces deux axes aboutissent à des positionnements épistémologiques différents et propose la typologie d'outils, ainsi classés, présentée dans le TABLEAU 6. [Roy & Garon \(2013\)](#) proposent une comparaison de plusieurs CAQDAS qu'ils fondent sur le continuum : analyse automatique (logiciel d'analyse lexicométrique), analyse semi-automatique (analyse lexicométrique et catégorisation sémantique basée sur un dictionnaire / thesaurus), et analyse manuelle. Nous proposons, de notre côté, – non dans l'optique de proposer une typologie mais simplement d'organiser notre présentation – de présenter séparément les outils qui entrent dans les catégories CAQDAS, et ceux qui mobilisent du TAL, choisissant ainsi le critère du degré d'automatisation proposé par les outils. Il est bien clair que la limite entre les deux peut parfois être floue, dans la mesure où certains outils CAQDAS proposent des fonctions qui incluent un certain degré d'automatisation. Nous entendons cependant distinguer surtout les outils de type logiciels qui consistent généralement en une assistance manuelle au codage, et les outils du TAL en tant que techniques décorrélées de quelque logiciel

²⁹ Listes d'indicateurs dans les textes analysés qui reviennent dans plusieurs passages et sont susceptibles de révéler un phénomène intéressant.

particulier et davantage destinés à produire du codage et/ou de l'extraction automatique d'informations.

	L'interface réagence le texte sans ajouter d'éléments d'interprétation	L'interface représente des éléments d'interprétation
Bavard	Lexicométrie	Dictionnaires, automates, registres
Muet	Concordanciers	Logiciels d'étiquetage réflexif

TABLEAU 6 : « Verbose et représentation de l'interprétation » dans les logiciels d'aides à l'analyse qualitative - Tableau de Lejeune (2017)

7.2.1. Les CAQDAS

Le sigle CAQDAS désigne des outils destinés à aider l'analyse en recherche qualitative et dont il existe une quantité notable ; NVivo³⁰, Tropes³¹, ATLAS.it³², HyperRESEARCH³³, QDA Miner³⁴, Alceste³⁵, Sphinx Quali³⁶, The Ethnograph³⁷, etc. sont autant de solutions de ce type, gratuites ou payantes, au service de l'analyste. Les premiers CAQDAS remontent aux années 1990³⁸ et sont initialement pensés pour aider à l'analyse d'entretiens transcrits ; on ne peut réellement parler, à propos de ces logiciels, d'automatisation du processus, en ce qu'ils consistent pour la plupart en une aide élémentaire au codage manuel, simple outillage permettant d'effectuer sur un ordinateur ce que l'analyse faisait déjà sur papier (surligner, découper, etc.) (Kuckartz (2010) cité par Wiedemann (2013)). Si certains incluent quelques fonctions supplémentaires (par exemple, recherches de mots, utilisation de dictionnaires, etc.), leur portée reste en majeure partie de l'ordre d'un traitement lexical plutôt que sémantique. En outre, ceux qui incluent des systèmes de classes sémantiques ont souvent l'inconvénient que ces classes soient construites *a priori* et, par là même, entrent en inadéquation avec la

³⁰ <https://ritme.com/software/nvivo/> (consulté le 05/08/2021).

³¹ <https://www.tropes.fr/> (consulté le 05/08/2021).

³² <https://atlasti.com/fr/> (consulté le 05/08/2021).

³³ <http://www.researchware.com/products/hyperresearch.html> (consulté le 05/08/2021).

³⁴ <https://provalisresearch.com/fr/produits/logiciel-d-analyse-qualitative/gratuit/> (consulté le 05/08/2021).

³⁵ <https://www.image-zafar.com/Logiciel.html> (consulté le 05/08/2021).

³⁶ <https://www.lesphinx-developpement.fr/logiciels/analyse-qualitative-sphinx-quali/> (consulté le 05/08/2021).

³⁷ <http://www.qualisresearch.com/> (consulté le 05/08/2021).

³⁸ On peut en réalité faire remonter leur invention aux années 1960 avec The General Inquirer (Stone et al., 1966, rapporté par Wiedemann, 2016, p. 5), mais le développement de nombreux logiciels aura lieu dans les années 1990.

recherche menée. Notons d'autre part que, même lorsque les CAQDAS proposent des fonctions d'analyse quelque peu avancées, une majorité de chercheurs tendent à les considérer tout de même comme des « classeurs électroniques »³⁹ (Fielding & Lee (2002) cités par Marathe & Toyama, (2018, p.2)), notamment dû à un sentiment de difficulté de maîtrise de ces fonctions.

7.2.1.1. CAQDAS et méthodologies sous-jacentes

Les logiciels CAQDAS, sont souvent pensés initialement par rapport à une certaine méthodologie. Leur degré d'émancipation possible par rapport à ces méthodologies sont sujets à débats entre les spécialistes. En effet, ils ont tendance, selon certains, à s'éloigner au fur et à mesure de leurs développements et à devenir relativement agnostiques, comme l'illustre la prise de position suivante :

« La plupart des CAQDAS sont relativement flexibles en ce qui concerne les méthodes de recherche avec lesquelles ils sont utilisés. Les premières versions développées avaient généralement en tête des méthodologies concrètes qui devaient être transposées dans un processus guidé par le programme. Les représentations de données et les fonctionnalités d'analyse dans ATLAS.ti, par exemple, reproduisaient principalement des concepts connus de la méthodologie de la Théorie Ancrée (Mühlmeyer-Mentzel, 2011). Plus tard, alors que les logiciels ont mûri et intégré de plus en plus de fonctions, ils ont perdu leurs relations strictes avec des méthodes qualitatives spécifiques. Bien que les différences soient marginales, les débats sur le logiciel qui convient le mieux à telle ou telle méthode persistent dans la communauté de la recherche qualitative. » (Wiedemann, 2013, p.7)⁴⁰

³⁹ Nous traduisons *electronic filing cabinets*.

⁴⁰ Nous traduisons : "Mostly CAQDAS packages are relatively flexible concerning the research methods they are used with. Early versions developed, usually had concrete methodologies in mind which should be mapped onto a program-guided process. Data representations and analysis functionalities in ATLAS.ti for example were mainly replicating concepts known from grounded theory methodology (Mühlmeyer-Mentzel, 2011). Later on, while the packages matured and integrated more and more functions, they lost their strict relations to specific qualitative methods. Although differences are marginal, debates on which software suits which method best persist in the qualitative research community." La référence citée renvoie à :

Mühlmeyer-Mentzel, A. (2011). Das Datenkonzept von ATLAS. ti und sein Gewinn für « Grounded-Theory »-Forschungsarbeiten. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(3), Article 17.

Au contraire, pour d'autres, le présupposé théorique sur lequel le logiciel est basé conserve une importance majeure, à l'instar de [Lejeune \(2017, p. 205\)](#) qui entend « désamorce[r] l'illusion de la « boîte magique » » et « dénonce[r] le mensonge de l'outil a-théorique et le mirage du logiciel universel ».

Estimant qu'une prise de position sur la question nécessite une connaissance et une pratique approfondies de différentes méthodes sur différents outils, nous ne prononcerons pas d'avis sur ce débat mais estimons qu'il méritait ici d'être signalé.

7.2.1.2. *Principes et fonctionnalités des CAQDAS*

Le principe général des CAQDAS est qu'ils permettent de coder les textes et d'effectuer une exploration de la représentation des classes dans les textes à l'aide de divers modules de visualisation : concordanciers, croisement matriciel des catégories codées avec des variables de segmentation du corpus⁴¹, graphes, nuages de mots, etc. Sur la phase de codage, des logiciels comme NVivo demandent à l'analyste de créer sa grille de codes et de coder le texte à la main, tandis que d'autres comme Alceste analysent le texte et proposent des catégories automatiquement (dans le cas d'Alceste, en effectuant des regroupements de mots ayant des contextes similaires). Ces deux types de logiciels n'ont donc pas les mêmes objectifs, et ne visent pas les mêmes types d'analyse. Les premiers, comme NVivo, sont une assistance au travail réalisé manuellement, et permettent d'accélérer le processus de remontée en synthèse et le comptage ([Wanlin, 2007](#)). Les seconds, quant à eux, visent à obtenir des résultats plus rapides, et des modalités d'exploration spécifiques ; par exemple, Tropes – qui est fondé sur l'Analyse Propositionnelle du Discours et l'Analyse Cognitivo-Discursive et mobilise des ressources sémantiques – permet entre autres d'identifier des « univers de référence » dans les textes, d'obtenir une analyse chronologique.

7.2.2. *Approches automatisées*

Le TAL et la fouille de textes proposent de nombreuses techniques qui peuvent aider à l'analyse qualitative. [Kaufmann et al. \(2020\)](#) répertorient ces techniques en fonction des tâches à réaliser dans un tableau⁴² que nous reproduisons ci-dessous (TABLEAU 7). Cet inventaire considère que pour coder davantage d'occurrences dans les

⁴¹ Par exemple, des variables qui caractérisent les différents locuteurs d'un corpus.

⁴² Tableau présenté une première fois dans le Mémoire de Master du premier auteur ([Kaufmann, 2014, p. 18](#)).

textes (objet des présents travaux), des méthodes à base de règles ou du *Machine Learning* (ML) peuvent être utilisées, que le ML est approprié pour générer de nouveaux codes, et que la génération de métadonnées structurées permet d'améliorer le codage réalisé manuellement⁴³. A ces méthodes sont associés huit types d'outils⁴⁴ : les règles manuelles, l'analyse de patrons, l'extraction de mots clefs, la reconnaissance des entités nommées, les bases de données lexicales, la linguistique statistique, l'analyse des sentiments et l'extraction de concepts. Précisons que cette liste est établie par les auteurs pour présenter quelles techniques *peuvent* être utilisées, et qu'il ne s'agit pas d'une recension de celles qui sont effectivement fréquemment citées dans la littérature portant sur l'analyse qualitative des données textuelles.

Task	Method	Tools
Find new codings	Rule based	Manually written rules tailored to specific codes Pattern analysis
	Corpus based (machine learning)	Keyword extraction Named entity recognition Lexical databases Statistical linguistics Sentiment analysis
Generate new codes	Machine learning	Concept extraction Keyword extraction Sentiment analysis
Enhance manual codes	Generation of structured metadata	Sentiment analysis Named entity recognition
Facilitate translations	Dictation	Voice recognition software

TABLEAU 7 : Techniques du TAL pour l'analyse des données qualitatives – Tableau de Kaufmann et al (2020, p. 754)

Wiedemann (2016, pp. 28-34), quant à lui, dans un ouvrage dédié à la fouille de texte pour l'analyse qualitative en Sciences Sociales, distingue trois types d'approches : 1) les approches lexicométriques, qui mobilisent des outils comme l'extraction de mots clefs

⁴³ La dernière ligne du tableau mentionne la reconnaissance vocale pour la transcription des données orales de type entretiens et *focus groups*. Nous ne nous attardons pas sur ce point car cela fait partie d'une phase amont à l'analyse à proprement parler.

⁴⁴ Pour la raison citée dans la note précédente, nous ne mentionnons pas la reconnaissance vocale.

ou l'analyse des cooccurrences, 2) les approches ML non-supervisées, avec la clusterisation de documents et le *topic modelling* et 3) les approches ML supervisées, avec la classification, la reconnaissance d'entités nommées et l'analyse des sentiments.

Nous présentons, pour notre part, dans la suite de ce chapitre, quelques approches utilisées à des fins d'analyse qualitative et n'avons pas ici pour ambition de lister toutes les possibilités offertes par le TAL ; d'autres approches et outils, notamment parmi ceux cités dans le tableau que nous venons de présenter, seront présentées de manière plus approfondie dans la suite de cette thèse. Nous commencerons ici par présenter les approches *fouilles de texte* en général, puis nous nous arrêterons sur des techniques plus spécifiques : l'extraction de mots clefs, la reconnaissance des entités nommées, l'analyse des sentiments et le *topic modelling*. Précisons que ces méthodes et techniques ne s'excluent pas mutuellement dans la pratique ; elles peuvent être utilisées conjointement, séparément ou dans une suite de chaîne de traitements.

En outre, précisons immédiatement – nous y reviendrons – que si des approches automatiques telles que celles que nous présentons ici sont bien éprouvées et sont utilisées dans le cadre de l'analyse qualitative, nous partageons le constat de [Kaufmann et al. \(2020\)](#) du fait que, étonnamment, les travaux visant à une (semi-)automatisation de la tâche de codage même sont rares (les auteurs ne répertorient qu'une seule publication à ce propos, celle de [Crowston et al. \(2012\)](#)).

7.2.2.1. La fouille de texte pour l'analyse qualitative

Les approches fouille de texte (*text mining* en anglais), qui reposent sur une logique statistique, englobent une certaine diversité de procédés⁴⁵ mais les méthodes employées sont généralement structurées selon une suite de traitements commune ([Antons et al., 2020](#); [Jo, 2019](#), pp. 3-10). Une première phase consiste à agréger les données. Dans une seconde phase, des prétraitements de type tokenisation, lemmatisation, etc., sont appliqués. La troisième phase consiste en une catégorisation (par règles ou par ML) ou clusterisation du contenu. Enfin, la quatrième phase consiste à intégrer les résultats à la recherche. La fouille de texte est assez naturellement considérée comme épistémologiquement compatible avec l'analyse qualitative, aussi bien pour ce qui concerne les approches *content analysis* que celles qui relèvent de la Théorie Ancrée

⁴⁵ Cf. techniques listées par [Wiedemann \(2016\)](#) citées au point précédent.

(Wiedemann, 2013; Yu et al., 2011) ; ce constat peut être considéré comme de plus en plus vrai au fur et à mesure que les technologies du TAL progressent, notamment de par le fait que les algorithmes sont de mieux en mieux capables de saisir le sens du texte en prenant en compte différents niveaux de contexte (Wiedemann, 2016, p. 22). L'objectif de cette approche est de révéler dans les textes des informations nouvelles, implicites, non perceptibles manuellement ; par exemple, Caputo et al. (2016) font émerger et analysent des clusters thématiques issus d'articles de presse italienne de deux périodes différentes pour étudier la question de la représentation sociale du SIDA et son évolution.

Notons que le terme *fouille de texte* regroupe une variété de procédés et techniques, et peut à ce titre être considéré comme assez générique. Aussi, d'autres approches que nous décrirons dans la suite de cette thèse peuvent être considérées comme y appartenant⁴⁶. D'autre part, il nous paraît important de souligner que si la fouille de texte est traditionnellement associée à un niveau lexical, elle tend à avoir de plus en plus recours à la sémantique (Sinoara et al., 2017).

De nombreuses études sont menées à l'aide de la fouille de texte ; la revue systématique de Antons et al. (2020) dans le domaine de la recherche sur l'innovation porte sur 124 publications traitant le sujet. De manière notable, la fouille de texte est particulièrement utilisée dans le domaine de la santé publique / médecine et ses différentes disciplines (voir par exemple la revue systématique de Abbe et al. (2016) pour la psychiatrie ou encore Zhu et al. (2013) pour l'oncologie), mais son utilisation s'étend bien au-delà des quelques domaines cités ; Jung & Lee (2020) recensent des publications mentionnant la fouille de texte, entre 2010 et 2019, dans 171 disciplines académiques. Enfin, elle est aussi beaucoup utilisée dans le cadre de méta-études pour dégager les tendances d'un certain domaine de recherche (*Ibid.*).

7.2.2.2. Techniques et outils du TAL

Nous présentons dans ce point plusieurs techniques et outils du TAL. Les deux premières – l'extraction de mots clefs et la reconnaissance des entités nommées – permettent de réaliser des extractions de segments de texte d'intérêt et constituent à ce titre des outils à intégrer dans des systèmes ou des méthodes d'analyse plus globales. Les deux suivantes, l'analyse de sentiments et le *topic modelling* permettent d'obtenir des

⁴⁶ Par exemple, l'analyse des sentiments et le *topic modelling* sont des techniques de la fouille de textes.

informations sur les textes qui dépassent la simple extraction et peuvent donc servir à assister pleinement, à elles seules, des analyses. Nous ne présentons pas à ce stade tout ce qui a trait à la classification, bien que ces techniques auraient tout à fait eu leur place ici, car ce sujet sera amplement commenté dans la suite de nos travaux.

1) L'extraction de mots clefs

L'extraction de mots clefs vise à obtenir une vue d'ensemble sur les données traitées en identifiant, dans les documents, les termes⁴⁷ importants – les plus représentatifs de ce dont il est question dans ces documents – et en comparant leurs fréquences avec celles issues d'un corpus de référence (Archer (2008), citée par Wiedemann (2016, p. 30)). Elle est souvent utilisée en phase amont d'autres tâches telles que le résumé automatique de texte, les systèmes de questions-réponses ou les analyses de similarité textuelle (Nasar et al., 2019). Cette tâche peut être réalisée par le biais de nombreux algorithmes, dans le cadre d'approches supervisées ou non-supervisées, et est parfois réalisée avec des systèmes basés sur des réseaux de neurones (*ibid.*).

L'extraction de mots clefs est souvent appliquée sur des corpus de presse (Das et al., 2020; Wan & Xiao, 2008), de revues scientifiques (Beliga et al., 2017; Lopez & Romary, 2010; Nguyen & Kan, 2007) ou de *posts* issus du Web de type *tweets* (Ray Chowdhury et al., 2019; Shaban et al., 2017). Les performances de ces systèmes sont relativement faibles par rapport à celles obtenues pour d'autres tâches du TAL (Hasan & Ng, 2014; Z. Liu et al., 2010) ; la très complète et détaillée revue de littérature de Nasar et al. (2019) recense une quarantaine de travaux sur le sujet publiés entre 1999 et 2019, et seuls six obtiennent un score de précision supérieur à 0,6⁴⁸.

2) La reconnaissance des entités nommées

La reconnaissance d'entités nommées consiste à identifier dans les textes des éléments tels que des noms de personnes, de lieux, d'organisation ou encore des dates dans des corpus généraux, ou bien des éléments spécifiques dans des corpus spécialisés, comme par exemple des noms de pathologies et de médicaments dans des corpus médicaux (Goyal et al., 2018). Elle profite donc aux analyses qualitatives qui ont besoin

⁴⁷ Monolexicaux ou polylexicaux (respectivement *keyword extraction* et *keyphrase extraction* dans la littérature anglophone).

⁴⁸ Les scores de rappel (et de f-scores) sont une donnée manquante dans un nombre notable de publications recensées.

de prendre en compte ce type d'informations. Elle peut servir, par exemple, à repérer les données à anonymiser dans un corpus si cela est nécessaire et peut intégrer une chaîne de prétraitements pour les systèmes de résumé automatique, de questions-réponses ou encore de traduction. Cette tâche a été longtemps réalisée par des systèmes de règles et de lexiques et est aujourd'hui principalement assurée par des systèmes de ML, notamment avec des réseaux de neurones.

Aujourd'hui, ces systèmes permettent des performances très satisfaisantes, avec des f-score de l'ordre de 0,85 à 0,90, même si ces performances peuvent varier en fonction de plusieurs facteurs tels que la langue cible, le genre ou domaine du corpus, ou encore la nature des entités nommées visées (Goyal et al., 2018; Yadav & Bethard, 2019).

3) L'analyse des sentiments

L'analyse des sentiments – *sentiment analysis* ou *opinion mining* en anglais – a pour objectif de repérer les sentiments exprimés par les locuteurs dans un corpus. Pour ce faire, est réalisé un étiquetage des opinions qui consiste souvent à assigner au texte une polarité (négative, neutre, positive) de l'opinion exprimée. On a pour cela souvent recours à un dictionnaire de mots auxquels sont assignées ces polarités (*excellent* : positif ; *immonde* : négatif). Le contenu ainsi étiqueté est ensuite classifié en fonction de la prévalence de ces étiquettes, les difficultés de cet exercice étant de déterminer la valeur à attribuer à un texte qui contient du vocabulaire à la fois positif et négatif, mais également de repérer l'ironie (« *On a attendu deux heures pour pouvoir entrer, super, belle organisation !* »). Les approches de ML font fi des dictionnaires et de l'étiquetage, mais se donnent les mêmes objectifs finaux (i.e., classifier du contenu textuel en fonction des sentiments exprimés). Un peu plus fins et ambitieux, les travaux en analyse des sentiments à base d'aspects s'efforcent de reconnaître également à quels attributs de l'objet à propos duquel l'opinion est exprimée la polarité extraite se rapporte ; par exemple, dans « *la chambre était correcte mais la tapisserie est vieillotte* », on cherchera à extraire l'aspect (*la tapisserie*, qui sera peut-être incluse dans une catégorie comme *décoration*) et on lui attribuera une polarité négative. Enfin, la détection d'émotion, autre sous-discipline de l'analyse de sentiment, se donne quant à elle l'ambition d'extraire les émotions précises exprimées (joie, colère, etc.) et pas seulement la polarité.

Ces approches sont surtout utilisées pour analyser des corpus de types retours clients, mais pas seulement ; on recense par exemple de nombreux travaux récents qui

analysent la perception de l'actuelle – à l'heure où nous écrivons ces lignes – crise sanitaire due au SARS-CoV-2 sur les réseaux sociaux (Manguri et al., 2020; Nemes & Kiss, 2021), et des travaux similaires ont été conduits par le passé sur d'autres épidémies (Alamoodi et al., 2020). En outre, si elles sont souvent déployées sur des volumes de textes importants de types collections de *tweets*, d'avis clients, etc., y avoir recours pour analyser des entretiens n'est pas exclu (Y.-W. Chen & Nakazawa, 2017). Elles conviennent, d'autre part, à des recherches portant sur des sujets proches du nôtre ; Weichbroth & Baj-Rogowska (2019) utilisent une méthode d'analyse des sentiments sur des données d'avis sur l'application *WhatsApp* afin de faire émerger des problématiques d'utilisabilité et d'UX.

4) Le *topic modelling*

Le *topic modelling* (ou *topic mining*) désigne une approche qui vise à identifier les thèmes prévalents dans un ensemble de documents et à calculer l'affinité de ces documents avec chacun d'entre eux. Il s'agit d'un domaine du TAL qui s'appuie sur des modèles probabilistes et appartient au champ plus large de l'analyse sémantique. Plus précisément, il s'appuie souvent sur de l'analyse sémantique latente (qui repose elle-même sur l'hypothèse distributionnelle qui stipule que des mots apparaissant fréquemment dans des contextes similaires ont des signifiés proches), et de l'Allocation de Dirichlet Latente (modèle génératif probabiliste) (Blei et al., 2003).

Le *topic modelling* sert donc à identifier ce dont parlent les utilisateurs d'une communauté Web donnée, d'en tirer des tendances, ainsi qu'à choisir les documents les plus pertinents en fonction de la question de recherche traitée. Il est utilisé dans des études de sociologie, pour de l'analyse d'opinions à visée marketing ou de l'analyse du discours dans les médias (Nikolenko et al., 2017). Cependant, il présente quelques inconvénients qui peuvent amener des réticences à l'utiliser pour de l'analyse qualitative de par le fait que, étant généralement non supervisé, il peut faire émerger des thèmes qui ne sont pas ceux qui intéressent le chercheur (une approche semi-supervisée peut être une réponse à ce problème) (*Ibid.*). En outre, de par sa nature statistique, le *topic modelling* est principalement utilisé sur des données volumineuses issues du Web, telles que des contenus de réseaux sociaux (Karami et al., 2020; Mountford, 2018), des évaluations de produits / critiques (He et al., 2020; Hu et al., 2019; Luo et al., 2020), des publications scientifiques et des brevets (Asmussen & Møller, 2019; Suominen et al.,

2017). Peu d'études mentionnent en revanche des analyses par *topic modelling* sur des données récoltées spécifiquement pour l'étude concernée, et les quelques-unes qui le font concernent des réponses ouvertes à des questionnaires en ligne (Chung et al., 2021; Pietsch & Lessmann, 2018) et non des transcriptions d'entretiens.

Si le *topic modelling* est beaucoup utilisé pour des analyses qualitatives, rares sont les études qui tendent à évaluer sa fiabilité pour le domaine (i.e., sa capacité à réellement révéler si les thèmes extraits sont vraiment représentatifs, si tous les thèmes sont bien extraits, ou si la proportion donnée pour les différents thèmes est correcte) ou encore son potentiel de reproductibilité des résultats (Koltcov et al., 2014). Quelques travaux à ce sujet pointent des faiblesses de ces approches, à savoir le manque de cohérence et de stabilité des thèmes extraits (*Ibid.*) ; les auteurs conseillent notamment aux analystes de lancer plusieurs fois les algorithmes pour identifier les thèmes stables entre les différents résultats obtenus. Ces faiblesses peuvent notamment avoir pour conséquences de causer une certaine confusion pour les analystes et, dans une certaine mesure, de réduire les performances générales de systèmes de classification qui s'appuient sur de l'extraction de thème (Agrawal et al., 2018).

Dans ce chapitre, nous avons donc présenté l'analyse qualitative et les outils qui l'accompagnent traditionnellement. Nous avons souligné la diversité des approches et pratiques en analyse qualitative en général et plus spécifiquement celles relatives au codage, tâche par nature interprétative et subjective. Nous avons ensuite présenté les outils logiciels communément utilisés par les analystes, les CAQDAS, puis nous avons mis en lumière quelques techniques du TAL souvent mises au service de l'analyse qualitative.

Nous constatons que, vu par le prisme de l'analyste en Sciences Sociales, l'automatisation de l'analyse est loin d'être une évidence, et que du chemin reste à parcourir pour mettre pleinement le TAL à leur service. D'autre part, il est évident que parmi les solutions communes existantes, aucune ne serait à même de résoudre le problème qui nous est posé dans cette thèse – construire un outil d'autocodage sur mesure pour une méthode particulière – bien que certaines techniques existantes puissent être utiles pour résoudre des points problématiques particuliers.

Dans le prochain chapitre, nous nous reconcentrons sur ce qui sera notre contribution à cet effort d'outillage de l'analyse qualitative en positionnant nos travaux et en présentant les défis qu'ils se donnent pour ambition de relever.

CHAPITRE III : POSITIONNEMENT ET APPROCHE

Ce chapitre propose de préciser le positionnement de nos travaux, leur originalité et leurs problématiques dans le paysage des pratiques de l'analyse qualitative outillée, ainsi que d'introduire la manière dont nous investiguons les possibilités de solutions à apporter pour répondre à ces problématiques. Dans cette optique, nous commençons par expliquer en quoi la demande d'Ixiade soulève des questions singulières et constitue un défi nouveau adressé à la recherche, puis nous présentons la structure des deux prochaines parties de cette thèse à travers l'opposition entre approches descendantes et ascendantes.

8. Originalité de la recherche

Si la recherche que nous menons dans le cadre de cette thèse CIFRE est motivée par un besoin industriel, elle soulève des problématiques sur lesquelles la recherche théorique se penche également et qui sont loin d'être résolues. Dans ce point, nous situons nos travaux et montrons leur originalité par rapport aux recherches sur l'analyse qualitative outillée, puis dans le cadre spécifique des tâches d'autocodage, et enfin du point de vue des exigences opérationnelles, méthodologiques et relatives à l'expérience utilisateur (UX).

8.1. Positionnement de la demande d'Ixiade dans le paysage de l'analyse qualitative outillée

La demande formulée par Ixiade, à savoir la construction d'un outil capable d'aller vers une automatisation de la tâche de codage en Critères de la méthode CAUTIC®, présente certaines caractéristiques bien particulières en termes de positionnement et de difficultés à adresser.

8.1.1. Positionnement et singularités de la méthode CAUTIC®

Contrairement à la grande majorité des méthodes d'analyse qualitative décrites dans la littérature que nous avons présentée, nous ne sommes pas, dans le cas de la méthode CAUTIC®, dans une démarche de recherche visant à de la découverte (de phénomènes sociaux par exemple) ou encore à de la construction théorique, mais dans celle de l'utilisation d'une méthode standardisée utilisée comme outil de diagnostic à

visée de conseils opérationnels et stratégiques (Anjembe, 2014, pp. 62-78). Difficile alors, de par sa nature singulière, de la positionner dans une des typologies d'approches et méthodes présentées au point 6.1.1 (p. 35) du précédent chapitre⁴⁹. Or, ce sont pour ces méthodes et approches qu'ont été conçues les différentes solutions d'automatisation d'analyse qualitative.

Dans l'optique de son automatisation, c'est l'aspect standardisé de la méthode, reflété à travers la grille d'analyse aux codes figés, qui va constituer un des éléments définitoires de la mission qui nous est confiée. En effet, le codage que déploie la méthode CAUTIC® est de type déductif⁵⁰, ce qui n'est pas rare, mais, en général, les chercheurs qui mobilisent ce type de codage établissent une liste de codes *ad hoc* pour chaque étude. La dénomination et la nature des codes employés par la méthode CAUTIC® sont également assez particulières ; les codes sont formulés sous forme de questions qui incarnent chacune un Critère dont il s'agit d'aller chercher dans les données s'il est ou non validé. La structure en Niveaux subdivisés en Critères, quant à elle, ne représente pas une caractéristique très singulière, dans la mesure où il est assez courant d'organiser les codes utilisés dans une structure hiérarchique.

8.1.2. Outiller l'évaluation de l'acceptabilité

Soulignons également la spécificité maîtresse de la méthode CAUTIC®, par rapport aux travaux d'automatisation et/ou aide au codage recensés dans la littérature, qui réside dans le fait que cette méthode soit centrée sur l'acceptabilité. En effet, le TAL est souvent mis au service de problématiques telles que la génération de scénarii de tests d'interface (Fischbach et al., 2020; Madhavan, 2014; C. Wang et al., 2020), ou encore d'analyses de retours utilisateurs orientées UX, avec des approches de type analyse des sentiments, et souvent pour des produits logiciels finalisés *a minima* dans une version Beta (Beniwal et al., 2021; Weichbroth & Baj-Rogowska, 2019). Il n'existe pas en revanche, à notre connaissance, de travaux qui mêlent acceptabilité et TAL selon les modalités suivantes :

- 1) En considérant l'acceptabilité comme un phénomène complexe dont l'étude implique de prendre en compte des dimensions variées.

⁴⁹ En l'absence de similarités évidentes entre la méthode CAUTIC® et les différentes approches présentées, et n'étant pas nous-même spécialiste de questions épistémologiques relatives aux Sciences Sociales, nous nous garderons d'essayer de faire des rapprochements qui pourraient être assez hasardeux.

⁵⁰ Par opposition au codage inductif (cf. point 6.2.2 (p. 42) de cette partie).

- 2) En prévoyant que l'objet testé peut être de n'importe quelle nature du moment qu'il apporte des changements dans les usages : produit, service, transformation organisationnelle, etc., et ce dans n'importe quel domaine d'activité.
- 3) En prenant en compte le fait que l'objet testé peut l'être à n'importe quel stade de maturation, y compris les plus précoces. C'est notamment le cas lorsque l'on se situe au stade de Concept et que les réactions exprimées par les utilisateurs ne consistent qu'en des projections dans l'usage.

La raison évoquée dans le point 1) de cette liste justifie en partie le fait que nous ayons abordé la demande d'Ixiade comme un problème d'autocodage d'analyse qualitative à part entière malgré la nature très opérationnelle, orientée conseils et évaluative de la méthode CAUTIC®. En effet, des problématiques proches de l'acceptabilité sont traitées par le TAL dans le cadre de travaux de type analyse de retours clients. Ces travaux partagent avec les nôtres l'objectif final de fournir à de potentiels commanditaires des évaluations de leurs produits ou services. Il aurait donc pu *a priori* être possible de considérer que notre tâche était de nature similaire et l'envisager à travers le même prisme. Cependant, les réponses qu'apporte le TAL dans ce cadre consistent généralement en de *l'opinion mining* qui fournit comme informations des associations de sentiments à des aspects de l'objet testé. Or, premièrement, cela n'est pas suffisant pour couvrir toutes les dimensions à prendre en compte dans le cadre d'une analyse CAUTIC® et, deuxièmement, notre travail se situe dans une phase de l'analyse qui est en amont de celle qui s'apparente⁵¹ à une analyse des sentiments (i.e., la deuxième phase de l'analyse CAUTIC® qui consiste à classifier les contenus en freins, motivations et conditions).

8.1.3. Tâches à réaliser vs tâches assurées par les outils existants

Parmi les approches et outils présentés au chapitre précédent, il ne semble qu'aucun n'est en mesure de résoudre à lui seul le problème qui nous est posé. Les CAQDAS sont majoritairement des outils d'assistance au codage manuel. Ils ne laissent pas, en outre, la liberté d'intégrer des modèles d'analyse et présentent donc l'inconvénient d'être difficilement personnalisables en fonction des besoins précis du chercheur (N.-C.

⁵¹ Nous simplifions ; la nature de cette tâche est en réalité différente de ce qui est généralement traité en analyse des sentiments car, s'il existe un parallélisme certain entre classification en polarités positive vs négative et classification en motivations, conditions et freins, des différences substantielles existent entre ces deux ensembles de catégories.

Chen et al., 2018). Les quelques-uns qui proposent des fonctions d'autocodage sont assez limités en termes de possibilités : MAXQDA permet de coder les segments de texte qui comportent des mots clefs saisis par l'utilisateur, et NVivo propose une fonction de codage automatique sur la base de thèmes ou de sentiments. Les approches fouilles de texte, et en particulier le *topic modelling*, servent à faire émerger des thèmes, leur distribution et leur relation, ce qui n'est pas l'objet de nos travaux. L'analyse des sentiments n'est pas non plus pertinente pour les raisons évoquées au point précédent.

Après ces clarifications quant au positionnement de nos travaux dans le domaine de l'analyse qualitative outillée en règle générale, nous abordons plus spécifiquement sa place parmi les propositions de solutions d'autocodage qui ont été formulées dans la littérature.

8.2. Positionnement au regard des tentatives récentes d'autocodage

Si l'outillage de l'analyse qualitative n'est pas une pratique nouvelle, il est étonnant de voir que les travaux visant, à proprement parler, à de l'autocodage, demeurent somme toute assez rares ; la recherche s'est principalement concentrée sur les approches de *topic modelling* sur de larges volumes de données (N.-C. Chen et al., 2018; Kaufmann et al., 2020; Marathe & Toyama, 2018). Des travaux édifiants et relativement récents (N.-C. Chen et al., 2018; Marathe & Toyama, 2018) mettent en avant les défis à relever pour parvenir à une adéquation entre les pratiques des chercheurs en Sciences Sociales et les possibilités du TAL et du ML. Ils constituent un appel à des recherches allant dans ce sens.

Nous présentons, dans les lignes suivantes, quelques travaux notables qui constituent des propositions de solutions d'autocodage. Nous abordons ici la question des techniques, architectures et résultats obtenus dans ces travaux en termes de performance telle qu'évaluée dans une logique de tâche à résoudre par une machine. Nous traiterons plus loin (au point 8.3 de cette partie (pp. 72-81)) tout ce qui a trait à l'adéquation entre ces systèmes et les pratiques des chercheurs.

8.2.1. Travaux notables : approches et résultats

Les ambitions, données, approches et résultats des travaux que nous présentons sont résumés dans le TABLEAU 8 et détaillés dans les lignes qui suivent. Les scores donnés dans ce tableau et dans notre commentaire correspondent aux mesures classiques de rappel (r) et précision (p).

	Crowston et al (2012)	Kaufmann et al (2020)	Marathe & Tomaya (2018)	Rietz (2021)
Ambitions	Utiliser le TAL pour le codage de le cadre du <i>content analysis</i>	- Augmenter la reproductibilité et la traçabilité du codage - Etre multi-domaine	- Explorer les possibilités de techniques simples de TAL pour l'autocodage	Créer une solution de codage (avec interface) qui réponde aux besoins et pratiques des chercheurs en Sciences Sociales
Données	Corpus de messages	Entretiens codés	Entretiens codés	Entretiens codés
Unités codées	Syntagmes, phrases*	Paragraphes	Paragraphes contextuels	Selon le choix de l'utilisateur
Type de codes	Codes qui décrivent des pratiques typographiques, des catégories morphosyntaxique ou des actes discursifs, considérés comme indicateurs de catégories plus larges	Codes de type rubriques** organisés en niveaux	Démographiques, in vivo, meta et descriptifs	Codes de type rubriques ou thèmes**
Exemple de codes	<i>Expressing agreement</i> (catégorie <i>Positive politeness</i>)	<i>Trust > Transparency > IP Rights</i>	<i>research vs pedagogy / drawbacks of collaboration with the art / research as creative practice</i>	<i>No oversight of quality of teaching</i>
Techniques utilisées	Règles construites manuellement	Pipeline TAL basée sur de l'extraction de mots clefs et des services ML	Similarité de mots clefs et requêtes manuelles	Règles manuelles et ML
Meilleurs résultats obtenus	r=0,76 - p=0,68***	- Codes de haut niveau d'abstraction : r=0,63 - p=0,5 - Code de tous niveaux d'abstraction : r=0,69 - p=0,3	r=0,83 - p=0,88	r=0,89 - p=0,92

* Jamais clairement explicité par les auteurs mais déduisible à partir des exemples donnés

** Selon la typologie de Paillé & Mucchielli (2016) présentée dans le point 6.2.3 de la partie I de cette thèse

*** Scores globaux que nous calculons à partir de ceux données par code dans la publication

Tableau 8 : Ambitions, données, techniques et résultats de quatre travaux d'autocodage (Crowston et al., 2012; Kaufmann et al., 2020; Marathe & Toyama, 2018; Rietz, 2021)

8.2.1.1. Les approches par règles puis par ML de Crowston et al. (2012)

Crowston et al. (2012) étudient les possibilités du TAL pour outiller les analyses de type *content analysis*. Leurs travaux se basent, du côté linguistique, sur la théorie des langues de spécialité⁵² et, du côté du TAL, sur une approche symbolique. Les données utilisées sont un corpus d'échanges de 797 messages entre des développeurs de logiciels libres. Les codes sont au nombre de 12 et sont répartis en trois grandes catégories (*Emotional expressions*, *Positive politeness* et *Negative politeness*) dont ils sont considérés comme étant des indicateurs. Nous remarquons que certains de ces codes consistent simplement à décrire une pratique typographique ou une catégorie morphosyntaxique

⁵² *Sublanguage Theory*.

(*Emoticons, Capitalization, Inclusive pronouns*), tandis que d'autres impliquent de repérer dans le texte des types d'actes discursifs (*Apologies, Encouraging participation, Disclaimers/self-deprecation*).

Règle :

Premise	<S> (\$it \$anypos) (do VBZ) (seem \$anypos) (\$anywd \$anypos)* (more \$anypos) (\$anywd \$anypos) (th[ae]n \$anypos) (\$i \$anypos) (\$anywd \$anypos)* </S>
Action	==> generic (\$&, 'entity', 'gm', 'agreement', sf(\$1,\$2,\$3,\$4,\$5,\$6,\$7,\$8,\$9));

Explications :

- La première ligne du tableau représente la chaîne de texte à repérer et la deuxième l'action à effectuer (i.e., ajouter l'étiquette *agreement*)
- <S> et </S> représentent les débuts et fin de phrase
- *do, seem, more* sont des lemmes
- *\$it* correspond à la classe sémantique de *it* (qui contient *it, this, that, these*)
- *VBZ* indique que le verbe doit être au présent
- *\$anypos* indique que le mot peut appartenir à n'importe quelle catégorie morphosyntaxique
- *\$anywd* représente n'importe quel mot

Cette règle permet de coder la phrase, donnée en exemple par les auteurs, "*It does seem to be more trouble than I thought at first*" (sic.)

ENCADRÉ 5 : Exemple de règle utilisée dans le système de Crowston et al. (2012)

Le système proposé se base sur un système de règles construites manuellement. Ces règles décrivent des chaînes à repérer par suites d'entrées lexicales, de lemmes ou de classes sémantiques ; l'exemple fourni par les auteurs est reproduit et expliqué dans l'ENCADRÉ 5⁵³. Nous constatons donc qu'il s'agit d'un système dont la construction est chronophage, d'autant plus que les exceptions aux règles doivent être rectifiées à l'aide d'autres règles ; les auteurs expliquent que pour éviter que la règle qui permet de coder *I'm sorry* sous l'étiquette *Apology* ne code également *I'm not sorry*, ils ajoutent une nouvelle règle pour neutraliser le codage dans ce cas particulier de négation.

Les résultats obtenus, sur les données de test, oscillent très fortement selon les codes, allant de $r=0,5$ à $r=1$ et de $p=0$ à $p=1$. Cependant, tous codes confondus, ces résultats sont assez satisfaisants avec $r=0,76$ et $p=0,68$ ⁵⁴. Dans une communication publiée

⁵³ Nous copions à l'identique le tableau contenant la règle (p. 534) et ajoutons son explication à partir de celles fournies par les auteurs dans le corps du texte de l'article.

⁵⁴ Nous calculons ces résultats globaux, non fournis dans la publication, à partir des données qui y sont présentées.

quelques temps avant cet article, Crowston et al. (2010) comparaient cette approche à une approche ML – qu'ils ne détaillaient pas – et obtenaient de moins bons résultats, notamment dû au fait que leur corpus d'apprentissage était de taille relativement restreinte⁵⁵. Plus tard, Crowston prend part à une autre collaboration (Yan et al., 2014) qui vise également à de l'autocodage d'emails de développeurs de logiciels libres⁵⁶. Ils utilisent pour cela une approche qui mobilise un classifieur de type *Support Machine Vector* (SVM) avec lesquels ils testent plusieurs combinaisons de caractéristiques. Les meilleures performances moyennes du système obtenues sont $r=0,87$ et $p=0,068$. La précision est donc très faible mais les auteurs observent que pour les codes dont il existe plus de 100 annotations dans le corpus, elle a tendance à augmenter significativement, atteignant $p>0,40$ ⁵⁷ pour un de ces derniers.

8.2.1.2. Preuve de concept d'autocodage par chaîne de traitements ML de Kaufmann et al. (2020)

Kaufmann et al. (2020), reprenant en substance les travaux réalisés par le premier auteur dans le cadre de son mémoire de Master (Kaufmann, 2014), entendent utiliser le TAL pour augmenter la reproductibilité et la traçabilité du codage dans le cadre de ce qu'ils définissent comme une recherche exploratoire de type *preuve de concept*. Ils utilisent pour cela comme données d'entraînement trois sets d'entretiens préalablement codés⁵⁸. Les codes utilisés sont organisés hiérarchiquement (par ex : *People > Parents*).

Le système imaginé consiste, dans les grandes lignes, à calculer la similarité entre les mots clefs⁵⁹ extraits dans le texte à coder et ceux précédemment extraits dans les données d'entraînement (i.e., listes de mots clefs par code). Pour effectuer ce calcul, les auteurs expérimentent plusieurs méthodes :

⁵⁵ Notons cependant que ces travaux ont été conduits il y a plus d'une décennie, et que des progrès considérables ont eu lieu depuis dans le domaine du ML.

⁵⁶ La grille de codes utilisée est composée de six dimensions : *Social/relationship*, *Task process*, *Task substance*, *Dual process and substance*, *Change behaviors* et *Networking* comprenant chacune entre 1 et 14 codes, pour un total de 35 codes (exemples de codes : *Approval* et *Suggest/Recommend* pour la dimension *dual process and substance*, *Humor* et *Criticism* dans la catégorie *Social/relationship*).

⁵⁷ Les scores par codes ne sont pas fournis en valeurs numériques mais sous forme de graphique et nous ne pouvons donc pas rapporter le score précis.

⁵⁸ Un set de six entretiens issus du domaine de l'ingénierie logicielle, un de 11 entretiens portant sur la satisfaction par rapport à la vie en général, et un dernier de six entretiens sur le bénévolat ponctuel des *community managers* dans le domaine des logiciels libres.

⁵⁹ Extraits automatiquement avec *AlchemyAPI* d'IBM.

- Un simple comptage de mots clefs communs entre les paragraphes du texte et ceux des différents codes. Cette approche est déclinée en deux possibilités : avec les listes des mots clefs extraits ou avec ces mêmes listes mais augmentées par ajout de cooccurrents fréquents des mots clefs dans le *British National Corpus*⁶⁰.
- L'utilisation des scores de pertinence des mots clefs pour sélectionner ceux à utiliser.
- Un calcul de similarité sémantique – avec l'outil DISCO⁶¹ – entre les mots clefs du paragraphe et le nom du code.
- Un calcul de similarité sémantique, également avec DISCO, entre chaque mot clef du texte et chaque mot clef contenu dans les listes.

Les résultats de chaque méthode ayant des scores de confiance hauts sont retenus et utilisés pour coder le texte.

Une première évaluation a consisté en la comparaison des codes produits automatiquement à ceux précédemment étiquetés dans le corpus de test. Les résultats obtenus sont des scores de $r=0,63$ et $p=0,5$. Cependant, les auteurs précisent que seuls les codes les plus hauts dans la hiérarchie (i.e., les plus généraux, par exemple *People*) ont été utilisés ; appliqués aux codes de tous les niveaux, la précision chutait à $p=0,3$, et le rappel était légèrement meilleur ($r=0,69$). Les auteurs attribuent cette chute de la précision au faible nombre de données d'entraînement par code, mais n'excluent pas que le niveau d'abstraction du code puisse également avoir un impact.

Une deuxième évaluation consistait à faire coder deux entretiens vierges à une chercheuse, puis de comparer son codage à celui proposé par chaque méthode testée⁶². L'approche qui a obtenu le meilleur score est celle par calcul de similarité entre les mots clefs du texte et les mots clefs des listes, avec 230 codes acceptés pour deux rejetés.

8.2.1.3. Essais de techniques simples du TAL de *Marathe & Toyama (2018)*

Marathe & Toyama (2018) explorent la question des possibilités de l'autocodage et de la propension des analystes à y avoir recours. Après une étude exploratoire auprès

⁶⁰ <http://www.natcorp.ox.ac.uk/> (Consulté le 01/09/2021).

⁶¹ <https://www.linguatools.de/disco/> (Consulté le 01/09/2021).

⁶² Séparément, puis combinées avec le système de sélection des sorties les plus pertinentes précédemment mentionné.

de chercheurs⁶³, ils proposent un prototype composé d'outils simples de TAL. Les données utilisées pour tester ce prototype sont des entretiens issus d'un projet qui visait à étudier la place de l'art à l'université, et qui ont été codés par les participants dans le cadre de l'étude exploratoire sus-citée. Les codes peuvent être de quatre types : *in vivo*, descriptifs⁶⁴, démographiques (âge, genre, catégorie socioprofessionnelle, etc.) et méta (codes descriptifs mais qui décrivent des phénomènes qui transparaissent au niveau du document dans son entièreté).

Le processus de segmentation des données utilisé est original ; les auteurs, partant du constat qu'une segmentation en tours de parole donnait des segments trop volumineux et que celle en paragraphes avait pour inconvénient le fait de décorrélérer la question posée (qui contient des mots clefs pour coder) de la réponse fournie (qui ne les contient pas forcément), proposent une segmentation en ce qu'ils appellent des « paragraphes contextuels ». Ces derniers consistent à accoler à chaque paragraphe de réponse dans le tour de parole la question posée. Les auteurs précisent que cette technique a considérablement augmenté les performances de leur système.

Le prototype proposé se base sur une représentation vectorielle de type *sacs de mots* et trois méthodes sont testées et comparées. La première consiste à calculer la similarité cosinus entre les paragraphes contextuels et les libellés des codes. La deuxième est similaire à la première mais inclut une augmentation de la liste des mots clefs des codes par des synonymes. Enfin, la troisième consiste à affecter à chaque code une requête composée de mots clefs combinés à des opérateurs booléens. Un exemple de requête complexe présenté par les auteurs pour le code *Art research examples* est « (video OR film OR animation OR (oral AND history) OR folktale OR poem OR poetic OR book OR anthology OR ((idea OR musical OR sketch) AND design) OR graphics) AND (project OR collection OR studio OR workshop OR performance OR (communication AND public)) ».

La troisième approche obtient les meilleurs résultats avec $p=0,88$ et $r=0,83$, scores bien supérieurs à ceux obtenus avec les deux premières méthodes ($p=0,15$ et $r=0,33$ pour la première, et $p=0,18$ et $r=0,43$ pour la deuxième). Sans surprise, ce sont les codes *in vivo*

⁶³ Cf. point 8.3.1.1 (p. 74).

⁶⁴ Cf. §4 du point 6.2.3, p. 43-42 pour l'opposition entre codes *in vivo* et codes descriptifs.

– qui par définition contiennent le même lexique que le texte codé – qui obtiennent les meilleurs scores ($p=0,94$ et $r=0,93$).

8.2.1.4. Système de codage par règles manuelles et suggestions ML avec interface complète de Rietz (2021)

Les travaux de thèse Rietz (2021), qui ont été réalisés sur une période sensiblement similaire aux nôtres, visent à mettre le ML au service de différentes composantes de l'analyse qualitative⁶⁵. Une partie de ces travaux, résumée dans la publication de Rietz & Maedche (2021), consistait à créer un système d'autocodage – appelé *Cody* – basé sur une logique de ML interactif assez complet et intégré dans une interface. L'objectif de ses travaux n'est pas tant de créer un système performant en termes de rappel et de précision, mais plutôt de parvenir à une solution de codage qui entre en adéquation avec les pratiques des analystes.

Cody, dont l'interface est présentée dans les FIGURES EN ANNEXE 1 à 3 (pp. 266-267), comporte deux modalités pour générer des suggestions de codes, une à base de règles et une à base de ML, et fonctionne selon une logique d'*active learning* (i.e., le système apprend les règles de codage au fur et à mesure que l'analyste code). Le principe de la modalité par règles est la suivante : lorsqu'un analyste crée un nouveau code, une suggestion de règle est générée à partir du texte codé⁶⁶. Un exemple de règle, présenté dans les captures d'écran de l'Annexe, est (*pay* OR paid OR money**) AND (*wasn* OR didn't OR no*). Il correspond au code *Motivation: Not financial*, et permet de coder les deux phrases données en exemple : « *I wanted to do the job and I didn't think about the money* »⁶⁷ et « *It wasn't very well paid to start with* »⁶⁸. La règle créée est ensuite immédiatement projetée sur le reste du texte pour coder les segments qui lui correspondent, et l'utilisateur peut voir quels mots clés ont déclenché le codage. Pour la partie ML, Rietz entraîne un modèle de régression logistique par descente de gradient stochastique à partir des données codées et des règles créées. Les suggestions ML sont mises à jour toutes les 10 actions sur des codes (addition, édition, suppression) ou par

⁶⁵ Une partie des travaux porte notamment sur l'utilisation de *chatbots* pour réaliser des entretiens.

⁶⁶ Par utilisation d'un calcul de score de similarité pour générer des synonymes des mots du texte et de distance de Levenshtein pour ajouter des mots proches.

⁶⁷ « Je voulais ce travail et je n'ai pas pensé à l'argent » (nous traduisons).

⁶⁸ « Première chose, ce n'était pas très bien payé » (nous traduisons).

action sur un bouton de rafraîchissement. Toutes les suggestions (par règles ou par ML) peuvent être supprimées.

Une première évaluation formative sur la perception de l'outil est menée et sera commentée plus loin. Une deuxième évaluation, sommative, est conduite pour évaluer l'efficacité du système. Pour cela, Rietz utilise des données d'entretiens codés qui s'inscrivent dans le domaine de l'éducation⁶⁹ et demande à des étudiants familiers avec les analyses qualitatives de les coder avec une grille d'analyse fournie. Les étudiants doivent suivre la grille d'analyse mais peuvent ajouter des codes s'ils le souhaitent⁷⁰. Les résultats sont évalués en termes de rappel et de précision pour les suggestions ML seulement, en comptant comme vrais positifs les codes qui coïncident avec les annotations manuelles et celles produites par les règles. Les meilleurs varient, selon les participants, entre $p=0,61$ et $p=0,92$ et entre $r=0,56$ et $r=0,89$.

8.2.2. Positionnement et ambitions

Ces quelques travaux que nous considérons significatifs ainsi présentés, nous situons à présent dans les nôtres et précisons nos ambitions en termes de résultats.

8.2.2.1. Niveau d'abstraction et complexité sémantique des codes

La spécificité majeure de nos travaux, au regard des études susmentionnées, réside dans le niveau d'abstraction et la complexité sémantique des Critères CAUTIC®. En effet, ces travaux utilisaient majoritairement des codes de types thématiques ou rubriques, pour lesquels le repérage de la présence d'un ou de quelques types de mots clefs dans un segment de texte permet d'annoter ce segment avec une probabilité considérable de réussite.

L'analyse CAUTIC® nécessite, quant à elle, de repérer des catégories conceptuelles plus ou moins abstraites, et qui peuvent appartenir à la connaissance commune ou bien être très spécifiques à la sociologie des usages voire à la méthode elle-même ; les concepts de PROBLÈME et de SIMPLICITÉ, mobilisés respectivement par les Critères 2.1 et 2.4⁷¹, tombent dans la première catégorie, ceux d'IMAGINAIRE D'APPROPRIATION et d'EXTENSIONS

⁶⁹ Harding, J. (2015). *Identifying Themes and Coding Interview Data: Reflective Practice in Higher Education*. SAGE Publications. (Consultable à <https://methods.sagepub.com/dataset/coding-education> (consulté le 03/09/21)).

⁷⁰ L'auteur ne précise pas comment ces codes ont été pris en compte dans les calculs de scores de rappel et de précision.

⁷¹ Pour rappel : 2.1 – « Le concept est-il simple à utiliser ? » et 2.4 – « Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ? ».

D'USAGE des Critères 3.4 et 3.5⁷², dans la deuxième. En outre, ces catégories conceptuelles peuvent avoir des réalisations lexicales très variées qui ne seraient pas nécessairement regroupées ensemble dans une logique thématique ou taxonomique. Enfin, le repérage d'une ou de plusieurs catégories peut ne pas être suffisant dans la mesure où la manière dont ces catégories sont articulées sémantiquement et discursivement peut participer à qualifier ou disqualifier un verbatim pour un Critère. Ces affirmations ne valent pas nécessairement avec la même force pour tous les Critères.

Pour illustrer notre propos, considérons l'un des Critères *a priori* les plus simples de la grille d'analyse, le Critère 2.1 – « *Le concept est-il simple à utiliser ?* ». Ce dernier, si nous imaginons qu'il eût été utilisé dans un des travaux que nous avons présentés, aurait certainement donné lieu à un libellé de code de type *simplicité d'utilisation du Concept*. Les segments à coder auraient été repérés par des mots clefs comme *simple, facile, utiliser*, et par des règles qui ressembleraient à « *(facile OR simpl* OR difficile OR compliqué) AND (utilis*)* ». Cela aurait effectivement permis de repérer les verbatim – présents dans notre corpus d'apprentissage que nous présenterons par la suite – « *Pour moi ça me paraît simple à utiliser.* » ou « *Tel qu'il est présenté là, il a l'air facile d'utilisation autant pour la mise en œuvre que pour avoir les résultats.* ». Cependant, notons premièrement que la règle ne correspond pas exactement à l'intitulé du code en ce qu'elle omet de préciser la mention du Concept, ce qui ne permet pas d'exclure les phrases qui mentionneraient une facilité ou une difficulté d'utilisation relative à un autre objet. On aurait pu alors ajouter à cette règle une liste de mots qui seraient potentiellement utilisés pour désigner le Concept. Mais, dans ce cas, les deux verbatim donnés en exemple n'auraient pas été repérés (aucun des verbatim de notre corpus ne l'aurait d'ailleurs été). En effet, le Concept étant l'objet principal de la discussion, les utilisateurs interrogés ne le mentionnent par un substantif que très rarement. Au-delà de ce problème, et bien plus important, existe celui du fait que, dans la méthode de CAUTIC®, ce Critère doit également permettre d'identifier les verbatim suivants :

- 1) « *ça ne demande pas des compétences extraordinaires* »

⁷² Pour rappel : 3.4 – « *Le concept suscite-t-il un imaginaire d'appropriation ?* » et 3.5 – « *Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?* ».

- 2) « *J'ai vu qu'on n'avait pas besoin d'éteindre le compteur, il faisait ça la lumière allumée pour changer de fonction. Il n'y a pas besoin de tournevis ou de pince, on a besoin de rien... »*
- 3) « *Je m'attendais à quelque chose de plus encastré et protégé. Si j'y vais trop fort, ça ne va pas entrer dans le rail et ça va se casser... »*
- 4) « *Ce n'est pas plug & play, cela va nécessiter une mise à jour »*

Or, ces verbatim ne contiennent pas de mentions explicites ni de facilité ou de difficulté d'utilisation. Lors d'une analyse humaine, ces verbatim sont codés comme appartenant au Critère 2.1 pour des raisons qui vont bien au-delà du repérage des mots clefs appartenant à des classes sémantiques similaires à celles présentes dans l'intitulé du code. Dans le cas du 1), par exemple, cela nécessite le fait de repérer que l'expression *ne demande pas des compétences extraordinaires* permet l'annotation, parce qu'elle contient les notions de NÉCESSITER + COMPÉTENCE, et parce que sa configuration sémantique est telle que l'utilisation du Concept – non verbalisée – et l'élément lexical qui dénote la COMPÉTENCE assument des rôles thématiques particuliers. En d'autres termes, cela nécessite de repérer le fait que l'idée de 'nécessiter des compétences' est en lien avec celle de 'facilité d'utilisation', et que la présence de cette idée, ainsi que la configuration syntaxico-sémantique dans laquelle elle se trouve dans le texte, permettent d'affirmer que l'on a trouvé une réponse à la question formulée par le Critère CAUTIC® ; nous sommes donc loin d'une extraction de mots clefs, en ce que la proximité sémantique de mots isolés ne suffit pas, et que la dimension compositionnelle de l'énoncé est primordiale.

D'autre part, la question des actants non verbalisés, ou bien lexicalisés par des pronoms, que nous avons évoquée est assez épineuse. En effet, dans les exemples 2) et 3), c'est leur sens qui permet de valider l'annotation. Dans le 2), il faut que ce soit le Concept qui fasse en sorte que l'action à réaliser ne nécessite plus d'outil, or, ni le Concept ni cette action ne sont verbalisés. Dans le 3), il faut comprendre que *ça* renvoie au Concept, et la résolution de cette anaphore est compliquée dans la mesure où il est probable que le mot *concept* ou son nom n'ait pas été verbalisé dans les lignes précédentes. La nécessité de la prise en compte d'une dimension pragmatique n'est donc pas négligeable.

Enfin, notons que dans le 4), c'est l'expression *plug & play* qui est rattachée à l'idée de 'facilité d'utilisation'. Ce lien réside dans le fait que cette facilité est un des bénéfices apportés par le type de systèmes désigné par cette expression. C'est donc un lien sémantiquement indirect avec un élément lexical spécifique à un domaine qui permet la classification. Cela indique que l'utilisation de ressources organisées en relations d'hyponymie ne serait pas suffisante pour augmenter une liste d'unités lexicales qui composeraient une règle.

Malgré ces considérations, il est possible d'imaginer que l'on pourrait, dans l'absolu, créer une règle lexicale par cas possible par Critère. Par exemple, pour le 1), nous pourrions créer la règle « (*nécessiter OR requérir OR demander*) and (*compétences OR connaissances*) ». Mais il est évident, au vu de la multitude de configurations possibles par Critère, qu'il s'agirait là d'un travail colossal s'il était réalisé à la main et, si ce travail était réalisé automatiquement à partir d'un corpus, il n'est absolument pas certain que ce corpus intègre une couverture suffisante de cas possibles, ni qu'il ne contienne un nombre notable de cas contradictoires difficilement modélisables.

8.2.2.2. Des codes constants pour des données multi-domaines

Certains codes utilisés par [Marathe & Toyama \(2018\)](#) et [Rietz \(2021\)](#)⁷³ sont un peu plus complexes que des codes de type thématiques ou rubriques. Mais, il faut noter qu'ils ont été construits dans le cadre d'une étude particulière et pour un jeu de données précis, et que les intitulés des codes sont tirés, ou au moins inspirés, par le lexique du texte. Cela n'est pas le cas dans le cadre de CAUTIC® qui consiste à projeter des catégories issues de la sociologie des usages et de la psychologie sur des corpus dont le vocabulaire peut être plus ou moins technique et appartenir à de nombreux domaines de spécialité différents. Ainsi, si le code *drawbacks of collaboration with the arts*⁷⁴ utilisé par [Marathe & Toyama \(2018\)](#) implique de repérer plusieurs catégories sémantiques, ces catégories sont lexicalisables par des ensembles beaucoup plus restreints que ceux qui sont mobilisés par CAUTIC®. Ces ensembles lexicaux sont, en outre, plus facilement repérables dans la mesure où ce code a été élaboré pour étiqueter un ensemble fini d'entretiens thématiquement cohérents. Si, lors d'une étude CAUTIC®, un participant venait à parler

⁷³ Pour rappel, des codes comme *research as a creative process* ou *no oversight of quality of teaching*.

⁷⁴ *Inconvénients de la collaboration avec les arts*. Ce code est utilisé pour « annoter les inconvénients à affronter dans le cadre de partenariat avec des artistes » (p. 348) (nous traduisons "to annotate disadvantages faced in partnering with artists").

de potentiels inconvénients dans sa collaboration avec des artistes, le verbatim serait probablement associé au Critère 4.2 – « *Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur ?* »⁷⁵. Or, ce Critère doit permettre d'identifier de nombreuses autres relations interpersonnelles possibles que celles de collaboration avec des artistes. La liste lexicale utilisée pour identifier les collaborateurs serait donc bien plus importante que dans le cas du code *drawbacks of collaboration with the arts*.

8.2.2.3. Impact sur les objectifs de performance et modalités d'évaluation

Au vu de ces caractéristiques spécifiques, et si l'objectif des travaux est bien entendu d'atteindre les meilleures performances possibles, il apparaît d'emblée qu'atteindre des scores comparables à ceux évoqués plus hauts semble très ambitieux. En outre, si l'idéal semble *a priori* de réussir à construire un système qui coderait aussi bien que l'humain, cette ambition apparaît comme très peu réaliste et n'est d'ailleurs jamais visée ni même souhaitée par les différents chercheurs que nous avons précédemment cités.

Parce qu'elle émane d'une entreprise, la demande qui nous est formulée par Ixiade est avant tout motivée par un désir d'économies de ressources passant par une réduction du temps de codage, avec la condition *sine qua non* que la qualité de l'analyse finale ne soit pas impactée négativement. Nous souhaiterions ainsi atteindre une réduction de ce temps de l'ordre de 50%. Il est cependant évident que la dimension temporelle ne peut pas être la seule métrique d'évaluation que nous utiliserons. Cette évaluation de performance devra dans un premier temps mesurer les écarts entre codage automatique et codage humain, ce qui n'est pas trivial pour plusieurs raisons.

On peut évaluer la validité d'un système d'autocodage selon deux axes. Le premier axe est celui de la performance pure, qui peut être évaluée par rapport à un corpus *gold* préétiqueté avec des scores de rappel et de précision, ou bien en considérant le système comme un codeur humain avec des scores d'accord inter-juge. Le deuxième axe est plus qualitatif et consiste à examiner le système à travers le prisme de l'expérience utilisateur afin d'évaluer sa pertinence opérationnelle ; un système qui obtiendrait de très bons scores de rappel et de précision mais ne permettrait pas, par exemple, de modifier

⁷⁵ Il pourrait aussi être classé dans les Critères 4.3 ou 4.4 selon son contenu précis.

aisément les codes faussement annotés perdrait beaucoup de son utilité et pourrait peut-être même ne pas être utilisé. Nous détaillerons les tenants et aboutissant de ce deuxième axe dans le point suivant.

Pour revenir au premier axe, l'élaboration et l'utilisation d'un corpus *gold* comportent leur lot de difficultés. Nous avons déjà abordé la question de la dimension subjective inhérente à l'analyse qualitative, et cet aspect peut résulter en un accord inter-juge potentiellement faible. Il s'agira donc de prendre en compte ce paramètre dans la constitution d'un corpus d'entraînement. D'autre part, comme l'ont rappelé [Crowston et al. \(2012\)](#) en tenant compte dans leur évaluation d'annotations « inspirantes », le corpus *gold* ne contiendra pas forcément toutes les annotations potentiellement acceptables. Pour ces raisons, une comparaison avec un corpus *gold* est, certes, incontournable, mais il faut tout de même garder à l'esprit qu'elle comporte des limites si on la met en perspective avec la finalité complète des présents travaux.

Dans ce point, nous avons présenté des travaux d'autocodage et avons sur cette base positionné les nôtres en termes de spécificité et difficultés d'analyse et de résultats obtenus. Nous abordons au point suivant la question des exigences opérationnelles, méthodologiques et d'UX qui s'imposent pour le système que nous entendons construire.

8.3. Exigences opérationnelles, méthodologiques et UX : un outil sur mesure au service de l'analyste

L'objectif central de nos travaux réside dans la réalisation d'un système de TAL efficace pour de l'autocodage et les aspects relatifs à l'interaction avec l'interface qui mobilisera ce système en *back-end* sont, tout au plus, périphériques. Précisons donc immédiatement que les considérations de *design* d'interface pures sont hors de la portée de cette thèse, et que ce n'est pas ce dont il est question dans les lignes qui suivent. Il n'en demeure pas moins que l'adéquation entre un système de codage et les attentes et besoins des utilisateurs se doit, à notre sens, d'être investiguée pour que nos choix en matière d'algorithmes, de ressources et de modélisation aillent pleinement dans le sens de cette adéquation. Pour cette raison, nous présentons dans ce point des travaux de chercheurs qui se sont penchés sur la question de la création d'outils qui puissent pleinement profiter aux Sciences Sociales, puis nous faisons le point sur les exigences que nous devons respecter pour aller, nous aussi, dans ce sens.

8.3.1. La volonté de construire des outils (réellement) au service des sciences sociales

Fournir un outil de TAL pour une équipe de chercheurs en Sciences Sociales est un défi profondément pluridisciplinaire et l'exigence d'un aspect sur mesure dans un tel cadre est rarement adressée ; en général, les chercheurs en TAL proposent, pour assister les chercheurs en Sciences Sociales, un ensemble d'outils existants avec lesquels composer, le développement d'une solution *ad hoc* nécessitant de déployer des moyens excessivement importants (Poibeau, 2014)⁷⁶. La situation qui en résulte est celle d'une sorte de compromis permanent qui limite l'intérêt des chercheurs pour des solutions automatisées, dû au fait que les possibilités offertes par certains outils provoquent parfois des résultats qui n'adressent pas directement la problématique de recherche. L'origine de cette inadéquation entre besoins et outils ne tient pas uniquement à un manque de moyens déployés, mais relève également des différences d'orientation entre les disciplines. En effet, s'il existe bel et bien des points communs qui rendent épistémologiquement compatibles certaines approches du TAL et du ML avec l'analyse qualitative (Wiedemann, 2013; Yu et al., 2011), des différences fondamentales entre ces disciplines ne sont pas pour autant négligeables ; par exemple, là où les approches de ML mettent l'accent sur l'obtention de bonnes performances de classification de texte dans des catégories, les chercheurs en Sciences Sociales sont quant à eux intéressés par « l'identification de concepts théoriques complexes, socialement construits et instables, dont les limites sont souvent mal définies »⁷⁷ (Nelson et al., 2021, p. 3). Or, la conciliation entre ces différents intérêts n'est pas une mince affaire.

Depuis quelques années, face aux constats d'un manque de collaboration interdisciplinaire (Wallach, 2016) et de la paucité des recherches sur la question de l'adaptation des outils aux pratiques et problématiques des chercheurs en Sciences Sociales (Marathe & Toyama, 2018), des efforts communs apparaissent pour pallier ce manque. Ces travaux se basent sur les nombreuses critiques adressées aux CAQDAS et déploient des approches très centrées utilisateurs qui mettent davantage l'accent sur des démarches d'exploration des pratiques que sur la question de la performance pure des outils. Les initiatives de N.-C. Chen et al. (2018) et de Marathe & Toyama (2018), que nous

⁷⁶ C'est notamment le cas, en France, comme le rapporte l'auteur, la mission assurée par la Très Grande Infrastructure de Recherche (TGIR) Huma-Num (<https://www.huma-num.fr/>).

⁷⁷ Nous traduisons : "identifying complex, socially constructed, and unsettled theoretical concepts, often with ill-defined boundaries"

présenterons plus en détails dans les lignes suivantes, sont emblématiques de cette nouvelle impulsion adressée aux Sciences de l'Information, et d'autres travaux notables partagent le même objectif ; à titre d'exemple, [Ganji et al. \(2018\)](#) s'intéressent au manque d'intérêt des logiciels de codage pour augmenter l'accord inter-juge (et donc la fiabilité de l'analyse). Les conclusions et observations tirées de ces travaux, que nous détaillons à présent de manière non exhaustive, sont récapitulées dans le TABLEAU 9.

8.3.1.1. Enquêtes de [Marathe & Toyama \(2018\)](#) sur les pratiques des chercheurs et recommandations pour les outils

[Marathe & Toyama \(2018\)](#) entendent pallier le manque constaté de recherches visant à cerner les besoins et pratiques des chercheurs en Sciences Sociales qui provoque la création d'outils peu adaptés. Ils formulent pour cela deux questions primordiales : 1) « Comment les chercheurs en analyse qualitative codent-ils ? » et 2) « Une première passe de codage peut-elle être automatisée et doit-elle l'être ? Si tel est le cas, jusqu'où les techniques de TAL les plus simples peuvent-elles aller ? ». Afin de tenter d'y répondre, ils sollicitent quinze chercheurs expérimentés en analyse qualitative dans le cadre de cinq enquêtes contextuelles puis d'une série de 10 entretiens semi-directifs.

Parmi les résultats qu'ils mettent en évidence, on trouve le constat du fait que les différentes disciplines des chercheurs tendent à changer la manière dont ils définissent les objectifs du codage et semblent influencer leurs choix d'outils. Les auteurs observent par exemple que les chercheurs en Sciences sociales utilisent NVivo ou Atlas.ti tandis que ceux des autres domaines (études des sciences et technologies, anthropologie et interaction homme-machine) utilisent des tableurs ou des outils de traitement de texte ; l'une d'entre eux justifie cela par la lourdeur et le manque de flexibilité des CAQDAS. D'autre part, pour les chercheurs qui les utilisent, les CAQDAS sont considérés comme peu intuitifs et frustrants, et la nécessité de leur utilisation semble être subie. Les auteurs résument ces opinions en indiquant que l'utilisation de ces logiciels est perçue comme « un mal nécessaire ». L'un des participants interrogés estime que les CAQDAS sont « tous mauvais de manière différente ». Au-delà de ce manque d'intuitivité, il leur est également reproché d'utiliser des formats propriétaires et de ne pas permettre le codage simultané par plusieurs codeurs, ce qui nuit à la collaboration.

Perception et critiques des outils par les chercheurs en Sciences Sociales		
i	Les CAQDAS sont considérés comme peu intuitifs et frustrants, et la nécessité de leur utilisation semble être subie (« un mal nécessaire »).	M&T18
ii	Les CAQDAS utilisent des formats propriétaires et ne permettent le codage simultané par plusieurs codeurs, ce qui nuit à la collaboration.	M&T18
iii	La confiance dans les outils est réduite par un manque d'intercompréhension entre les disciplines, dû au fait que les chercheurs en Sciences Sociales ne sont pas suffisamment formés aux techniques du ML, et vice versa.	C18
iv	Le fonctionnement de type « boîte noire » des outils de ML n'inspire pas confiance aux chercheurs	C18
v	Les systèmes à base de règles provoquent des suggestions de codes inutiles parce qu'elles ne tiennent pas compte d'une sémantique dynamique.	R21
Avantages perçus de l'automatisation		
vi	Une augmentation de la qualité du codage est rendue possible par le fait que les utilisateurs doivent définir des règles, ce qui les force à se projeter dans le contenu possible des textes.	R21
Écarts épistémologiques et méthodologiques entre les TAL/ML et les Sciences Sociales		
vii	Etablir la causalité en répondant aux questions <i>pourquoi ?</i> et <i>comment ?</i> est un des buts premiers des Sciences Sociales, tandis que l'objectif du ML est la prédiction du comportement des données sans emphase sur la causalité.	C18
viii	Les chercheurs en Sciences Sociales cherchent à étudier des phénomènes complexes, pour lesquels ils ne disposent parfois pas d'un vocabulaire bien déterminé, ce qui est difficilement compatible avec le ML.	C18
ix	La construction d'un modèle apprenable par le ML n'est pas le but premier dans une approche qui mêle Sciences Sociales et ML : le ML a besoin d'un volume important de données étiquetées, tandis que les chercheurs en Sciences Sociales tendent à arrêter de coder lorsque l'information arrive à saturation.	C18
x	Les éléments récurrents dans les données renforcent les systèmes de ML mais ne constituent pas nécessairement ce qui intéresse le plus les chercheurs en Sciences Sociales	C18
Recommandations méthodologiques pour la création d'outils		
xi	Il est préférable d'impliquer les chercheurs dans un codage qui ait du sens pour eux (i.e., qui ne constitue pas un exercice de production de données d'apprentissage, mais leur soit utile pour traiter leurs problématiques)	C18
xii	Les suggestions produites automatiquement doivent prioriser la précision par rapport au rappel, car le fait de ne pas avoir à rejeter trop de suggestions de codes inutiles est apprécié.	R21

Attentes méthodologiques des chercheurs en Sciences Sociales		
xiii	Un désir d'automatisation est présent chez les chercheurs, mais seulement après avoir manuellement construit la grille d'analyse et avoir codé un extrait des données	M&T18
Attentes des chercheurs en termes d'interactions avec la machine		
xiv	Les chercheurs souhaitent un niveau élevé de transparence de l'outil en termes de recommandations de codage ; le système doit fournir des explications quant à la raison pour laquelle un code est suggéré, sans quoi une certaine confiance ne peut être accordée.	M&T18
xv	Les utilisateurs d'un logiciel qui propose des suggestions de codes souhaitent un niveau de contrôle élevé sur ces dernières, avec un système d'acceptation ou de rejet simple et un aperçu facile et rapide du nombre de suggestions générées.	R21
xvi	Les utilisateurs apprécient avoir plusieurs choix de codes proposés, mais leur nombre ne doit pas être excessif.	R21
xvii	Des explications sur les choix réalisés par le système sont désirées mais globalement ignorées.	R21
xviii	Les scores de confiance des suggestions automatiques sont généralement considérés comme peu utiles par les chercheurs qui ignorent comment ils sont calculés	R21
Facteurs de variations des attentes		
xix	Les différentes disciplines des chercheurs tendent à changer la manière dont ils définissent les objectifs du codage et semblent influencer leurs choix d'outils.	M&T18

C18 : N.-C. Chen et al (2018) – M&T18 : Marathe & Tomaya (2018) – R21 : Rietz (2021)

TABLEAU 9 : Retours d'expériences et recommandations pour l'adéquation entre pratiques du codage en Sciences Sociales et outils

Deux conclusions principales en termes d'attentes sont tirées de cette enquête. La première est qu'un désir d'automatisation est présent, mais seulement après avoir manuellement construit la grille d'analyse et avoir codé un extrait des données, l'autocodage du document entier semblant trop radical et irréversible. La deuxième est que les chercheurs souhaitent un niveau élevé de transparence de l'outil en termes de recommandations de codage ; le système doit fournir des explications quant à la raison pour laquelle un code est suggéré, sans quoi une certaine confiance ne peut être accordée.

8.3.1.2. *Identification des écarts épistémologiques et méthodologiques entre le ML et les pratiques en Sciences Sociales par N.-C. Chen et al. (2018)*

N.-C. Chen et al., (2018) souhaitent mettre en évidence les écarts entre le ML et les pratiques en Sciences Sociales et lancent un appel à réfléchir à des méthodes de ML davantage centrées sur l'humain qui pourraient « avoir le potentiel de transformer la recherche en Sciences Sociales »⁷⁸. Ils soulignent notamment le fait qu'établir la causalité en répondant aux questions *pourquoi?* et *comment?* est un des buts premiers des Sciences Sociales, tandis que l'objectif du ML est la prédiction du comportement des données sans emphase sur la causalité. Ceci constitue une différence fondamentale entre les deux domaines. D'autre part, ils constatent également que les chercheurs en Sciences Sociales cherchent à étudier des phénomènes complexes, pour lesquels ils ne disposent parfois pas d'un vocabulaire bien déterminé, ce qui est difficilement traitable par des systèmes qui reposent sur des caractéristiques décontextualisées (par exemple, des mots clefs) tels que ceux que le ML propose.

Les auteurs se donnent pour objectif d'identifier les principaux obstacles et défis à relever pour construire des applications ML destinées au codage qualitatif. Ils se basent pour cela sur des études formatives conduites sur l'outil *Aedonium* (Drouhard et al., 2017). Celui-ci est conçu pour répondre aux problèmes d'interprétation des données dans un cadre collaboratif en proposant un système qui permet de repérer les ambiguïtés dans les textes. Ils établissent plusieurs constats, dont celui du fait que la confiance dans les outils est réduite par un manque d'intercompréhension entre les disciplines, dû au fait que les chercheurs en Sciences Sociales ne sont pas suffisamment formés aux techniques du ML, et *vice versa*.

Il faut ajouter à cela le fait que leurs pratiques sont sur certains points difficilement conciliables. Notamment, la construction d'un modèle apprenable par le ML n'est pas le but premier dans une approche qui mêle Sciences Sociales et ML ; la production d'un volume important de données étiquetées par des catégories bien définies est essentiel pour le ML, mais cela n'est pas l'objectif que les chercheurs en Sciences Sociales souhaitent atteindre, préférant arrêter de coder lorsqu'ils atteignent un niveau de saturation de l'information. Pour cette raison, il est préférable d'impliquer les chercheurs dans un

⁷⁸ Nous traduisons (p. 9:2) : "have the potential to transform social science research".

codage qui ait du sens pour eux, c'est-à-dire qui ne constitue pas simplement un exercice de production de données d'apprentissage mais permette de mettre en exergue des problématiques qui les intéressent⁷⁹. En outre, les éléments récurrents dans les données renforcent les systèmes de ML mais ne constituent pas nécessairement ce qui intéresse le plus les chercheurs en Sciences Sociales. Ainsi, ce qui est considéré comme du bruit pour le ML peut être une information non négligeable pour les Sciences Sociales. Enfin, de manière plus générale, le fonctionnement de type « boîte noire » des outils de ML n'inspire pas confiance aux chercheurs.

8.3.1.3. Retours d'utilisateurs sur la transparence et les avantages et inconvénients d'un système de suggestions par règles et ML (Rietz, 2021)

Parmi les travaux précédemment évoqués, Rietz (2021) traite la question de la transparence d'un système d'autocodage pour augmenter la confiance des analystes. Il s'inspire principalement des travaux de Marathe & Toyama (2018)⁸⁰ pour fixer six prérequis auxquels le logiciel doit répondre⁸¹ (p. 64-65). L'évaluation formative qu'il réalise pour son système consiste à demander à six docteurs ou doctorants familiers avec la pratique de l'analyse qualitative sur CAQDAS de coder leurs propres données – selon leur propre méthode et avec leurs propres codes – sur *Cody*. Une deuxième évaluation⁸², sommative cette fois, est également réalisée. L'analyse des retours utilisateurs ainsi collectés révèle quelques points intéressants dont nous présentons certains en les

⁷⁹ C'est pour cette raison que les auteurs ont proposé un système de codage collaboratif centré sur le repérage des ambiguïtés, exercice qui permet d'adresser le problème de l'accord inter-juge et de mieux comprendre les données.

⁸⁰ Mais également d'autres travaux, à savoir, respectivement aux prérequis cités à la note suivante : 1) (Crowston et al., 2012), 2) (Ganji et al., 2018), (Paredes et al., 2017) et (Grimmer & Stewart, 2013), 3) (N.-C. Chen et al., 2018) et (Basit, 2003), 5) (Richards, 2002) et (N.-C. Chen et al., 2016), 6) (N.-C. Chen et al., 2018) et (Cheng et al., 2019).

⁸¹ Ces prérequis sont :

- 1) Le système doit permettre de choisir l'unité d'analyse souhaitée (syntagmes, phrases, paragraphes, etc.), pour viser à une meilleure cohérence inter-analystes.
- 2) Il doit inciter les analystes à former des règles par combinaison de mots clefs, afin que cette pratique constitue une occasion d'approfondir leur compréhension des données.
- 3) N'étant pas l'objectif des analystes, l'entraînement du modèle ML ne doit pas causer d'accrocs et sa complexité doit être cachée.
- 4) Les suggestions doivent être basées sur des annotations manuelles pour conserver l'intérêt du codage pour comprendre les données tout en n'utilisant les suggestions que pour réduire la répétitivité.
- 5) Le système doit favoriser la réflexion avec un système de suggestions qui fait office de second codeur qui suit les règles prédéfinies et permet donc de repérer, parce qu'il applique ces règles, celles qui ont été mal définies.
- 6) Le système doit être en mesure de fournir des explications des suggestions qu'il propose afin que le codeur puisse comprendre comment ses actions affectent les suggestions et ce, sans qu'aucune connaissance technique ne soit nécessaire.

⁸² Cf. point 8.2.1.4, p. 66-67.

reformulant parfois pour les généraliser sous forme d'enseignements à retenir plutôt qu'en simples retours d'expérience.

Concernant l'interaction avec les suggestions de codage émises par le logiciel, nous notons que les utilisateurs d'un logiciel qui propose des suggestions de codes souhaitent un niveau de contrôle élevé sur ces dernières (p. 71), notamment avec un système d'acceptation ou de rejet simple, ainsi qu'un aperçu facile et rapide du nombre de suggestions générées. Concernant cette problématique de contrôle et de confiance, les utilisateurs apprécient avoir plusieurs choix de codes proposés (p. 78) car cela montre que le système a bien envisagé toutes les options possibles. Cela force également l'utilisateur à rester vigilant et à ne pas prendre le risque, à cause de la fatigue en fin d'exercice, d'accepter les suggestions de codes sans y réfléchir suffisamment. Cependant, ce nombre de choix possibles ne doit pas être excessif (p. 77). Enfin, concernant la transparence des choix du système, Rietz observe que des explications sur les choix réalisés par ce dernier sont désirés mais globalement ignorés (p. 78), les utilisateurs trouvant suffisant le fait de constater que les suggestions émises sont pertinentes. Sur ce point, les scores de confiance sont généralement considérés comme peu utiles (p. 78), de par le fait que les utilisateurs ne savent pas comment ils sont calculés.

D'autre part, à propos du fait de pouvoir définir manuellement des règles, notons qu'une augmentation de la qualité du codage est rendue possible par le fait que les utilisateurs doivent définir des règles (p. 75). En effet, cela demande à l'analyste d'imaginer ce en quoi peut consister un contenu textuel pertinent, le forçant ainsi à un niveau de réflexion plus poussé sur son codage ; notamment, cette réflexion et l'observation de codes erronés dus à des règles mal définies permettent de faire des choix qui limitent les chevauchements de codes. A ce propos, et plus anecdotiquement, les utilisateurs appelés à définir leurs règles de codage se sentent responsables des suggestions erronées (p. 71). En outre, les utilisateurs constatent que, parce que les règles ne tiennent pas compte d'une sémantique dynamique⁸³, elles provoquent des suggestions

⁸³ « La sémantique dynamique est une perspective sur la sémantique du langage naturel qui met l'accent sur la croissance de l'information dans le temps. Il s'agit d'une approche de la représentation du sens dans laquelle les éléments du texte ou du discours sont considérés comme des instructions pour mettre à jour un contexte existant avec de nouvelles informations, le résultat étant un contexte actualisé. » (Nouwen et al., 2010)
(Nous traduisons : "Dynamic semantics is a perspective on natural language semantics that emphasizes the growth of information in time. It is an approach to meaning representation where pieces of text or discourse are viewed as instructions to update an existing context with new information, the result of which is an updated context.")

inutiles (p.75). En revanche, mieux l'entretien est structuré, plus il est facile de créer des règles qui donnent des résultats précis (p. 75). Globalement, ils estiment qu'utiliser des règles de codage permet effectivement d'accélérer le processus (p. 77).

Concernant la partie ML, les retours des utilisateurs montrent que les suggestions produites doivent prioriser la précision par rapport au rappel (p. 76), car le fait de ne pas avoir à rejeter trop de propositions inutiles a été très apprécié.

8.3.2. Exigences dans le cadre des présents travaux

Nous situons à présent les exigences auxquelles nous souhaitons répondre, au-delà et au service du gain de temps d'analyse qui constitue l'objectif premier de nos travaux. Ces exigences sont en premier lieu déterminées par la demande formulée par l'entreprise Ixiade, mais aussi par de nombreux échanges avec nos collègues au sein de l'entreprise, notre propre expérience du codage, et la littérature que nous venons de présenter. Sur ce dernier point, certaines considérations évoquées dans la recherche ne nous concernent pas directement – en particulier tout ce qui a trait à l'intérêt des outils pour augmenter ou stabiliser la liste de codes ou aider à la construction théorique – mais la confrontation entre notre expérience et les travaux exploratoires précédemment cités tend à nous faire penser que la majeure partie des constatations qui y sont formulées est tout à fait pertinente dans le cas qui nous intéresse. Pour nous y référer, nous utilisons les chiffres romains minuscules entre parenthèses qui renvoient aux retours d'expérience et recommandations listées dans le [TABLEAU 9](#).

Tout d'abord, nous partageons avec les recherches citées l'ambition de ne pas proposer un outil qui soit perçu comme une contrainte (i) et qui tende à aider à la collaboration plutôt que de la contraindre (ii). D'autre part, nous souhaitons, autant que faire se peut, proposer un système qui ne soit pas perçu par les analystes comme une « boîte à outils » de modules de TAL avec lesquels composer. En d'autres termes, nous ne souhaitons pas imposer un nouveau paradigme d'analyse guidée par le TAL, mais au contraire de contraindre nos modèles et algorithmes à épouser les modèles théoriques et méthodologiques utilisés par les analystes. En cela, nous voulons autant que possible nous confronter aux défis que constituent les écarts épistémologiques et méthodologiques entre les deux domaines (vii, viii, ix, x). Pour cela, nous entendons encourager à l'intercompréhension (iii) et à la communication sur les méthodologies des différentes disciplines, leurs tenants et aboutissants, ainsi que leurs potentielles limites (v). Nous

entendons également tenir compte du fait que, si un effort demandé aux analystes dans l'élaboration de l'outil et de ses ressources peut être bénéfique pour toutes les parties (vi), la manière dont ils envisagent le codage n'est pas similaire à celle dont on conçoit la constitution d'un corpus annoté en TAL (xi). Enfin, quant aux attentes en termes d'outil final, nous entendons satisfaire les besoins de transparence (iv, xiv) nécessaires à la confiance (xvii), de lisibilité (xviii), de contrôle (xv, xvi) et de fluidité d'utilisation (xii).

Après avoir situé nos travaux, souligné leurs particularités et leurs difficultés, et précisé nos ambitions, nous abordons à présent les grandes directions structurantes que nous suivons par la suite.

9. Approches ascendantes vs approches descendantes

Dans ce point nous abordons la question de la dialectique entre approches ascendantes et descendantes qui va structurer nos travaux ainsi que les grandes orientations que ces derniers vont suivre. Puis, nous résumons la problématique et présentons les approches testées dans cette thèse.

9.1. Dialectique approches ascendantes et descendantes et grandes directions

Nous présentons ici les différences entre approches ascendantes et descendantes, puis nous nous arrêtons plus particulièrement sur l'opposition entre ML et approches symboliques.

9.1.1. Approches ascendantes et descendantes

Le travail que nous devons réaliser mêle exploitation de corpus et implémentation d'une méthode experte basée sur des considérations théoriques qui n'ont pas été originellement construites sur la base de données textuelles. Pour cette raison, mais aussi dans une logique exploratoire, nous utiliserons deux grandes approches ; la première est l'approche *ascendante*, c'est-à-dire qui part des données pour aller vers un modèle, et la deuxième est *descendante*, c'est-à-dire qui part d'une modélisation experte vers une projection de cette modélisation sur les données. Cette dialectique est similaire à celle des approches *corpus-driven* et *corpus-based* traditionnellement mise en avant en linguistique de corpus (Tognini-Bonelli, 2001). Comme dans le cas de beaucoup d'oppositions méthodologiques en TAL et dans d'autres disciplines, cette dialectique implique une

certaine complémentarité des approches, qui peuvent être et sont utilisées de manière hybride pour traiter la même problématique (Gomez-Perez et al., 2020).

9.1.2. ML et approches symboliques

L'opposition ascendant / descendant transparaît dans le TAL à travers celle traditionnellement formulée entre méthodes symboliques et statistiques. Depuis plusieurs années, les techniques de ML, et en particulier de *Deep Learning*, ont fait des progrès considérables permettant d'atteindre des performances remarquables. Ces performances atteintes par le ML et les promesses qu'il formule ont eu tendance à l'imposer comme l'approche à privilégier. Il permet en effet d'atteindre des scores élevés sur des tâches de classification similaires à celles que nous devons réaliser dans le cadre de nos travaux et peut être efficace sur des corpus multi-domaines. On pourrait alors estimer *a priori* qu'il est la voie à suivre. Cependant, au moins trois raisons nous poussent à ne pas suivre cette orientation comme étant l'unique possible. La première est le faible volume de données d'apprentissage dont nous disposons par rapport au grand nombre de configurations sémantiques et lexicales possibles par Critère. La deuxième est l'aspect « boîte noire » d'une telle solution qui nuit au désir de transparence des outils que souhaitent les analystes. Enfin, la troisième est que le fonctionnement d'outils purement basés sur du ML empêche des ajustements fins des algorithmes sur des paramètres facilement intelligibles par des néophytes et l'ajout de règles correctives pour des phénomènes précis. Cela peut pousser l'analyste vers un certain désengagement, de par le fait que son expertise ne puisse pas être prise en compte et que l'essentiel de sa contribution se résume à coder laborieusement de gros volumes de textes, ce qui ne lui apporte personnellement que peu voire aucun bénéfice. Difficile alors de faciliter la communication interdisciplinaire nécessaire à la création d'un outil sur mesure lorsque cet outil que l'on propose n'est pas capable de justifier les décisions qu'il prend et dans lequel on ne peut pas ajouter à souhait des exigences précises en termes de choix de classification.

Les approches symboliques, quant à elles, comportent des inconvénients bien connus, notamment celui de demander des efforts considérables pour la modélisation linguistique et la création de règles. Elles ont en revanche des avantages comme celui, très intéressant pour les raisons que nous avons évoquées dans le point 8.2.2.1 (pp. 67-70), d'être conformes au principe de compositionnalité dans les textes, ce qui permet la

reconnaissance d'entités dans des combinaisons nouvelles et, ainsi, d'atteindre un certain niveau d'abstraction et de généralisation⁸⁴.

Quoi qu'il en soit, chacune de ces deux approches comporte ses points faibles et points forts et il serait très caricatural, et même faux, de considérer que le ML a complètement supplanté les approches symboliques ; au contraire, les apports mutuels que peuvent entretenir ces deux paradigmes sont bien identifiés, et leur conciliation est un sujet très actuel et prometteur (Garnelo & Shanahan, 2019; Gomez-Perez et al., 2020).

En conclusion, il ne s'agit donc pas pour nous de choisir une orientation, ascendante ou descendante sur la base d'*a priori*, mais bien d'aller dans la direction que choisissent de nombreux autres chercheurs, à savoir d'explorer les possibles apports de ces deux approches, avant d'en choisir l'une ou l'autre ou bien de se diriger vers un modèle hybride. Dans le point suivant, après avoir résumé les problématiques soulignées dans ce chapitre, nous présentons le plan des approches testées.

9.2. Résumé de la problématique et approches testées

Au vu des différents éléments que nous avons mis en évidence dans ce chapitre et, plus généralement, dans la première partie de cette thèse, nous pouvons à présent formuler une problématique et présenter les approches que nous allons tester.

Sur le premier point, nous pouvons dire que nos travaux consisteront donc à explorer les possibilités offertes par le TAL pour tenter de répondre aux questions suivantes :

⁸⁴ Comme l'expliquent très bien Garnelo & Shanahan (2019, p.18) : « En IA symbolique, les représentations [...] obéissent à un principe de compositionnalité dans la mesure où la dénotation d'une représentation est fonction de la dénotation de ses parties et de la manière dont ces parties sont combinées. [...] Pour un agent confronté à un monde qui présente lui-même une structure combinatoire, un système compositionnel de représentation a le potentiel de conférer la capacité de former des abstractions et de généraliser bien au-delà de sa propre expérience, car les représentations d'objets et de relations familiers peuvent entrer dans de nouvelles combinaisons. Une enfant n'a peut-être jamais vu d'ours en peluche dans un seau, mais si elle a déjà rencontré des seaux, des ours en peluche et des objets dans des objets, elle peut imaginer un ours en peluche dans un seau. Cela peut s'avérer utile si jamais elle a l'occasion, par exemple, de transporter une collection d'ours en peluche de la pataugeoire à la balançoire. »

(Nous traduisons : "In symbolic AI, representations [...] conform to a principle of compositionality to the extent that the denotation of a representation is a function of the denotation of its parts and the way those parts are combined. [...] For an agent confronted by a world that itself exhibits combinatorial structure, a compositional system of representation has the potential to confer the ability to form abstractions and to generalise far beyond its own experience, because representations of familiar objects and relations can enter into novel combinations. A child may never have seen a teddy bear in a bucket before, but if she has encountered buckets and teddy bears and things in things then she can imagine a teddy bear in a bucket. This could come in handy if ever she has occasion, say, to transport a collection of teddy bears from the paddling pool to the swing.")

- 1) Comment mobiliser le TAL pour créer un système d'autocodage en analyse qualitative selon les spécificités suivantes ? :
 - a. pour une analyse dans le domaine spécifique de l'évaluation de l'acceptabilité et qui, si elle mobilise des concepts et méthodes de sociologie, est à visée diagnostic et non de construction théorique.
 - b. qui utilise une grille d'analyse prédéfinie constituée de codes qui mobilisent une complexité sémantique certaine et doit pouvoir être appliquée à des corpus multi-domaines.
 - c. qui dispose de petits corpus étiquetés.
- 2) Comment faire en sorte que le système créé permette de répondre à des exigences :
 - a. d'efficacité en termes de diminution de temps de codage ?
 - b. de transparence de l'algorithme et de possibilité de contrôle ?
 - c. de non-imposition, par le TAL, d'un nouveau paradigme d'analyse ?

Cette exploration des possibilités devra donc tenir compte des difficultés d'adéquation épistémologique et méthodologique entre le TAL et la pratique du codage en analyse qualitative, ainsi que des exigences opérationnelles et UX formulées par Ixiade. Sur ces points, nous pourrions nous appuyer sur les quelques travaux exploratoires que nous avons cités, mais aussi sur nos nombreuses interactions avec les analystes d'Ixiade, ainsi que sur nos propres connaissances et expériences pratiques dans les deux domaines.

Dans ce chapitre, nous avons mis en avant le positionnement spécifique de la tâche demandée par Ixiade et objet de cette thèse ; il s'agit d'une demande d'automatisation d'analyse qualitative de nature particulière et qui utilise des représentations sémantiques complexes. L'outil construit pour cela devra satisfaire de nombreuses exigences de performance et d'adéquation aux pratiques. Afin de répondre à cette demande, nous explorons les possibilités offertes par deux grandes directions : une approche descendante et une ascendante, que nous présentons en premier lieu dans la partie suivante de cette thèse.

PARTIE II – APPROCHE

ASCENDANTE

CHAPITRE IV : CORPUS D'APPRENTISSAGE

Cette seconde partie de thèse s'attèle à décrire les investigations menées quant aux possibilités d'approches dites *ascendantes* à résoudre le problème que nous est posé. Ces approches s'appuyant par définition sur des données, le présent chapitre décrit la manière dont nous avons construit un corpus d'apprentissage à partir de contenus d'études passées. Nous présenterons dans les deux suivants chapitres des expérimentations qui permettent une évaluation du potentiel des méthodes statistiques puis celle de la possibilité de recours à une ressource sémantique pour enrichir le corpus et aider à son analyse.

Nous décrivons donc, pour l'heure, le processus de constitution du corpus ainsi que ses caractéristiques. Les deux étapes qui ont été nécessaires à sa création, à savoir l'agrégation des données d'études passées et leur filtrage par annotation experte avec accord inter-juge, structurent ce chapitre.

1. Agrégation et structuration des données

Le processus de constitution de corpus que nous mettons en œuvre consiste à extraire, agréger et formater des verbatim codés lors d'études précédemment menées par l'entreprise Ixiade et qui ont été archivées. L'objectif est d'obtenir des données d'entraînement du type de celles utilisées pour entraîner des classifieurs, i.e. constituées d'unités textuelles étiquetées avec des classes (ici, les Niveaux et Critères CAUTIC®). Nous choisissons en effet, dans cette partie, de considérer le problème comme une tâche de classification d'unités textuelles préalablement extraites, tout en conservant, lorsque possible, la trace des textes entiers qui contiennent toutes les données des entretiens/*focus groups* y compris celles considérées comme non pertinentes pour l'analyse. Cette décision est motivée par le fait que nous cherchons principalement à extraire des données les caractéristiques propres au contenu attendu pour chaque Critère, réservant à un second temps la distinction entre données pertinentes et non pertinentes dans les transcriptions intégrales. Elle est également induite par des considérations pratiques imposées par le contenu des archives : les données non pertinentes ne sont pas systématiquement conservées. Dans la suite, nous décrivons les

données disponibles, expliquons le processus d'agrégation et de formatage, et exposons la volumétrie du corpus obtenu.

1.1. Données disponibles : structure et sélection

Nous commençons par explorer les archives de l'entreprise pour sélectionner les études passées pertinentes dans le cadre de nos travaux. Nous retenons celles pour lesquelles la méthode CAUTIC® a été déployée et dont les données textuelles sont au moins en partie en français. Précisons immédiatement qu'au moment où nous avons constitué ce corpus la plateforme de collecte *yoomaneo* n'était pas encore lancée et qu'il ne contient donc aucune donnée saisie en ligne par des participants mais uniquement des transcriptions d'entretiens et de *focus groups*. Par étude contenue dans les archives, trois types de fichiers, représentés dans la FIGURE 7, nous intéressent :

- Les transcriptions ou prises de notes des entretiens et des *focus groups* sont des fichiers *.docx* qui contiennent tout le matériau collecté. Si les entretiens donnent lieu à des transcriptions intégrales de l'échange, les *focus groups* font parfois seulement l'objet de prises de notes dans lesquelles coexistent des paroles transcrites, des résumés synthétiques (par exemple, « *Intéressant sur les hottes de désenfumage.* »), ou encore des pseudo-titres (« *Jeu : si c'était un jeu, ce serait un jeu un peu long...* »). Toutes les données – entretiens et *focus groups* – sont susceptibles de contenir des didascalies (« *La bonne nouvelle, c'est qu'il y a quelque chose pour l'accrocher [il fait le geste d'accrocher à un crochet à son étagère au-dessus].* ») ou des indications de désambiguïsation occasionnelles (« *ça m'intéresserait plus (+)* », « *Je pense que ça [le praliné] va peut-être rester souple* »). Cependant, ces indications d'aide à la compréhension sont assez rares.
- La mise à plat est contenue dans un fichier Excel ou Access et consiste en l'agrégation de tous les verbatim des différents entretiens et *focus groups* de l'étude qui sont classifiés dans les Niveaux et/ou Critères CAUTIC®.
- Le livrable est un fichier *.ppt* ou *.pdf* qui constitue le rapport restitué au client. Il contient souvent des verbatim sélectionnés parmi ceux de la mise à plat pour illustrer les conclusions de l'analyse. Cette sélection se fait sur la base soit de la représentativité du verbatim par rapport au Critère traité,

soit pour sa pertinence pour un point précis du projet, soit parce qu'il est considéré comme particulièrement percutant en termes communicationnels⁸⁵. Ces verbatim, seront appelés *verbatim illustratifs* en opposition aux *verbatim analyse* de la mise à plat.

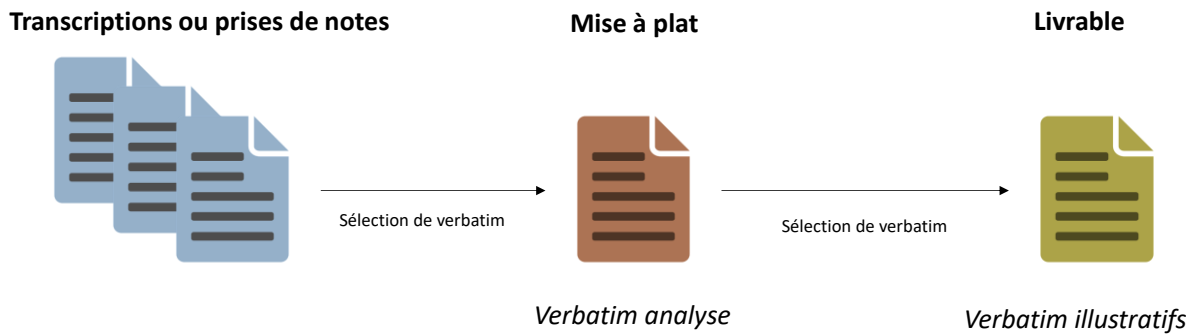


FIGURE 7 : Fichiers des données d'études dans les archives d'Ixiade

Précisons qu'un verbatim illustratif peut être l'intégralité du verbatim analyse source, ou bien en être seulement un extrait, par souci de concision dans le livrable. Le verbatim analyse source est lui-même extrait de la transcription par un processus de segmentation arbitraire ; paragraphe entier, quelques phrases, une seule phrase ou encore seulement un syntagme, voire un seul mot. On a donc affaire à un processus en entonnoir qui permet de « rogner » progressivement une partie de la transcription en une sous-partie dans la mise à plat, puis en une sous-sous-partie dans le livrable, comme illustré dans l'exemple ci-dessous :

- 1) « [Euh... Négatif... non, non, faut... Après, il faut... C'est... Donc il y avait ce concept de... d'achat des modules qui est... qu'il faut voir, la... la pertinence de... de ce mode de fonctionnement. [[Il ne faut pas que ce soit trop... trop restrictif et trop bridé. Il faut quand même que ce soit une solution assez ouverte]verbatim illustratif pour que ce soit facilement utilisable par l'industriel. Euh... Après, sinon, en point négatif... Non, je ne vois pas de gros points négatifs sur ce type d'appareil.]verbatim analyse]paragraphe transcription »

⁸⁵ Ces derniers sont souvent appelés par les analystes, de manière tout à fait informelle, des *punchlines*.

Aux critères de sélection relatifs au déploiement de CAUTIC® et à la langue, nous ajoutons le fait que ne sont retenues que les études pour lesquelles ont été conservés soit les verbatim analyse, soit les verbatim illustratifs, soit les deux.

Pour mettre en œuvre cette sélection, parmi une liste de près de 500 projets archivés, nous demandons à nos collègues de pointer ceux qui, selon leurs souvenirs, ont nécessité le déploiement de la méthode CAUTIC®. Nous explorons les dossiers ainsi désignés et éliminons les études qui ne répondent pas au critère de langue ou dont les données textuelles n'ont pas (ou peu) été conservées. Nous obtenons ainsi 29 études potentiellement exploitables.

L'entreprise Ixiade ayant réalisé de nombreuses études sur des sujets relatifs à l'électricité, nous distinguons celles qui relèvent de ce domaine de celles qui relèvent d'un autre. Seules neuf études sont dans le deuxième cas. Partant de l'idée qu'une sur-représentation du domaine électrique dans le corpus est en adéquation avec l'activité d'Ixiade, mais que nos données doivent rester multi-domaines, nous choisissons de ne conserver que huit études dans le domaine électrique, afin d'obtenir un ensemble équilibré. A l'issue de cette sélection, nous conservons donc 17 études. Elles ont été réalisées entre 2005 et 2018 par au moins neuf analystes⁸⁶ différents et comprennent comme matériau brut un total de 67 entretiens transcrits⁸⁷ (soit 406 000 mots) et 16 transcriptions de *focus groups* (soit 108 000 mots). Les verbatim sont, quant à eux, au nombre de 3 825 pour les verbatim analyse et 1 608 pour les verbatim illustratifs.

1.2. Agrégation et formatage

A partir de ces documents, nous agrégeons donc manuellement⁸⁸ les verbatim disponibles, puis les formatons sous formats tabulaire et XML. Il arrive que certaines analyses n'aient pas nécessité une granularité fine du codage CAUTIC®, et que les verbatim se soient vu assigner seulement un Niveau et non un Critère précis ; nous décidons de les conserver. Pour chaque verbatim, nous consignons : un code

⁸⁶ Le nom de l'analyste n'était parfois pas identifiable.

⁸⁷ Sur 114 réalisés ; les transcriptions manquantes ont soit été perdues, soit elles n'ont jamais été réalisées, les analystes travaillant parfois directement avec l'audio lorsque les délais du projet ne permettent pas la transcription. Autrement dit, pour un nombre notable de verbatim de notre corpus récupérés dans les mises à plat, nous ne disposons pas de la transcription intégrale dont ils ont été extraits.

⁸⁸ La diversité des formats de fichiers et des pratiques des différents analystes à travers le temps (pas de *template type*) a rendu difficile, voire régulièrement impossible, l'automatisation complète ou partielle de l'extraction de la majorité des données, nous forçant ainsi à avoir recours à des copier-coller et des saisies d'informations manuelles pour chaque verbatim.

d'identification de l'étude, un code d'identification du ou des participants⁸⁹, le type de collecte (entretien ou *focus group*) et le type de verbatim (analyse ou illustratif). S'il s'agit d'un verbatim illustratif qui est identique à un segment d'un verbatim analyse, nous renseignons également la référence du verbatim analyse source. Enfin, nous renseignons le Niveau et Critère CAUTIC® auquel il a été attribué⁹⁰. Trois verbatim du corpus XML obtenu sont présentés dans l'ENCADRÉ 6 (les balises *<part>* indiquent un tour de parole d'un participant, les balises *<iwer>* celui d'un interviewer).

```

<corpus>
  <verbatim id="1" etude="EC1_Elec" elec="oui" niveau="1" critere="1.2" origine="IW" type="analyse">
    <part code="EC1_Elec_IW02_0">Je n'ai pas bien compris si c'était destiné à rester ad-vitam dans le
    bâtiment ou si c'est juste à la livraison sur une période donnée. Parce que si c'est un bâtiment
    neuf, sauf à dire qu'il y a des malfaçons... j'ai du mal à comprendre</part>
  </verbatim>
  [...]
  <verbatim id="346" etude="EC1_Elec" elec="oui" niveau="2" critere="2.4" origine="IW"
  type="illustratif" idDisc="130">
    <part code="EC1_Elec_IW02_0">Savoir que sur 3 mètres carrés, il y a un mur qui a pris l'eau et que
    sa résistance n'est pas bonne, si c'est un bâtiment qui a 3000 mètres carrés de façades, je trouve
    que c'est un peu lourd pour détecter ça. </part>
  </verbatim>
  [...]
  <verbatim id="1866" etude="EC5_Elec" elec="oui" niveau="1" critere="1.3" origine="IW" type="discours"
  fmc="condition">
    <iwer>le variateur, il remonte quel type d'information ?</iwer>
    <part code="EC5_Elec_IW04_1">Ah ben on peut récupérer ce qu'on veut. On peut remonter les
    surintensités, l'intensité...</part>
    <part code="EC5_Elec_IW04_2">Le courant absorbé... tous les paramètres, en fait. Tout, tout, tout,
    tout. On peut tout remonter.</part>
    <part code="EC5_Elec_IW04_1">La puissance consommée.</part>
    <part code="EC5_Elec_IW04_2">Courant consommé, puissance consommée. Après, tout le paramétrage,
    euh... On peut tout remonter.</part>
    <iwer>Est-ce qu'il peut par exemple remonter la température de l'environnement moteur ?</iwer>
    <part code="EC5_Elec_IW04_2">Tout à fait, oui.</part>
    <iwer>Avec un capteur ?</iwer>
    <part code="EC5_Elec_IW04_2">Tout à fait. Un capteur qui va sur le bornier de... du variateur, qui
    est défini comme fonction de... température moteur et on le relaie sur le réseau.</part>
  </verbatim>
  [...]
</corpus>

```

ENCADRÉ 6 : Extrait du corpus de verbatim d'études passées

1.3. Volumétrie du corpus non filtré

Ce premier corpus est constitué de 5 433 verbatim qui correspondent à environ 205 000 mots. Le verbatim le plus court est constitué d'un seul mot, le plus long de 876, ce qui représente donc une variabilité élevée. Cette variabilité est cependant relative, car 68% des verbatim comportent moins de 50 mots et seuls 7,4% en ont plus de 100. Nous

⁸⁹ Un verbatim peut contenir des paroles de plusieurs locuteurs dans le cas des *focus groups*.

⁹⁰ Nous renseignons également sa classe en termes de freins, motivations et conditions, pour les besoins des travaux de notre collègue dont la thèse porte sur cette problématique.

présentons brièvement ici sa composition et renvoyons le lecteur aux [TABLEAUX EN ANNEXE 1 à 3](#) (pp. 271-272) pour consulter plus en détails les chiffres mentionnés.

1.3.1. Répartition par études et domaines

Comme détaillé dans le [TABLEAU EN ANNEXE 1](#) (p. 271), 59,58% des verbatim sont issus du domaine de l'électricité, et 40,42% portent sur d'autres domaines parmi lesquels la médecine est fortement représentée (29% des verbatim de la catégorie *Non Electrique*). On constate d'autre part que le nombre de verbatim est très inégal d'une étude à l'autre ; une seule d'entre elle représente 20,06% des verbatim, tandis qu'une autre n'en représente que 0,79%.

1.3.2. Répartition par Niveaux et Critères

La répartition des verbatim par Niveaux et Critères attribués (voir [TABLEAU EN ANNEXE 2](#), p. 271) est plutôt inégale ; le nombre de verbatim par Critère varie de 81 à 508, et le Niveau 2 comporte deux fois plus de verbatim que le Niveau 4. Cette disparité n'est pas due à des données manquantes mais reflète les contenus des données collectées ; en d'autres termes, les participants s'expriment davantage sur des sujets relatifs au Niveau 2 (i.e. à leurs pratiques) qu'aux autres Niveaux.

1.3.3. Répartition par types de participants

Comme indiqué dans le [TABLEAU EN ANNEXE 3](#) (p. 272), la majorité des participants sont des hommes (62% contre 25% de femmes et 13 % de participants pour lesquels cette information n'était pas présente dans les données). Cette différence de représentation entre hommes et femmes varie selon le titre auquel ils ou elles sont interrogé(e)s ; seuls 15% des personnes participant à l'étude en leur qualité de professionnels sont des femmes, tandis qu'elles représentent 52% des participants qui s'expriment en tant que particuliers⁹¹.

En somme, le désir d'équilibre de notre corpus de verbatim n'est pas pleinement satisfait, mais la quantité de données disponibles étant limitée, nous sommes contrainte de chercher à en exploiter le maximum au détriment de cet équilibre. De plus, les disparités présentes dans le corpus, notamment en termes de domaines d'application et

⁹¹ Au total, 73% des participants ont été impliqués dans des études dites *B2B* (*business to business*), c'est-à-dire pour des projets donc la cible de l'innovation testée est professionnelle, contre 27% pour des études dites *B2C* (*business to consumer*), dans lesquelles les personnes sont interrogées en tant que consommateurs particuliers. Sur les 17 études comprises dans le corpus, 12 sont *B2B*, deux sont *B2C*, et trois incluent à la fois des participant professionnels et des participants particuliers. Cette répartition est assez représentative de l'activité de l'entreprise Ixiade.

de répartition par Critères, ne sont pas liées à des données absentes dans les archives mais bien représentatives de l'activité et des pratiques d'études de l'entreprise. A défaut, donc, d'être équilibré, notre corpus peut prétendre à une certaine représentativité. Nous présentons dans le point suivant le processus de filtrage expert que nous avons utilisé sur ce dernier.

2. Filtrage expert

A la lecture des verbatim du corpus, nous nous interrogeons quant à leur réelle représentativité pour chaque Critère. En effet, pour un nombre notable d'entre eux, nous trouvons que le lien entre le contenu et le Critère auquel il est associé est difficilement saisissable. Plusieurs raisons peuvent expliquer cela. Les plus évidentes sont l'erreur humaine et les potentielles disparités de maîtrise de la méthode par les analystes qui se sont succédé au fil des années pendant lesquelles les études ont été menées. Au-delà de potentielles erreurs, la difficulté de la tâche et sa dimension hautement interprétative peuvent amener à considérer le codage d'un verbatim comme sujet à discussion sans qu'il ne constitue à proprement parler une erreur. Une autre raison réside dans le fait que les données issues des mises à plat que nous utilisons n'avaient pas vocation à être comprises et réutilisées par une autre personne que l'analyste ; les fichiers utilisés dans l'analyse étaient ses brouillons, et sa manière de segmenter et organiser l'information n'étaient pas destinés à en faciliter la transmission. Ainsi, par exemple, il nous semble que les analystes qui utilisaient dans leur mise à plat des verbatim qui étaient sous forme de longs paragraphes ne cherchaient en fait qu'à en classifier un petit extrait et ne laissaient un long contexte qu'afin de mieux le comprendre lors de l'étape d'interprétation et de synthèse. Mais, surtout, la raison la plus problématique relève de la nécessité d'une dimension pragmatique pour interpréter les verbatim. En effet, en amont de notre lecture des verbatim, nous ne nous étions familiarisée que très superficiellement avec les tenants et aboutissants des projets dans le cadre desquels ils avaient été analysés. Nous ne pouvons donc pas interpréter le texte en déployant une dimension pragmatique permise par la connaissance du sujet traité ; il arrive en effet que, sans la connaissance de ce sujet – le Concept testé mais aussi certaines problématiques particulières de l'étude – il soit difficile de comprendre les raisons sous-jacentes qui poussent à considérer un verbatim comme pertinent ou comme étant en adéquation avec un des Critères. Or, la machine ne disposera pas de cette connaissance, et viser un système qui prenne en compte des

paramètres pragmatiques aussi complexes (en plus de la complexité sémantique) semble peu réaliste. En l'état, donc, le corpus contient trop de données qui vont constituer du bruit pour l'apprentissage. Il nous a alors paru nécessaire, pour construire un corpus d'apprentissage en adéquation avec nos exigences et ambitions, d'éliminer les verbatim dont la classification :

- est clairement erronée (due à une erreur humaine)
- n'est pas conditionnée par l'interprétation subjective d'une seule personne, si compétente soit-elle
- présente la nécessité absolue d'une interprétation pragmatique permise par la connaissance du projet

Nous décidons donc de solliciter les membres de l'entreprise pour qu'ils recodent les verbatim du corpus, présentés isolément, afin d'éliminer ceux qui ne répondent pas à ces critères. En cela, nous allons à l'encontre de la recommandation de [N.-C. Chen et al. \(2018\)](#) (cf. CHAPITRE III, 8.3.1.2, §3, p. 77), précédemment évoquée, de ne pas chercher à impliquer les analystes dans une tâche de codage dont le seul objectif est une production de données pour un apprentissage de la machine (et qui ne présente donc pour eux un aucun intérêt). Cependant, nous estimons que le corpus est trop bruité, souhaitons privilégier la qualité à la quantité – qui est, quoi qu'il en soit, relativement faible – et estimons donc que cette étape est nécessaire et qu'une alternative est difficilement envisageable⁹².

2.1. Méthodologie de la classification inter-juge

Nous commençons par supprimer les doublons ou quasi doublons dus aux redondances entre verbatim analyse et verbatim illustratifs, en ne conservant que les premiers. Nous en éliminons ainsi 1 066 et en conservons 4 367, que nous ordonnons de façon aléatoire, et soumettons par vagues aux membres de l'équipe, en leur demandant d'attribuer à chacun un Niveau et un Critère, selon les modalités suivantes :

- Un verbatim peut recevoir la classe *Inclassifiable* pour le Niveau et sera donc automatiquement classifié en tant qu'*Inclassifiable* également pour le Critère

⁹² Bien qu'étant nous-même formée à l'analyse CAUTIC® et l'ayant utilisée à l'occasion de plusieurs projets, nous nous refusons à effectuer personnellement un filtrage arbitraire des données qui ne permettrait pas de diminuer le biais de subjectivité induit par la dimension interprétative de l'exercice et nuirait à la dimension scientifique de la démarche.

- Un verbatim peut être attribué à un Niveau précis mais peut recevoir la classe *Inclassifiable* pour le Critère

Huit collaborateurs de l'entreprise (nous-même incluse) sont compétents pour classer du contenu selon la méthode CAUTIC®, et nous ne participons pas à l'évaluation en Critères pour conserver notre neutralité dans cette démarche⁹³. Nous désignerons ces collaborateurs analystes par des codes qui reflètent leur expérience avec la méthode : nous utiliserons les codes *S1, S2, S3, S4* et *S5* pour les seniors et *J1* et *J2* pour les juniors. Nous entendons par *seniors* les personnes qui sont expérimentées et autonomes dans la pratique de l'analyse et qui peuvent soit analyser l'intégralité des données d'un projet ou bien superviser la réalisation de cette analyse par d'autres personnes. Les *juniors* sont des personnes qui sont formées à la méthode, l'ont déjà utilisée dans le cadre de projets, mais pas en autonomie complète. En outre, les chiffres dans les codes sont un classement par ordre chronologique d'ancienneté (le 1 représente l'ancienneté la plus importante).

Lors d'une première vague d'évaluation servant à ajuster les modalités méthodologiques et pratiques de notre évaluation, nous demandons à six collaborateurs de classer les 100 mêmes verbatim. Suite à cela, nous constituons deux groupes d'évaluateurs auxquels nous soumettons parallèlement deux sets de verbatim différents par vagues de 200. Nous considérons comme complètement évalués les verbatim ayant été classifiés par au moins trois personnes. Cette tâche est particulièrement coûteuse en ressources et présente une forte incompatibilité avec la charge d'activité de l'entreprise qui contraint régulièrement à espacer les vagues d'évaluation ou à accumuler du retard sur celles précédemment initiées. Aussi, nous ne pouvons soumettre l'intégralité du corpus à l'évaluation, et n'obtenons finalement que 1 658 verbatim complètement évalués.

⁹³ Nous participons cependant ponctuellement à l'évaluation en freins, motivations et conditions, lorsqu'un ou une de nos collègues ne dispose pas d'assez de temps pour participer à une vague d'évaluation. Réciproquement, notre collègue taliste est amenée à coder des verbatim en Niveaux et Critères pour les mêmes raisons. Cela n'est pas idéal, mais nous sommes toutes deux formées à la méthode et la mettons en pratique dans le cadre de projets clients. De plus, la tâche étant particulièrement chronophage et ne pouvant constituer une priorité pour nos collègues bien occupés par leurs autres activités, cela a constitué une façon d'accélérer le processus de filtrage qui a dû être étalé sur de longs mois.

2.2. Critères de filtrage et volumétrie finale du corpus filtré obtenu

Afin d'obtenir un corpus contenant uniquement les verbatim sur lesquels le moins d'ambiguïté possible est présente, nous réalisons un filtrage qui consiste à ne conserver que ceux auxquels au moins 50% des analystes ont attribué le même code (en cas d'évaluation par un nombre pair d'évaluateurs donnant un résultat également réparti en 50% sur une classe et 50% sur une autre, nous considérons le verbatim comme étant à éliminer). Par souci de concision, nous indiquerons comme *validés* les verbatim qui ont obtenu un accord inter-juge répondant à ce critère. Lorsqu'un verbatim est validé pour le Niveau mais pas pour le Critère, nous ne conservons que l'indication du Niveau. Dans les cas où nous trouvons un accord de 50-50 sur les Niveaux, mais que le verbatim est validé sur les Critères d'un de ces Niveaux, nous conservons le Niveau associé au Critère retenu ; autrement dit, si, par exemple, sur quatre évaluateurs, deux assignent le Niveau 3 et deux le Niveau 4, mais que les deux premiers choisissent tous deux le Critère 3.1 tandis que les seconds attribuent, pour l'un, le Critère 4.1, et pour l'autre le Critère 4.2, alors nous considérons comme validé le Niveau 3.

Filtrage en Niveaux		
	n	%
Conserver les verbatim		
- Conserver le verbatim avec le même Niveau que celui indiqué dans les archives (1)	616	37,15%
- Conserver le verbatim mais lui attribuer un Niveau différent de celui contenu dans les archives (2)	469	28,29%
Total conservé	1085	65,44%
Supprimer les verbatim		
- Pas d'accord interjuge (3)	252	15,20%
- Accord interjuge sur catégorie <i>Inclassifiable</i> (4)	321	19,36%
Total supprimé	573	34,56%
Filtrage en Critères		
	n	%
Conserver les verbatim		
- Conserver le verbatim avec le même Critère que celui indiqué dans les archives (1)	277	16,71%
- Conserver le verbatim mais lui attribuer un Critère différent de celui contenu dans les archives (2)	460	27,74%
Total conservé	737	44,45%
Supprimer les verbatim		
- Pas d'accord interjuge (3)	572	34,50%
- Accord interjuge sur catégorie <i>Inclassifiable</i> (4)	349	21,05%
Total supprimé	921	55,55%

(1) Accord interjuge $\geq 50\%$ sur le même Niveau/Critère que celui qui avait été consigné dans les archives

(2) Accord interjuge ≥ 50 sur un Niveau/Critère différent de celui qui avait été consigné dans les archives

(3) Pas d'accord interjuge $\geq 50\%$ sur un même Niveau ou Critère

(4) Accord interjuge $\geq 50\%$ sur le fait que le verbatim soit *inclassifiable*

TABLEAU 10 : Résultats du filtrage du corpus par accord inter-juge

Le résultat de ce filtrage donne lieu à quatre cas de figure possibles qui sont présentés dans le TABLEAU 10 avec les résultats chiffrés de cette évaluation. Les cas qui donnent lieu à une validation du verbatim sont de deux ordres : soit l'accord inter-juge coïncide avec la classe initialement attribuée dans les archives, et nous conservons donc intacte la classification (validation par conservation), soit un accord inter-juge satisfaisant est atteint, mais sur une classe différente de celle trouvée dans les archives. Dans le second cas, nous choisissons de réassigner au verbatim la classe sur laquelle les évaluateurs se sont trouvés en accord⁹⁴ (validation par réassignation). La validation par conservation concerne ainsi seulement 37,15% des verbatim pour les Niveaux, et 16,71% pour les Critères. La validation par réassignation est effectuée, pour les Niveaux, dans 28,26% des cas, et 27,74% pour les Critères. Au total, nous ne pouvons donc garder dans notre corpus que 1 085 verbatim (soit 65,44% de ceux évalués), dont 348 n'ont qu'un Niveau d'attribué et pas de Critère.

En termes de volume, le corpus est donc considérablement réduit. En revanche, le filtrage n'a pas altéré de manière significative la répartition des verbatim par Niveau/Critère, par domaine d'activité, par type de verbatim analyse/illustratif, ou par type de répondants, ce qui semble indiquer qu'aucun de ces paramètres n'influe sur le degré d'ambiguïté des verbatim hors contexte. Cette répartition est présentée dans les TABLEAUX 11, 12 et 13. On observe cependant un petit écart au niveau de la répartition entre verbatim analyse et illustratifs ; les seconds représentent 30% du corpus issu des archives et 38% du corpus filtré. Deux tendances distinguent ces deux types de verbatim : la représentativité du verbatim par rapport au Critère est potentiellement supérieure dans le cas des verbatim illustratifs, et ceux de l'analyse sont sensiblement plus longs (donc, comportent davantage de contexte). C'est certainement la première qui provoque un accord inter-juge plus important, mais l'influence de la seconde n'est peut-être pas négligeable ; les verbatim de plus de 200 mots sont inexistantes parmi les verbatim conservés avec le même Critère que celui consigné dans les archives et seuls 20% d'entre eux sont réassignés, laissant les 80% restants considérés comme inclassifiables ou ne parvenant pas à susciter un accord inter-juge satisfaisant (cf. [FIGURE EN ANNEXE 7, p. 272](#)).

⁹⁴ Une alternative aurait été de considérer la classification initiale comme le travail d'un évaluateur supplémentaire. Cependant, nous considérons que ce travail, parce qu'il a été effectué en connaissance des contextes et par rapport aux exigences des projets traités, ne peut pas être comparé à la tâche que nous demandons ici aux analystes, et ne souhaitons donc pas les traiter de manière identique.

Cela indique que des segments longs, hors contexte entier du texte, n'aident pas à la classification mais, au contraire, la compliquent, notamment dû au fait que ces segments sont susceptibles de contenir des informations qui peuvent être attendues par plusieurs Critères.

Etude	Analyse	Illustratif	Total	Pourcentage
EC1_Elec	74	0	74	6,82%
EC2_Elec	72	2	74	6,82%
EC3_Elec	37	10	47	4,33%
EC4_Elec	55	4	59	5,44%
EC5_Elec	244	1	245	22,58%
EC6_Elec	64	1	65	5,99%
EC7_Elec	13	6	19	1,75%
EC8_Elec	11	28	39	3,59%
Total électrique	570	52	622	57,33%
EC1_Non Elec (électroménager)	147	3	150	13,82%
EC2_Non Elec (médecine)	45	5	50	4,61%
EC3_Non Elec (gérontologie)	53	0	53	4,88%
EC4_Non Elec (pâtisserie)	85	16	101	9,31%
ENC1_Non Elec (médecine)	15	2	17	1,57%
ENC2_Non Elec (médecine)	15	0	15	1,38%
ENC3_Non Elec (automatisme)	0	7	7	0,65%
ENC4_Non Elec (médecine)	13	44	57	5,25%
ENC5_Non Elec (pâtisserie)	0	13	13	1,20%
Total Non Electrique	373	90	463	42,67%

TABLEAU 11 : Répartition des verbatim par études et domaines, et par type analyse/illustratif dans le corpus final filtré

		Verbatim analyse	Verbatim illustratifs	Total			Verbatim analyse	Verbatim illustratifs	Total
NIVEAU 1	1.1	62	8	70	NIVEAU 3	3.1	71	10	81
	1.2	32	5	37		3.2	4	0	4
	1.3	51	5	56		3.3	16	2	18
	1.4	51	3	54		3.4	25	19	44
	1.5	6	2	8		3.5	7	2	9
	Non précisé	103	17	120		Non précisé	22	2	24
Total niveau		305	40	345	Total niveau		145	35	180
NIVEAU 2	2.1	30	5	35	NIVEAU 4	4.1	23	3	26
	2.2	25	2	27		4.2	11	1	12
	2.3	65	11	76		4.3	4	0	4
	2.4	72	16	88		4.4	13	2	15
	2.5	27	4	31		4.5	37	5	42
	Non précisé	150	15	165		Non précisé	36	3	39
	Total niveau		369	53		422	Total niveau		124

TABLEAU 12 : Répartition des verbatim par critère et niveau dans le corpus final filtré

	Femmes	Hommes	Inconnu
Répartition	21,03%	43,65%	35,32%
Professionnels	17,39%	65,22%	17,39%
Particuliers	59,18%	40,82%	0,00%
Moyenne d'âge	45,47	43,29	-

TABLEAU 13 : Moyenne d'âge et répartition des participants en genres, statuts professionnels/particuliers dans le corpus final filtré

2.3. Mesure de l'accord inter-juge

Nous avons abordé jusqu'ici la question de l'accord inter-juge comme une mesure utile en tant que seuil de filtrage sans nous intéresser à ce que ces scores signifient et peuvent révéler. Dans ce point, nous détaillons ces scores puis tentons d'identifier la nature des ambiguïtés qui tendent à les faire baisser.

2.3.1. Scores inter-juges obtenus

Nous commentons ici les mesures inter-juges et scores obtenus dans le cadre spécifique de l'évaluation que nous avons réalisée, puis en les mettant en perspective avec les pratiques et échelles utilisées dans le domaine de l'analyse qualitative.

2.3.1.1. Accord inter-juge dans le cadre spécifique de l'évaluation

Dans le cadre spécifique de cette évaluation, la mesure de l'accord inter-juge est utilisée comme seuil de filtrage, que nous avons établi selon les modalités que nous avons évoquées précédemment. Comme on peut le calculer à partir des chiffres donnés dans le [TABLEAU 10](#), 84,8% des verbatim parviennent à susciter un accord majoritaire entre analystes concernant son Niveau d'appartenance. Ce chiffre s'élève à 65,5% en ce qui concerne les Critères. Dans les deux cas, cet accord majoritaire peut porter sur le fait que le verbatim soit inclassifiable. Un accord absolu, c'est-à-dire pour lequel tous les évaluateurs attribuent la même classe, et que cette classe correspond à celle consignée dans les archives n'est atteint que pour 11,15% des verbatim en ce qui concerne les Niveaux, et 6,09% pour les Critères. Un accord unanime entre les évaluateurs (différent ou non de la classification des archives) n'est atteint que sur 21,71% des verbatim pour les Niveaux, et 12,49% pour les Critères. Ces scores sont donc bas et signifient que moins d'un cinquième des verbatim utilisés dans les études sont classifiables hors contexte avec un degré de certitude élevé.

2.3.1.2. Accord inter-juge tel que calculé en général pour les études qualitatives

Nous avons souhaité évaluer l'accord inter-juge avec une métrique communément utilisée dans le cadre des analyses qualitatives, afin d'estimer, dans une certaine mesure, la stabilité de l'annotation qui est permise par la méthode. L'expression « dans une certaine mesure », ici, est importante ; si le score obtenu permet d'avoir une idée de cette stabilité, il ne faut pas oublier que, si nous avons mené cette évaluation dans le cadre d'une étude précise, les analystes auraient tous eu pleinement connaissance du projet et des textes entiers, comme c'est le cas dans la pratique normale de l'analyse, et cela aurait

très certainement donné lieu à un score bien plus élevé. Nous ne cherchons donc pas, ici, à évaluer la stabilité générale du codage permise par la grille CAUTIC® et sa maîtrise experte, mais bien celle qu'il est permis d'atteindre avec des contextes restreints, une absence de connaissance pragmatique du sujet du projet et des cerveaux humains experts.

En analyse qualitative, l'accord inter-codeur est utilisé en tant que gage de fiabilité de l'analyse (on parle d'ailleurs d'*Intercorder Reliability* dans la littérature anglophone). Calculer ce score n'est pas une pratique systématique et on constate un manque de lignes directrices de recommandations consensuelles à ce sujet (O'Connor & Joffe, 2020). La variabilité de la fréquence du recours à cette pratique est principalement due aux différents courants d'analyse ; elle est souvent considérée comme faisant partie intégrante de la pratique du *content analysis*, mais est beaucoup moins appliquée dans les approches de type Théorie Ancrée car considérée comme incompatible avec la nature incrémentale et récursive du codage qui y est employé (*Ibid.*). Les mesures utilisées sont également variables selon les particularités des études⁹⁵, mais il est communément admis que les méthodes à base de pourcentage simple ne sont pas des plus appropriées car elles ne prennent pas en compte le facteur chance (*Ibid.*⁹⁶).

Nous choisissons d'utiliser la mesure de l'Alpha de Krippendorff, qui tend à être recommandée depuis quelques années, principalement dans le cadre du *content analysis* (Allen, 2017, p. 151). Cette mesure prend en compte le facteur chance, et a pour avantages : de pouvoir calculer un score quel que soit le nombre de codeurs et de codes, de pouvoir utiliser des données incomplètes (i.e. tous les codeurs n'ont pas tous codés exactement les mêmes segments), et de ne pas requérir un minimum de volume de données (Krippendorff, 2011). Cela convient donc tout à fait aux paramètres de notre évaluation. L'Alpha de Krippendorff s'interprète de la manière suivante : 1 correspond à un accord parfait, 0 à un désaccord parfait, et les scores négatifs à un désaccord systématique, supérieur à celui obtenu par chance. On considère les scores >0,80 comme garants d'une fiabilité certaine, et 0,667 est considéré comme le plus bas score acceptable (Allen, 2017, p. 146).

⁹⁵ Principalement : Kappa de Cohen, Kappa de Fleiss, Méthode Holsti et Pi de Scott.

⁹⁶ Citant Cohen (1960), Hallgren (2012) et Lombard et al. (2002).

Nous calculons ainsi ce score⁹⁷ selon les modalités suivantes : nous prenons en compte tous les verbatim qui ont été évalués (même si ayant reçu moins de trois évaluations), comptons la classe *Inclassifiable* comme une classe à part entière, et calculons ces scores pour l'ensemble des analystes, puis par groupes d'ancienneté, pour l'accord sur les Niveaux et celui sur les Critères. Les résultats obtenus sont présentés dans le TABLEAU 14. Ces scores sont donc très bas mais ils sont, comme ceux présentés au point précédent, à interpréter en prenant en compte que l'exercice demandé ne correspond pas aux conditions normales d'analyse. Ils sont toutefois un indicateur de la difficulté de la tâche et du degré d'interprétation qui est nécessaire à sa réalisation. D'autre part, sans surprise, le groupe *séniors* obtient des scores significativement plus élevés (d'ailleurs, deux des analystes séniors – S1 et S4 – obtiennent entre eux un score de 0,60). Cela semble indiquer que l'acquisition progressive de l'expertise tend à uniformiser le processus d'interprétation plutôt que de pousser à prendre de plus en plus de libertés interprétatives.

	Niveaux Critères	
Accord inter-codeur global	0,3	0,22
Accord inter-codeur groupe <i>séniors</i>	0,36	0,26
Accord inter-codeur groupe <i>juniors</i>	0,21	0,16

TABLEAU 14 : Score Alpha de Krippendorff pour l'accord inter-juge obtenu sur l'annotation des verbatim issus des archives dans le processus de filtrage du corpus

2.3.2. Nature des ambiguïtés qui diminuent l'accord inter-juge

L'accord inter-juge étant très bas, nous cherchons à identifier les ambiguïtés qui tendent à le faire baisser.

2.3.2.1. Ambiguïtés inhérentes aux classes

Nous souhaitons évaluer si, à ce stade de notre travail, il est déjà possible de repérer des ambiguïtés inhérentes aux classes, ce qui constituerait un des éléments d'explication de l'accord inter-juge relativement faible. En effet, si des couples de Niveaux ou Critères appellent naturellement des données proches, la difficulté de choisir entre les deux devrait être traduite par des désaccords inter-juges plus significatifs sur ces derniers. Nous créons les matrices de confusion présentées dans les TABLEAUX EN ANNEXE 4

⁹⁷ Nous utilisons pour cela le module Python *fast-krippendorff* (<https://github.com/pln-fing-udelar/fast-krippendorff>, consulté le 15/09/2021).

à 7 (pp. 272-274). Les deux premières confrontent les classifications originales contenues dans les archives avec celles obtenues suite au filtrage, et les deux dernières comparent deux à deux toutes les évaluations réalisées sur les mêmes verbatim par les analystes.

En ce qui concerne les Niveaux, c'est entre le Niveau 1 et le Niveau 2⁹⁸ que l'on observe le plus de confusion, que ce soit par comparaison avec les classifications des archives ou au niveau de l'accord inter-juge global entre évaluateurs. Une partie de cette confusion est certainement explicable par la proximité entre les Critères 1.3 et 2.3 et celle entre les Critères 1.4 et 2.4⁹⁹, qui transparaissent sensiblement dans les matrices qui portent sur les Critères, mais pas de manière nettement significative. D'autre part, il est intéressant de constater que les verbatim classés dans le Niveau 3 – qui porte sur l'identité et les valeurs du participant – dans les archives ont tendance à être significativement classés dans le Niveau 2 lors de l'évaluation hors contexte, alors que ces deux Niveaux ne sont pas significativement confondus entre les évaluateurs. Il est probable que cela soit dû au fait que l'assignation à ce Niveau 3 nécessite une interprétation qui fait appel à la perception générale du positionnement de l'individu sur certains sujets qui n'est saisissable qu'à travers le texte entier. Cette perception générale est mobilisée sur des phrases qui, isolément, ne permettent pas de déterminer si la motivation sous-jacente à l'opinion ou au fait énoncé est d'ordre pratique ou éthique ; le verbatim « *Si on fait appel qu'à l'Intelligence Artificielle sur la mise à la terre, c'est niet.* » peut être motivé par un aspect pratique (l'utilisateur croit que ça ne marchera pas) ou relatif aux valeurs (l'utilisateur n'a pas confiance en l'Intelligence Artificielle et favorisera toujours l'expertise humaine). Or, dans le deuxième cas, il est utile d'avoir précédemment rencontré dans le texte des énoncés qui précisent l'opinion de l'utilisateur sur l'Intelligence Artificielle, ou plus généralement son rapport aux nouvelles technologies, pour pouvoir identifier la motivation de ses propos dans le verbatim en question.

⁹⁸ Pour rappel, le Niveau 1 porte sur la compréhension et le Niveau 2 sur l'adéquation aux pratiques.

⁹⁹ Pour rappel :

1.3 – « *L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?* »

2.3 – « *La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?* »

1.4 – « *L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?* »

2.4 – « *Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?* »

Pour ce qui est des Critères, on constate notamment des signes d'ambiguïté légèrement significatifs entre les Critères 1.1 et 1.2¹⁰⁰ ; cela est tout à fait compréhensible de par le fait qu'ils portent tous deux sur la compréhension du Concept, et que distinguer ce qui relève de la compréhension du Concept global (1.1) ou de la compréhension du principe technique en particulier (1.2) n'est pas chose aisée, d'autant plus quand le Concept n'est pas connu de l'évaluateur. On remarque également une certaine confusion entre les Critères 2.3 et 2.4¹⁰¹, qui peut s'expliquer par le fait que la comparaison avec l'existant (2.3) et l'expression d'une potentielle résolution de problème (2.4) impliquent tous deux une notion d'avant/après. Par exemple, dans « *ça semble sans risque, car plus de manipulation de fil à effectuer* », on perçoit une comparaison avec une situation passée dans laquelle la manipulation de fils était nécessaire, mais savoir si cette manipulation constituait ou non un problème, hors contexte, est plus compliqué. D'autres signes d'ambiguïtés sont identifiables (1.1-3.3, 3.1-3.2, 4.3-3.2 par exemple), mais, quoi qu'il en soit, nous n'observons pas, que ce soit pour les Niveaux ou les Critères, de signaux très forts qui impliqueraient, par exemple, de considérer que l'ambiguïté est telle qu'il vaudrait mieux, dans le cadre de ces travaux, fusionner deux classes. Les proximités observées devront toutefois être confirmées et rester un point d'attention particulier.

2.3.2.2. Phénomènes linguistiques qui influent sur l'accord inter-juge

Après ce point relatif à l'ambiguïté induite par les Critères-mêmes, intéressons-nous à celle inhérente aux textes qui influe sur l'accord inter-juge. Le lecteur pourra trouver des illustrations de nos prochains propos dans les [TABLEAUX EN ANNEXE 8 à 10](#) (pp. 275-285) qui fournissent des exemples commentés de, respectivement, verbatim qui suscitent un accord inter-juge fort, verbatim unanimement considérés comme inclassifiables, et verbatim qui provoquent un fort désaccord entre les analystes.

Au regard de ces exemples, il semble, sans surprise, que le manque de contexte et la nécessité non satisfaite d'une interprétation pragmatique soient les éléments qui poussent les analystes à évaluer le verbatim comme inclassifiable. Cependant, ces deux

¹⁰⁰ Pour rappel :

1.1 - « *L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?* »

1.2 - « *Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?* »

¹⁰¹ Pour rappel :

2.3 - « *La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?* »

2.4 - « *Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?* »

caractéristiques ne sont pas uniquement relatives aux verbatim considérés comme inclassifiables, et il est intéressant de regarder, par comparaison, celles qui provoquent un fort désaccord. En effet, certains verbatim ont manifestement poussé les analystes à faire des suppositions malgré le manque manifeste d'informations, et le fait que ces suppositions soient différentes selon les analystes montre qu'une forte dimension interprétative subjective a été mobilisée sur ces verbatim en particulier. Précisons que, dans certains cas, il est très probable que le verbatim aurait pu être classé dans plusieurs Critères même en toute connaissance de l'étude et du texte entier, ce qui justifie les différentes suppositions. Par exemple, on comprend que le verbatim « *Il faut un enregistrement derrière et alors, les gens seront motivés. Ils seront plus responsables.* » puisse être classifié dans des Critères différents qui traitent des fonctions du Concept et de la relation à l'environnement social. Cela constitue d'ailleurs la source principale d'ambiguïté des verbatim ; le même segment contient, tout simplement, des éléments qui renvoient légitimement à plusieurs Critères.

Dans les cas où des suppositions sont faites par des analystes, il est intéressant de constater qu'il semble que ces derniers tentent de compenser le manque de contexte et de connaissance pragmatique du sujet de l'étude par la recherche d'indices sémantiques et discursifs et par leur connaissance pragmatique de la méthode, de son application et des réponses fréquentes des participants. Par exemple, le rattachement de « *C'est plus attractif que ce que je connais : ce qui n'est pas négligeable quand on fait une maladie comme la mienne* » à 2.4 – « *Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?* » se fait par le lien sémantique entre *maladie* et *problème*, et par l'identification d'une comparaison avec une situation actuelle. Celui entre « *Ça va nous enlever du boulot.* » et 3.2 – « *Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?* » est pragmatique et lié à la connaissance des problématiques générales de terrain ; l'analyste imagine, parce qu'il l'a souvent entendu lors d'études, que l'interprétation de cette phrase ne s'arrête pas à ce que la sémantique pure permet de comprendre (i.e., une charge de travail réduite), mais que cette modification de la répartition de la charge de travail pourrait avoir un impact sur le rôle professionnel de l'utilisateur (par exemple, cela pourrait entraîner des licenciements, ou une réorganisation des tâches assignées au personnel qui pourrait impacter directement le rôle de l'utilisateur dans son entreprise). Cependant, il est possible qu'il s'agisse là d'une

surinterprétation, et que le verbatim soit simplement l'expression du fait que l'utilisateur se réjouisse d'avoir moins de travail à accomplir.

En somme, d'une part, les phénomènes qui amoindrissent l'accord inter-juge sont bien, comme nous l'avons identifié *a priori*, le manque de contexte et l'impossibilité d'une interprétation pragmatique permise par la connaissance du projet. D'autre part, il semble que pour contrer ces difficultés, les analystes développent une stratégie qui mobilise la reconnaissance d'indices sémantiques et discursifs, mais aussi leur capacité d'interprétation pragmatique permise non pas par la connaissance du sujet traité mais par celle acquise par la pratique de l'analyse CAUTIC® au fil du temps. Or, cette dernière est à double tranchant car elle peut permettre d'identifier, effectivement, des intentions implicites, mais peut aussi provoquer des surinterprétations et des classifications qui paraîtraient erronées si l'on avait accès au texte entier.

En conclusion de ce chapitre, nous soulignons le fait que le corpus qui sera utilisé dans cette partie sera, pour des raisons pragmatiques, de taille assez restreinte. Cependant, les données les plus ambiguës ont été éliminées, et la répartition des verbatim qu'il contient est représentative de l'activité de l'entreprise et des pratiques des analystes. En outre, les scores inter-juges qui ont été calculés lors de la phase de filtrage sont significativement bas, mais doivent être interprétés en prenant en compte le fait que l'exercice demandé ne correspond pas aux conditions normales d'analyse. En revanche les paramètres de cet exercice sont assez similaires à ceux dont disposera le système.

En effet, la difficulté d'une interprétation – et donc de la classification – des verbatim sans pouvoir mobiliser une dimension pragmatique permise par la connaissance du sujet est majeure, mais constituera nécessairement une caractéristique du système ; s'il est possible de lui donner en entrée quelques éléments d'information (domaine d'activité par exemple), il est inenvisageable que l'on puisse parvenir à un niveau de connaissance et de compréhension similaire à celui d'un analyste humain. La stratégie d'utilisation de connaissances pragmatiques liées à la pratique de la méthode permettant de repérer quels types d'énoncés sont susceptibles de sous-entendre des éléments implicites liés à un certain Critère est davantage envisageable, mais cette stratégie a ses limites même pour un codeur humain. La deuxième difficulté d'interprétation et de classification réside dans la nécessité de prise en compte du texte entier pour certains aspects. S'il est possible d'appliquer des méthodes, telles que la

résolution d'anaphores, pour rendre plus explicite le contenu de quelques phrases à partir de l'intégralité du texte, établir des liens entre les différentes parties du texte qui permettent de prendre en compte des particularités du participant pour analyser un segment semble complexe. Par exemple, si pour interpréter « *il faut 10m² d'espace libre pour l'installer* » comme 'je n'ai pas la place chez moi' il faut avoir lu précédemment « *j'habite dans un petit appartement* », il faudrait que le système soit en mesure d'identifier, comprendre et mémoriser tous les éléments d'information de ce type et soit capable de les réutiliser pour l'analyse. Or, la multitude des formes et des sujets possibles abordés à repérer, ainsi que l'identification des segments sur lesquels ils peuvent aider à l'analyse, paraissent très complexes. Il est donc attendu que cette double difficulté constitue les principaux freins et limites du système.

CHAPITRE V : EXPÉRIMENTATIONS BASÉES SUR DES APPROCHES STATISTIQUES

Traditionnellement opposées aux approches symboliques, les méthodes statistiques et probabilistes (regroupées sous le terme de méthodes *stochastiques*) commencent à susciter un intérêt particulier au début des années 1990 ; les nouvelles possibilités permises par les progrès technologiques en informatique provoquent des demandes à visée pratique. Ce type de demande avait initialement motivé la création de la discipline du TAL dans les années 1940, puis avait été abandonné au profit de recherches à visée théorique suite au rapport de l'*Automatic Language Processing Advisory Committee* (ALPAC) en 1964 qui exprimait une vision très critique et peu optimiste quant au potentiel de réussite du TAL pour la traduction automatique (Cori, 2008). Dès lors, on assiste à un engouement pour des méthodes de « TAL robuste » – opposées, donc au « TAL théorique » – dont les méthodes stochastiques constituent une partie centrale (*Ibid.*). Parmi les applications possibles, la classification de texte va pleinement profiter de cet élan. Aujourd'hui, les approches statistiques sont prédominantes dans le TAL.

Dans ce chapitre, nous présentons une exploration des possibilités permises par ces méthodes pour la tâche qui nous incombe. Plus en détails, il s'agira :

- a) De mesurer l'impact précis de limites pré-identifiées inhérentes à nos données ou à la tâche qui nous est confiée sur la possibilité de recours à des méthodes statistiques (ces limites étant la petite taille du corpus par rapport au nombre de classes, son déséquilibre ainsi que la complexité sémantique de la tâche de classification). Par cette démarche, nous cherchons à apporter un début de réponse aux interrogations transverses de cette thèse 1) b. et c. soulevées au point 9.2 de la Partie I (p. 84).
- b) D'initier une caractérisation de nos données en prévision d'expérimentations futures.
- c) Potentiellement, de disposer d'un premier outil de classification.

Pour cela, nous commençons par une exploration textométrique simple de notre corpus afin d'étudier ce que la statistique lexicale peut révéler en termes de

caractérisation du contenu textuel attendu pour chaque Critère CAUTIC®. Puis, nous testons différentes techniques de classification textuelle, en présentant dans un premier temps ce domaine puis, et dans un deuxième temps, nos expérimentations.

3. Exploration textométrique

Le textométrie (également nommée *lexicométrie* ou *logométrie*) repose sur « des procédures de tris et de calculs statistiques pour l'étude d'un corpus de textes numérisés » (Pincemin, 2011, p. 259). Elle s'inscrit dans une logique de proximité avec le texte et est au croisement de la statistique lexicale et de l'analyse multidimensionnelle lexicale (Mayaffre et al., 2019). Pour le résumer d'une manière simple, une analyse textométrique va consister à appliquer une analyse quantitative sur les textes en mobilisant du décomptage d'occurrences pour effectuer des calculs statistiques divers ; on cherchera par exemple à extraire des fréquences d'occurrences, identifier des co-occurrences ou encore à estimer le degré de spécificité d'un terme au corpus. Par nature, la textométrie a une double visée descriptive et contrastive (Buhler et al., 2018; Comby, 2016). En cela, elle permet d'identifier des caractéristiques propres à un texte et de mettre ainsi en évidence ses particularités par rapport à d'autres (entre différents corpus, ou bien entre différentes parties d'un même corpus). Dans le cadre de la tâche qui nous intéresse, une telle analyse peut donc permettre de discerner des particularités textuelles attendues par chaque classe dans une tâche de classification, et ce de manière humainement explicitable et compréhensible. Cela rend donc *a minima* possible un processus descriptif, mais les informations ainsi recueillies peuvent également être mobilisées de manière opérationnelle ; Eensoo & Valette (2012) utilisent les résultats d'analyses lexicométriques, transformés en critères, comme entrées à des algorithmes de classification pour l'analyse des sentiments et obtiennent ainsi de meilleures performances qu'avec des entrées de type mots simples, lemmes et n-grams¹⁰².

3.1. Méthodologie

Dans ce qui constitue notre première tentative pour « faire parler nos données », nous décidons donc d'effectuer une analyse textométrique relativement simple sur le

¹⁰² La tâche qu'il réalisent ne comporte cependant que deux classes (tonalités gaie ou triste d'un récit) dont on peut supposer qu'elles contiennent des vocabulaires nettement plus distincts l'un de l'autre que dans le cas qui nous intéresse.

corpus que nous avons présenté au chapitre précédent. Nous entendons par cela profiter des vertus descriptives et contrastives d'une telle analyse, évoquées ci-dessus, pour essayer de discerner ce qui caractérise le contenu textuel attendu pour chaque Critère et, par là même, identifier des éléments discriminants et évaluer leur force.

Pour cela, nous lemmatisons le corpus avec *TreeTagger*¹⁰³ (Schmid, 1994), puis conservons les lemmes des verbes, noms, adjectifs, adverbes, pronoms, prépositions et conjonctions. Nous remplaçons les déterminants, les signes de ponctuations et les chiffres par leurs étiquettes (respectivement *DET*, *PUN* et *NUM*). Ensuite, nous extrayons les unigrams, bigrams, trigrams et quadrigrams. Pour chacun, nous calculons leurs fréquences dans chaque Critère, ainsi que leur répartition entre les Critères en termes de pourcentages (i.e., un n-gram présent quatre fois dans le corpus dont trois pour le Critère 1.2 et une dans le 1.4 aura un score de 0,75 pour le Critère 1.2 et 0,25 pour le 1.4). Nous extrayons également la mesure de dispersion des n-grams en fonction du nombre d'études dans le contexte desquelles ils ont été produits ; en d'autres termes, chaque verbatim étant attribué à une étude d'origine, si le bigram *poser problème* apparaît dans des verbatim issus de quatre études différentes, le score de dispersion sera de 4. A partir de ces données, nous conservons et consignons, pour chaque Critère, les n-grams les plus fréquents, avec des seuils minimaux de fréquence = 5 et dispersion = 3. Nous consignons également les n-grams les plus représentatifs du Critère, i.e. ceux dont au moins 50% des occurrences dans le corpus interviennent dans ce Critère, avec des seuils minimaux de fréquence = 2 et dispersion = 2. Enfin, pour chaque Critère également, nous extrayons les unigrams les plus spécifiques par rapport à un corpus de français général, à l'aide de l'outil *Anatext*¹⁰⁴ qui dispose de fréquences de référence pour le français¹⁰⁵ et permet de restituer cette information pour chaque lemme (mesure de *loglike*). Un exemple de ces informations ainsi extraites est présenté dans le **TABLEAU EN ANNEXE 11** (pp. 286-290) ; dans ce tableau, pour plus de lisibilité, nous avons retiré les n-grams qui ne présentaient pas d'intérêt pour l'analyse, comme les unigrams */être/* et */avoir/*, les prépositions seules ou bien les n-grams sémantiquement vides tels */de ce que/* ou */et qui être/*. Nous y distinguons par des codes couleurs, toujours dans un souci de lisibilité, les n-grams :

¹⁰³ <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/> (consulté le 15/10/21).

¹⁰⁴ <http://phraseotext.univ-grenoble-alpes.fr/anaText/> (consulté le 15/10/21).

¹⁰⁵ Issues de Lexique.org (New et al., 2001) (<http://www.lexique.org/>, consulté le 15/10/21).

- qui appartiennent à un champ lexical qui semble être appelé par le Critère : par exemple, pour le Critère 4.1 – « *L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?* », nous attendons des champs lexicaux qui gravitent autour des concepts de MARCHÉ, TENDANCES et PLAUSIBILITÉ et identifions donc des n-grams tels que */marché/*, */produit/*, */évolution/*.
- courants dans la langue : */pouvoir/*, */cela sembler/*, */chose comme cela/*
- qui servent potentiellement à exprimer une opinion/un jugement mais qui ne sont pas relatifs aux champs lexicaux du Critère : */ce être pas mal/* ou */être bien/*
- qui appartiennent à un socio-dialecte technique : */câbler/*, */matériau/*, */diagnostic/*
- qui appartiennent à un jargon commercial : */coût/*, */produit/*, */consommateur/*

Sur la base de ces tableaux, nous tentons ensuite de dégager des caractéristiques lexicales discriminantes pour chaque Critère.

De plus, afin de nous intéresser à la pertinence de chaque catégorie morphosyntaxique en termes de pouvoir discriminant entre les Critères, nous effectuons un comptage du nombre d'occurrences de chaque catégorie parmi les n-grams les plus spécifiques à chaque Critère. Pour cela, nous supprimons les doublons par enchevêtrement, c'est-à-dire, par exemple, lorsqu'une liste contient */je venir/* avec trois occurrences et */je venir de/* avec deux occurrences, nous ne conservons que */je venir/*. Nous assignons à chaque n-gram restant une catégorie morphosyntaxique selon les modalités suivantes¹⁰⁶ :

- De manière générale, nous prenons en compte la tête du groupe : */à comprendre/* reçoit la catégorie *V*.
- Les constructions *Verbe + Adj* et *Verbe + Nom* sont comptées comme appartenant aux deux catégories, sauf si le verbe est une copule : */gagner*

¹⁰⁶ Bien entendu, ce processus est soumis aux limites interprétatives que nous soulignerons au point suivant ; nous devons parfois faire des choix arbitraires, par exemple, choisir entre *N* et *Adj* pour */concurrent/*. Cela aurait pu être évité si nous avions extrait nos n-grams avec leurs étiquettes. Cependant, effectuer cette distinction aurait réduit le nombre de n-grams extraits ; les occurrences de */concurrent/ = Adj* et */concurrent/ = N* auraient donné lieu à deux n-grams différents, avec par conséquent des fréquences réduites. Or, la taille de notre corpus ne permet déjà d'extraire que peu de n-grams. De surcroît, des erreurs de désambiguïsation de l'étiqueteur auraient également inclus leur part d'approximation dans l'analyse.

du temps/ est étiqueté comme *V+N*, et */être intéressant/* est étiqueté comme *Adj* seulement.

- Nous ignorons les pronoms : */je prendre/* est étiqueté comme *V*.
- Les locutions clairement identifiables en tant que telles sont assignées à la catégorie à laquelle elles appartiennent : */pour DET instant/* est étiqueté comme *Adv*.
- Les constructions à verbe support avec nom prédicatif reçoivent une étiquette spécifique : */avoir envie/* est étiqueté comme *VSup*.
- Les constructions à verbe auxiliaire + infinitif sont comptées comme un seul verbe : */faire penser/* est étiqueté comme *V*.
- Nous supprimons les locutions prépositionnelles (peu fréquentes) : */DET certain nombre de/* est supprimé.

Sur la base de ces étiquetages et comptages, nous observons quelles catégories morphosyntaxiques semblent être les plus discriminantes pour chaque Critère, en termes de nombre de lemmes différents et de nombre d'occurrences de ces lemmes dans le corpus. Les résultats sont consultables dans les [FIGURES EN ANNEXE 8 et 9](#) (p. 291).

3.2. Résultats

Nous avertissons ici le lecteur quant aux limites de la démarche d'interprétation des résultats que nous proposons, puis présentons ces résultats en commençant par une synthèse par Critère et générale des phénomènes observés, puis par une synthèse de l'analyse par catégorie morphosyntaxique.

3.2.1. Limites de l'interprétation

En préambule à la présentation des résultats, nous souhaitons souligner deux limites à leur interprétation. La première est l'impossibilité de prise en compte de la polysémie ; en observant des n-grams isolés, il nous est impossible de distinguer leurs différentes acceptions dans le corpus, et nous sommes donc contrainte de procéder par supposition et d'interpréter chaque n-gram comme étant monosémique, et avec le sens qu'il est le plus susceptible d'avoir. Nous sommes pour cela parfois guidée – mais aussi potentiellement trompée – par notre connaissance *a priori* du contenu du corpus. Ainsi, nous interprétons */réseau/* comme un terme technique et non comme 'tissu de relations', car nous savons que des problématiques de réseaux (de communication informatique) sont abordées dans les projets desquels nous avons tiré nos verbatim. Il est

cependant tout à fait possible que d'autres acceptions soient présentes dans le corpus. Cette approximation aurait pu être évitée en utilisant un concordancier et en réalisant un filtrage manuel pour les n-grams qui nous semblaient revêtir un fort potentiel d'ambiguïté. Cependant, nous avons estimé que le temps nécessaire pour cela ne méritait pas d'être investi, au regard du nombre raisonnable de n-grams ambigus, et parce qu'il ne s'agit que d'une démarche exploratoire préliminaire à d'autres travaux pour laquelle une marge d'erreur est permise.

D'autre part, les fréquences des n-grams qui semblent intéressants pour discriminer les Critères sont somme toute assez faibles ; les seuils très bas utilisés pour filtrer les n-grams spécifiques à un Critère (fréquence = 2, dispersion = 2 et au moins 50% des occurrences dans le Critère) ont permis d'identifier, au total pour tous les Critères, 725 n-grams. Or, seuls 74 d'entre eux ont une fréquence >5, dont 14 avec une fréquence >10¹⁰⁷. Nous avons cependant dû utiliser ces seuils pour conserver un nombre convenable de n-grams à observer, mais les conclusions que nous pouvons tirer de cette observation, de par ces fréquences basses sont à considérer avec précaution à cause d'un risque non négligeable de phénomènes accidentels. Ce problème est principalement dû à la faible taille de notre corpus. Cette taille est d'ailleurs extrêmement réduite lorsque ce corpus est divisé en sous-corpus par Critère qui sont, de surcroît, très déséquilibrés. Cela a notamment conduit au fait que certains Critères (1.5, 3.2 et 4.3) ne peuvent tout simplement pas donner lieu à des observations car aucun – ou très peu – de n-grams ne leur sont spécifiques ; cela n'est pas étonnant car ils ne comportent que, respectivement, huit, quatre et quatre verbatim (cf. TABLEAU 11, p. 97), ce qui n'est pas suffisant pour voir émerger des phénomènes contrastifs.

3.2.2. Synthèses par Critères et synthèse générale

Les synthèses par Critères sont présentées dans les TABLEAUX 15 à 18 des pages suivantes (pp. 112 à 115). Le premier constat que l'on peut dresser à leur propos est que la proportion d'indices lexicaux discriminants est très inégale d'un Critère à l'autre. En effet, certains d'entre eux révèlent un vocabulaire relativement cohérent et sémantiquement autoporteur. Par exemple, à la lecture des grilles de n-grams du Critère

¹⁰⁷ En tout, nous avons extrait du corpus 74 192 n-grams.

Critère et champs lexicaux attendus	Synthèse	Signaux émergents
<p>1.1 – « L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ? » => intérêt, compréhension</p>	<ul style="list-style-type: none"> Le champ lexical de l'intérêt est bien représenté dans les n-grams les plus fréquents (/intérêt/, /intéressant/, /être intéressant/) mais pas dans les n-grams spécifiques au Critère. Ils apparaissent cependant comme les plus spécifiques par rapport au corpus de référence, mais cela est aussi le cas pour d'autres Critères (ex : 1.4 et 2.5) ce qui tend à montrer que la sur-représentation de ce champ lexical est davantage inhérente au corpus qu'à un Critère en particulier. Le champ lexical de la compréhension est pour sa part bien présent parmi les n-grams spécifiques au Critère (/paraître clair/, /que je retenir/). 	<ul style="list-style-type: none"> Adverbes qui indiquent une comparaison avec les pratiques existantes (/pour DET instant on/). Adjectifs descriptifs du Concept (/esthétique/, /ludique/).
<p>1.2 – « Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ? » => compréhension, éléments techniques</p>	<ul style="list-style-type: none"> Le champ lexical de la compréhension est bien présent dans les n-grams spécifiques au Critère, mais presque seulement incarnés par le lexème compréhension. De nombreux n-grams relatifs à la technique (/appareil/, /kilo/, /moteur/, /réseau/) sont bien présents dans les mots fréquents et spécifiques, mais cela est généralement le cas pour les autres Critères, ce qui indique une fois de plus une spécificité du corpus et non du Critère. 	<ul style="list-style-type: none"> Indices sur la présence de questions (/question/, /combien de/).
<p>1.3 – « L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ? » => comparaison, choses existantes</p>	<ul style="list-style-type: none"> De nombreux indices relatifs à la comparaison (/ressembler/, /différent/, /DET même chose/, /faire penser/, /équivalent/) et aux produit/services existants (/chose qui exister déjà/, /ce que on retrouver/, /sur DET marché/) sont présents dans les n-grams fréquents comme dans les n-grams spécifiques au Critère. 	<ul style="list-style-type: none"> Champ lexical émergent (autre que directement attendu par le Critère) : existence/sortie sur le marché
<p>1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? » => choix, fonctions, intérêt</p>	<ul style="list-style-type: none"> Le champ lexical de l'intérêt est fortement représenté parmi les n-grams fréquents comme dans les n-grams spécifiques (/cela pouvoir être intéressant/, /être aussi intéressant/), mais on retrouve ce champ lexical également pour d'autres Critères. La mention explicite des fonctions est également bien présente dans les n-grams fréquents et spécifiques. Cependant, au-delà de cette mention explicite, on ne retrouve pas de nom de fonction spécifique (on aurait pu s'attendre à des mots tels que <i>enregistrement, tracking, reconnaissance vocale</i>, etc. qui n'apparaissent aucunement). 	<ul style="list-style-type: none"> Possibles désignations indirectes des fonctions par référence à une partie du Concept ou de sa présentation (/tout DET partie/ sur-représenté).
<p>1.5 – « L'usage du concept est-il envisageable comme un système d'usage courant ? » => utilisation, usage courant</p>	<p>[aucune donnée pertinente]</p>	

TABLEAU 15 : Synthèse de l'analyse textométrique pour les Critères du Niveau 1

Critère et champs lexicaux attendus	Synthèse	Signaux émergents
2.1 – « Le concept est-il simple à utiliser ? » => simplicité/complexité, utilisation	<ul style="list-style-type: none"> Le champ lexical de la facilité est bien présent dans les n-grams fréquents et spécifiques (<i>/air simple/, /facile/</i>). En revanche, celui de la complexité est complètement absent. Les n-grams faisant référence à l'utilisation sont plus discrets, mais apparaissent en revanche deux fois dans les n-grams spécifiques par rapport à un corpus de référence, ce qui laisse penser qu'il s'agit d'avantage d'une spécificité du corpus que du Critère. 	<ul style="list-style-type: none"> L'expression avoir l'air est fortement représentée, ce qui laisse imaginer une certaine récurrence de syntagmes du type <i>/ça a l'air facile à utiliser/</i>.
2.2 – « Le concept complèterait-il sans les concurrencer les habitudes existantes ? » => habitudes, concurrence/remplacement, nouvelles possibilités	<ul style="list-style-type: none"> Le champ lexical de la possibilité (<i>/permettre de/, /on pouvoir/</i>) est bien présent dans les n-grams fréquents mais non dans les n-grams spécifiques, laissant penser qu'il n'est pas discriminant. Le champ lexical du remplacement (<i>/remplacer/, /substitution/</i>) est discrètement présent dans les n-grams fréquents et spécifiques au Critère, et il est complété par l'idée d'ajout/cumul (<i>/de rajouter du/, /associer/</i>) et du changement (<i>/ce être DET évolution/</i>). 	[aucun]
2.3 – « La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ? » => comparaison, pratiques existantes, crédibilité/légitimité, nouvelles pratiques	<ul style="list-style-type: none"> Le champ lexical de la comparaison est bien présent et principalement actualisé par des locutions conjonctives dans les n-grams fréquents et spécifiques au Critère (<i>/ce être comme DET/, /quand même beaucoup plus/, /tandis que là/</i>). Les autres champs lexicaux attendus ne sont pas représentés. 	<ul style="list-style-type: none"> Champs lexicaux émergents (autres que directement attendus par le Critère) : questions, mouvement
2.4 – « Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ? » => problèmes, résolution	<ul style="list-style-type: none"> Les deux champs lexicaux des problèmes et de leur résolution sont fortement représentés dans les n-grams spécifiques au Critère (<i>/cela ne résoudre pas/, /risque de/, /cela faciliter/, /régler/, /aider/</i>). L'actualisation lexicale du concept de PROBLEME n'est pas régie que par des relations de synonymie avec problème (<i>/difficulté/</i>), mais aussi par hyponymie (<i>/moins de perte/</i>). 	<ul style="list-style-type: none"> La notion de gain de quelque chose et d'amélioration apparaît comme spécifique (<i>/faire gagner du temps/, /gagner vachement de/, /être plus précis/</i>) : cela va au-delà de la résolution de problèmes Autres champs lexicaux émergents : aide, espace disponible
2.5 – « L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ? » => organisation	<ul style="list-style-type: none"> Les n-grams extraits pour ce Critère forment un ensemble qui n'a aucune cohérence interne, et aucun lien particulier avec le Critère n'est identifiable. 	[aucun]

TABLEAU 16 : Synthèse de l'analyse textométrique pour les Critères du Niveau 2

Critère et champs lexicaux attendus	Synthèse	Signaux émergents
<p>3.1 – « L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ? » => identité, cible, type de personnes/de secteurs</p>	<ul style="list-style-type: none"> • Les champs lexicaux du ciblage (/se adresser à/, /destiner à/) et des désignations de (types de) personnes et de secteurs (/type de clientèle/, /DET public/, /tertiaire/), sont bien représentés dans les n-grams spécifiques au Critère. • Certaines désignations de types de personnes/secteurs (/bricoleur/, /chirurgie/) apparaissent comme spécifiques au corpus mais non spécifiques au Critère. Cela est dû à la faible dispersion de ces n-grams et indique qu'ils sont très variables d'une étude à l'autre. 	<ul style="list-style-type: none"> • On observe une présence significative de la préposition pour mais dans des syntagmes ambigus, qui peuvent indiquer ou non l'identification d'une cible (/pour DET personne/, /être pour DET/, /pas pour/, /pour moi non/, /pouvoir être intéressant pour/, /pour nous/).
<p>3.2 – « Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ? » => identité, rôle, statut</p>	[aucune donnée pertinente]	
<p>3.3 – « Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ? » => valeurs/morale/éthique</p>	<ul style="list-style-type: none"> • Les seuls n-grams en lien avec le Critère sont présents uniquement dans la liste des n-grams spécifiques par rapport à un corpus de référence (/recyclable/, /durable/, /qualitatif/, /onde/, /individualité/). Ces n-grams sont bien spécifiques au Critère (presque toutes ont 100% de leurs occurrences dans ce Critère) dans le corpus, mais leur dispersion est trop basse (une seule étude) pour passer le filtre que nous avons défini. Cela signifie que ce type d'unités lexicales semblent bien discriminantes mais sont très variables d'une étude à l'autre. 	[aucun]
<p>3.4 – « Le concept suscite-t-il un imaginaire d'appropriation ? » => imaginaire, noms d'animaux</p>	<ul style="list-style-type: none"> • Les noms d'animaux sont bien présents comme spécifiques au Critère (/renard/, /chien/, /chat/). 	<ul style="list-style-type: none"> • La locution verbale faire penser à est présente dans les n-grams spécifiques au Critère.
<p>3.5 – « Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ? » => nouvelles idées</p>	<ul style="list-style-type: none"> • Seule la présence du mot extension dans les n-grams spécifiques semble significative, et seulement deux occurrences sont présentes. Hormis cela, les n-grams extraits pour ce Critère forment un ensemble très restreint qui n'a aucune cohérence interne. 	[aucun]

TABLEAU 17 : Synthèse de l'analyse textométrique pour les Critères du Niveau 3

Critère et champs lexicaux attendus	Synthèse	Signaux émergents
<p>4.1 – « L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ? » => marché, tendances/air du temps, plausibilité</p>	<ul style="list-style-type: none"> • Les n-grams appartenant au champ lexical des marchés sont seulement présents dans les n-grams fréquents, ce qui indique qu'ils ne sont pas discriminants. • L'ensemble formé par les n-grams extraits est assez pauvre. 	<ul style="list-style-type: none"> • Les adverbes tard et trop tard semblent discriminants pour ce Critère mais ils sont ambigus.
<p>4.2 – « Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ? » => relations et rôles en entreprise/cercles proches</p>	<ul style="list-style-type: none"> • Seul le verbe discuter dans les n-grams spécifiques au Critère semble en lien avec le Critère. Le reste de l'ensemble des n-grams extraits est assez pauvre. • On trouve dans les n-grams spécifiques par rapport à un corpus de référence des désignations de rôles/métiers tels que /cardiologue/, /propriétaire/, /fournisseur/ et /locataire/. Ces n-grams sont bien spécifiques au Critère dans le corpus, mais leur dispersion est trop basse (une seule étude) pour passer le filtre que nous avons défini. Cela signifie que ce type d'unités lexicales semblent bien discriminantes mais sont très variables d'une étude à l'autre. 	<p>[aucun]</p>
<p>4.3 – « Le concept est-il adapté à la place qu'occupent les professionnels dans la filière économique / aux relations familiales de l'utilisateur et à leur évolution ? » => place/statut, réseau professionnel, relations familiales/amicales, évolution</p>	<p>[aucune donnée pertinente]</p>	<p>[aucun]</p>
<p>4.4 – « Le concept est-il adapté aux types d'organisation et à leur capacité d'évolution / aux manières de vivre de l'utilisateur et à leur évolution ? » => organisations, évolution, habitudes de vie</p>	<ul style="list-style-type: none"> • Le champ lexical de l'organisation (au travail) est présent dans les n-grams spécifiques au Critère, mais seulement à travers les noms /équipe/ et /organisation/. Le reste des n-grams extraits pour ce Critère forme un ensemble assez pauvre. 	<p>[aucun]</p>
<p>4.5 – « Quel prix l'utilisateur consent-il à payer pour ce concept ? » => argent, sommes, coût</p>	<ul style="list-style-type: none"> • Le champ lexical de l'argent est largement représenté dans les n-grams fréquents et spécifiques (/cela valoir/, /NUM NUM €/, /beaucoup plus cher/, /investissement/, /coût/). 	<p>[aucun]</p>

TABLEAU 18 : Synthèse de l'analyse textométrique pour les Critères du Niveau 4

2.4, sans même connaître ce Critère, on peut facilement se faire une idée de ce dont il s'agit ; il contient notamment de nombreuses occurrences de */résoudre/* et */problème/*. D'autres, en revanche, donnent lieu à des ensembles incohérents qui n'ont aucun potentiel discriminant. C'est par exemple le cas pour le Critère 2.5, dont les n-grams extraits comme spécifiques sont : */être derrière/*, */que je faire/*, */qui être derrière/*, */DET outil de/*, */moment je/*, */on être bien/*, */quelque chose qui pouvoir/*, */outil de/*, */se dire/*, */à jour/*, */aller fonctionner/*, */bloquer/*, */pouvoir DET/*, */pouvoir en/* et */traitement/*.

Nous observons en outre que des n-grams de certains champs lexicaux sont identifiés comme spécifiques par rapport à un corpus de référence mais pas pour un Critère en particulier par rapport aux autres. Cela démontre qu'ils sont en réalité spécifiques à notre corpus mais difficilement discriminants au sein de ce dernier ; *a minima*, leurs occurrences sont réparties entre deux Critères. A titre d'exemple, les lexies relatives à l'intérêt sont identifiées comme spécifiques par *AnaText* dans les sous-corpus des Critères 1.1 – « *L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?* » et 1.4 – « *L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?* ». Cependant, on ne trouve des occurrences considérées comme spécifiques au Critère que pour le 1.4 (alors qu'elles sont fréquentes également pour le 1.1 et font pleinement partie des informations que l'on est en droit d'espérer rencontrer pour identifier du contenu textuel pour celui-ci).

D'autre part, il semble, au vu de cette analyse, que certains Critères mobilisent des concepts qui peuvent être actualisés par un ensemble trop varié de lexies¹⁰⁸ pour que ces dernières ne puissent être identifiées par l'analyse. Cette variation est par ailleurs relative aux différents projets, ce qui est notamment mis en évidence par le fait qu'elles soient

¹⁰⁸ L'idée d'envisager le problème qui nous est posé sous l'angle d'une actualisation lexicale de concepts inhérents aux Critères est un parti pris ici qui peut paraître contradictoire avec la logique ascendante que nous utilisons dans cette partie ; une approche ascendante stricte consisterait à identifier, dans un premier temps, les concepts qui émergent des listes lexicales et à considérer ces concepts comme ceux associés aux Critères plutôt que d'assumer une posture descendante qui consiste à associer *a priori* des concepts aux Critères. C'est, en fait, ce que nous avons fait dans une étape préliminaire aux synthèses que nous proposons ; nous avons regroupé les n-grams en champs lexicaux et avons identifié ceux qui émergeaient fréquemment pour chaque Critère. Cependant, nous n'avons trouvé que peu de champs lexicaux significatifs non reliés directement aux concepts prévisibles à la lecture des intitulés des Critères ; les quelques-uns ainsi identifiés sont indiqués dans les colonnes *Signaux émergents* de nos tableaux de synthèse (au-delà de ceux-ci, seuls émergeaient des champs lexicaux inhérents au corpus en général : principalement du jargon technique et le champ lexical de la possibilité – omniprésent dans le corpus de par le fait que le Concept permet de nouvelles pratiques). Ainsi, nous choisissons cet angle plutôt descendant ici pour la simple raison qu'une démarche complètement ascendante n'apporte que très peu d'informations, tandis que l'idée d'actualisation lexicale de concepts *a priori* inhérents aux classes permet de mettre en exergue davantage de phénomènes (notamment des « carences » entre l'attendu *a priori* et la réalité du corpus, mais aussi l'identification de vocabulaires possibles que la simple analyse de fréquences associée à un regroupement en champs lexicaux ne permet pas d'identifier).

souvent identifiées comme spécifiques par *Anatext*, mais ne passent pas le filtre de dispersion que nous avons imposé. Par exemple, les noms de fonctions du Concept, utiles pour discriminer les verbatim du Critère 1.4 sont extrêmement variables d'un projet à l'autre, et aucune d'entre elles n'apparaît dans les résultats filtrés (on aurait pourtant pu imaginer que des noms de grands types de fonctions tels que *l'enregistrement* ou le *téléchargement*¹⁰⁹, par exemple, soient présents dans les grilles filtrées). C'est également le cas pour les noms et adjectifs qui ont une connotation relative aux valeurs morales/éthiques/culturelles attendues pour le Critère 3.1 – « *Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?* » ; parmi les n-grams repérés comme spécifiques par *AnaText*, on trouve */recyclable/*, */durable/*, */qualitatif/*, */onde/*¹¹⁰ et */individualité/*. Or, ces n-grams ne sont pas présents dans la liste de ceux spécifiques au Critère car ils ne sont, chacun, présents que dans une seule étude. De manière générale, en outre, tout ce qui a trait aux pratiques et aux habitudes – concepts centraux de l'analyse CAUTIC® – donne lieu à un vocabulaire très variable, ce qui rend très difficile leur identification dans le corpus sur une base statistique.

Degré de stabilité d'une étude à l'autre	Relation avec les lexies exprimées dans l'intitulé du Critère	Exemples
Stable	Dérivation	<i>intéressant</i> → <i>intérêt</i> , <i>intéressant</i>
	Synonymie	<i>simple</i> → <i>facile</i>
	Antonymie ?	?
Variable	Hyponymie	<i>problème</i> → <i>perte</i> **
	Connotation*	<i>valeur culturelle (positive)</i> ← <i>recyclable</i>

*i.e. le concept du Critère fait partie de la connotation de la lexie du texte

** Exemple : une perte de réseau ou une perte financière sont des types de problème

TABLEAU 19 : Relations lexicales entre les intitulés des Critères et les lexies du corpus

Il semble donc que, comme illustré dans le TABLEAU 19, selon le Critère, on soit en présence d'un vocabulaire discriminant plutôt stable lorsque ce Critère appelle des lexies qui sont reliées à son intitulé par des relations de dérivation, synonymie ou, potentiellement, antonymie, et très variable d'une étude à l'autre lorsque ces relations sont hyponymiques ou connotatives. Pour la première catégorie, l'absence de n-grams

¹⁰⁹ Par exemple, des fonctions spécifiques pourraient être *enregistrement de mon parcours*, *enregistrement du nombre de pas*, *enregistrement des entrées et sorties*, *téléchargement de mon rapport mensuel de progression*, *téléchargement des données clients*, ou *téléchargement de mes photos*.

¹¹⁰ Comme précédemment expliqué (Partie I, Chapitre I, point 4.2.4, §3, p. 29), la présence du mot *onde* peut parfois renvoyer à une crainte de leur impact sur la santé par les participants aux études, et leur omniprésence peut donc être un point qui entre en contradiction avec leur système de valeurs.

reliés par une relation d'antonymie aux lexies exprimées dans les intitulés des Critères nous interpelle. En effet, cela met *a priori* en évidence des « carences » particulières dans les données. Par exemple, le fait que, pour le Critère 2.1 – « *Le concept est-il simple à utiliser ?* » nous identifions de nombreuses lexies relatives à la simplicité et aucune pour la complexité paraît pour le moins étrange ; l'explication la plus évidente est que le corpus n'est pas assez représentatif pour ce Critère car il ne contient pas assez de verbatim du type « *ça a l'air complexe d'utilisation* ». Cependant, une seconde explication possible est aussi que, par gêne, les utilisateurs ont tendance à ne pas avouer qu'ils trouvent le Concept présenté complexe, ou tout du moins pas de manière si explicite, auquel cas le corpus serait alors, au contraire, particulièrement représentatif de la réalité des contenus textuels collectés dans le cadre des études. Une autre hypothèse serait que les verbatim qui contiennent des antonymes des mots présents dans les intitulés des Critères seraient plus difficiles à repérer pour l'analyste et susciteraient un accord inter-juge moindre – qui les aurait éliminés de notre corpus –, mais nous ne pensons pas que cela soit le cas car la complexité du processus d'interprétation pour associer « *c'est facile à utiliser* » et « *c'est difficile à utiliser* » au Critère 2.1 est identique dans les deux cas. Une observation des fréquences et répartitions de */difficile/*, */compliqué/*, */compliquer/* et */complexe/* dans les données non filtrées révèle que ces lexies sont bien présentes dans le corpus mais davantage réparties entre différents Critères que */simple/* et */facile/*. Quelle qu'en soit la raison, donc, et sur la base du corpus dont nous disposons, il semble qu'en l'état les antonymes soient moins discriminants que les synonymes.

Enfin, nous constatons des épiphénomènes, consignés dans les colonnes *Signaux émergents* des TABLEAUX 15 à 18 qui peuvent potentiellement servir d'indices de discrimination, et dont certains paraissent assez intuitifs et d'autres plus surprenants et potentiellement accidentels et/ou particulièrement ambigus. Pour les premiers, on peut par exemple citer la désignation indirecte des fonctions du Concept par des syntagmes tels que */tout DET partie/* pour le Critère 1.4 – « *L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?* » ; on imagine bien des verbatim comme « *je trouve toute la partie des alertes en cas de pannes très intéressante* » ou « *toute cette partie, là, je n'y vois aucun intérêt* », la *partie* pouvant faire référence à un ensemble de fonctions du Concept. Pour les seconds, on remarque que */pour DET instant/* est surreprésenté pour le Critère 1.1 – « *L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?* ». On peut supposer que cet adverbe est employé dans le cadre de comparaison avec

l'existant (« *pour l'instant, on fait ça à la main* »). Or, la comparaison avec l'existant est attendue pour d'autres Critères, et il semble donc que l'utilisation de cet indice pourrait entraîner des erreurs. Enfin, la surreprésentation de *pour* dans le Critère 3.1 – « *L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?* », relatif au sentiment d'être concerné ou non par le Concept, est intéressante mais problématique car cet indice est très ambigu ; */pouvoir être intéressant pour/* peut en effet permettre de repérer « *ça peut être intéressant pour des mères de famille* », mais une telle recherche donnerait également lieu à un bruit considérable dû à des verbatim tels que « *ça peut être intéressant pour gagner du temps* ». Le syntagme *pour moi non* est d'ailleurs emblématique de ce problème, dû à la potentielle nature adverbiale de *pour moi* lorsqu'il s'agit d'une expression non-compositionnelle (« *c'est peut-être intéressant pour les mères de famille, mais pour moi non* » – 'cela ne m'intéresse pas' – vs « *je ne sais pas ce qu'en ont pensé les autres mais pour moi non, ce n'est pas joli* » – 'à mon sens ce n'est pas joli' –).

3.2.3. Analyse par catégories morphosyntaxiques

Comme on peut le constater sur les FIGURES EN ANNEXE 8 et 9 (p. 291), les lexies qui semblent les plus discriminantes sont les lexies verbales, suivies des nominales puis des adjectivales et adverbiales dans les mêmes proportions. Plus précisément, davantage de verbes que de noms semblent discriminants, mais les seconds donnent lieu à sensiblement plus d'occurrences que les premiers.

Parmi les 17 Critères pour lesquels nous avons pu extraire des n-grams spécifiques, 10 entrent dans la catégorie de ceux pour lesquels on compte davantage de verbes que d'autres catégories spécifiques. On retrouve par exemple le 2.4 – « *Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?* » avec des verbes comme *faire gagner, supporter, apporter, changer* ou *perdre*. Trois Critères comportent davantage de noms discriminants, comme le 3.4 – « *Le concept suscite-t-il un imaginaire d'appropriation ?* » pour lequel on retrouve des noms d'animaux. Deux Critères comportent autant de noms que de verbes spécifiques. Les deux derniers ont pour catégories les plus représentées *V+N+Adj* (2.1) et *N+Conj* (3.3). Cependant, les quatre derniers Critères que nous venons d'évoquer ne sont porteurs que d'un nombre très restreint de lexies spécifiques, et cette répartition n'a donc que peu de signifiante. Nous observons ainsi que les adjectifs, pourtant *a priori* importants dans un corpus d'opinions,

semblent avoir un potentiel discriminant relativement faible pour la tâche qui nous importe comparé à d'autres catégories morphosyntaxiques.

Notons que l'on trouve peu de constructions à verbe support, mais cela est possiblement dû à la manière dont nous avons extrait les n-grams ; les adverbes ne sont pas neutralisés par l'extraction, si bien que *avoir envie*, *avoir souvent envie* et *avoir très envie*, s'ils faisaient partie du corpus, n'auraient pas été identifiés comme plusieurs occurrences d'*avoir envie*. La fréquence de la lexie aurait donc été basse, ce qui lui aurait potentiellement empêché de passer les seuils que nous avons fixés. Notons également que pour le Critère 4.5 – « *Quel prix l'utilisateur consent-il à payer pour ce concept ?* », l'étiquette NUM, que nous n'avons pas prise en compte, aurait été une information discriminante importante.

3.3. Conclusions

L'analyse textométrique que nous avons réalisée permet de mettre en évidence de fortes disparités entre les Critères concernant la présence d'indices discriminants dans les contenus textuels qu'ils appellent. Ces disparités sont dues en premier lieu aux déséquilibres de représentation des Critères dans le corpus qui provoquent, pour les Critères sous-représentés, une impossibilité d'extraction d'informations pertinentes. En second lieu, elles sont aussi dues au fait que certains Critères appellent des actualisations lexicales des concepts mobilisés par synonymie et dérivation qui donnent lieu à des vocabulaires qui sont assez stables et potentiellement finis (par exemple, PRIX → *coût*, *somme*, *tarif*, *coûter*, *valoir*, etc.). D'autres Critères, en revanche, appellent, par hyponymie et relation connotative, des vocabulaires très étendus et variables d'une étude à l'autre (par exemple, MÉTIER → *cardiologue*, *électricien*, etc.). Pour ce deuxième type d'actualisation lexicale, une approche par statistique lexicale basée sur apprentissage sur un corpus de taille aussi restreinte que le nôtre, et sans mobilisation de ressources sémantique et/ou d'apprentissage/modélisation *ad hoc* de relations lexicales, paraît peu prometteur. Or, nous estimons qu'au moins sept¹¹¹ des 20 Critères nécessitent au moins en partie l'usage d'un vocabulaire variable et étendu.

D'autre part, les données que nous avons pu observer dans cette analyse nous permettent de conclure – toujours avec prudence étant donné la taille du corpus – que les

¹¹¹ 1.4, 2.2, 2.3, 2.4, 3.3, 4.2 et 4.4.

catégories morphosyntaxiques qui semblent les plus discriminantes sont les verbes puis, dans une moindre mesure, les noms.

Cette exploration textométrique nous permet donc de mieux caractériser nos données mais elle ne nous fournit pas assez d'indices discriminants pour construire l'ébauche de règles de classification ; cet ensemble de règles serait relativement pauvre, très déséquilibré et ne reposerait que sur des quantités d'observations restreintes difficilement généralisables. La mise à l'épreuve de systèmes de classification sur des bases statistiques plus complexes que de la simple fréquence d'occurrences sera l'objet des deux prochains points de ce chapitre.

4. Les méthodes de classification de texte

La classification de texte est une des tâches majeures du TAL dont les applications sont aujourd'hui très nombreuses. La – relativement récente – revue de littérature de [Mirończuk & Protasiewicz \(2018\)](#) met en évidence de manière assez représentative toute cette variété. Les auteurs y indiquent que certains travaux exploitent la classification de texte dans un seul domaine en particulier. Les domaines représentés pour cette catégorie de travaux sont l'industrie, la finance, la médecine, les données issues d'internet¹¹² et les bases de données de brevets. Ils inventorient également les différents objectifs donnés par les chercheurs dont, par exemple, l'identification de sites web commerciaux ou non commerciaux, d'articles portant sur le terrorisme ou encore de la détection d'ironie¹¹³. L'objectif largement surreprésenté est l'analyse des sentiments. Un autre champ d'application fréquemment cité dans la littérature est la détection de *spams* (par exemple, [Alom et al. \(2020\)](#) ou encore [Xia & Chen \(2020\)](#)). La tâche de classification est inhérente au sujet de cette thèse, et nous nous situons donc dans le domaine de l'analyse qualitative et avons pour objectif la détection d'opinions caractérisées par une grille d'analyse psychosociologique spécifique.

Plusieurs caractéristiques de nos données d'apprentissage (taille restreinte, classes nombreuses et déséquilibre) semblent prédire des résultats peu satisfaisants par des méthodes stochastiques de classification. De plus, de telles méthodes, de par leur aspect « boîte noire », vont à l'encontre du désir de transparence de l'outil évoqué dans le

¹¹² Emails, logs de serveurs, pages web, tweets.

¹¹³ Respectivement : [Qi & Davison \(2009\)](#), [Choi et al. \(2014\)](#) et [Charalampakis et al. \(2016\)](#).

Chapitre III. Cependant, au regard des performances de tels systèmes, nous nous devons de les tester afin soit de les éliminer en cas de résultats non satisfaisants soit d'envisager leur place dans un système plus complet dans le cas où ils s'avéreraient être prometteurs. En préambule à ces tests, nous faisons dans ce point un bref état des lieux de ces méthodes en exposant tout d'abord les principes généraux, puis différentes approches, et enfin quelques repères quant aux résultats qu'elles permettent d'obtenir et des facteurs qui influent sur ces résultats.

4.1. Principes généraux

L'objectif général de la classification de texte est d'attribuer des classes prédéterminées à des textes entiers ou des segments plus courts. Par simplification, on appelle *documents* les unités textuelles à classer quelle que soit leur taille. La classification peut être binaire (classifier revient à choisir entre deux classes) ou multi-classe (plusieurs classes par document sont possibles), monolabel (une seule classe par document) ou multilabel (plusieurs classes pour le même document), et hiérarchique (les classes sont organisées hiérarchiquement) ou non.

La procédure générale consiste à : 1) prétraiter les données, 2) les transformer en une représentation adaptée à un algorithme qui manipule des données chiffrées, 3) entraîner le modèle, et 4) l'évaluer. L'étape 1) consiste à tokeniser le texte, en retirer les mots fréquents dans la langue – mots outils comme *est, donc, que* ou *alors* –, et appliquer une lemmatisation ou racinisation. La 2) consiste à transformer les données en une représentation adaptée à un algorithme de classification. Deux stratégies sont possibles : la plus commune est la représentation dans un espace vectoriel, où les unités à classer sont représentées par des vecteurs, et la seconde, bien plus rare, est la représentation en graphes, dans laquelle les mots sont les nœuds du graphe (Mirończuk & Protasiewicz, 2018). Différents types de descripteurs peuvent être exploités pour cela (par exemple, des mots simples, des informations issues d'ontologies ou encore des éléments sémantiques simples comme des entités nommées) (*Ibid.*). Il est ensuite usage de réduire la dimension des espaces vectoriels afin de contrer ce qui est désigné comme *fléau de la dimension* (*curse of dimensionality*) (Verleysen & François, 2005). On procède par la suite à l'entraînement – 3) – qui consiste à fournir les données transformées à un algorithme de classification afin que ce dernier apprenne à les classer, construisant ainsi un modèle. Ce modèle est ensuite appliqué sur des données de test afin d'en évaluer les performances.

Il est courant pour cela d'utiliser la méthode dite de validation croisée (*cross-validation*) qui consiste à segmenter en plusieurs partitions les données et à entraîner itérativement l'algorithme sur toutes les partitions sauf une qui sert de test, en changeant à chaque itération la partition qui servira pour ce test¹¹⁴. La forme la plus poussée est appelée *leave one out* et consiste, à chaque itération, à n'utiliser qu'une seule entrée comme test, en entraînant donc le modèle sur toutes les autres données¹¹⁵.

4.2. Approches

Au-delà des principes généraux que nous venons de décrire, la classification de texte fait appel à une variété d'approches, d'architectures et d'algorithmes. Dans ce point, nous n'entendons pas faire l'inventaire de toutes ces possibilités, mais mettons en avant les distinctions qui seront utiles pour la suite de ce chapitre. Nous commencerons ainsi par introduire les différentes méthodes de vectorisation possibles, puis nous présenterons plusieurs types d'algorithmes de classification.

4.2.1. Sacs de mots, TF-IDF et embedding

Il existe plusieurs manières de vectoriser les unités textuelles à classifier. Le choix des caractéristiques à exploiter (mots, éléments sémantiques, etc.) est un premier paramètre, mais il existe aussi des alternatives dans la manière-même de représenter ces caractéristiques. Parmi ces choix possibles, trois approches, en particulier, sont possibles : les sacs de mots, le *Term-Frequency – Inverse Document Frequency* (TF-IDF) et l'*embedding*.

La première est une méthode de représentation très courante. Elle consiste à construire un vocabulaire qui contient tous les mots présents dans l'intégralité du corpus, et à représenter chaque unité textuelle par un vecteur dans lequel la fréquence de chaque mot est indiquée, en créant ainsi un histogramme des mots du texte (Qader et al., 2019). De manière très simple, imaginons que l'on dispose d'un corpus de deux documents : « la tarte n'a pas un goût de tarte » et « la tarte a un goût de poire ». Le dictionnaire comporterait neuf mots : [1:la 2:tarte 3:n' 4:a 5:pas 6:un 7:goût 8:de 9:poire]. Chaque

¹¹⁴ Par exemple, on segmente un corpus d'entraînement de 1 000 documents en quatre ensembles de 250 documents A, B, C et D. Lors de la première itération, on entraînera le modèle sur A, B et C, et on le testera sur D. Lors de la seconde itération, on entraînera le modèle sur B, C, D et on le testera sur A, et ainsi de suite jusqu'à ce que chaque sous-ensemble ait servi de corpus de test.

¹¹⁵ Par exemple, sur un ensemble de 1 000 documents, on en utilisera à chaque itération 999 pour l'entraînement et 1 pour le test, jusqu'à ce que chaque document ait servi pour le test.

document serait donc représenté par un vecteur de neuf éléments, dont la valeur correspond au nombre d'occurrences du mot. Ainsi, on aurait pour la première phrase $[1\ 2\ 1\ 1\ 1\ 1\ 1\ 0]$, et pour la deuxième $[1\ 1\ 0\ 1\ 0\ 1\ 1\ 1\ 1]$ ¹¹⁶. Cette représentation repose sur la simple fréquence d'occurrences et ne tient compte ni de la syntaxe ni de la sémantique (synonymie, polysémie, proximité sémantique). Il en découle notamment le fait que « la pomme a un goût de poire » et « la poire a un goût de pomme » seront représentés par le même vecteur. Toutefois, si l'ordre linéaire des mots est perdu, il est possible de tenir compte de phénomènes de cooccurrence en considérant comme unités non pas des mots mais des n-grams¹¹⁷. D'autre part, des méthodes existent pour introduire une dimension sémantique dans les sacs de mots, comme la méthode des sacs de concepts, dans laquelle, donc, les caractéristiques exploitées ne sont plus lexicales mais conceptuelles (Mouriño-García et al., 2016).

Le TF-IDF (Spärck Jones, 1972) repose sur le même principe général que l'approche par sacs de mots¹¹⁸ – la première partie du nom, *Term Frequency*, est un autre nom de l'approche par sacs de mots –, combiné à celui de *Inverse Document Frequency*, et vise à minimiser l'impact des mots courants dans le corpus en assignant des poids aux mots (Kowsari et al., 2019). Plus précisément, il s'agit de repérer les mots à haute fréquence dans un document (TF), tout en les pondérant selon le principe que, si ces mots sont courants dans l'ensemble du corpus, ils n'ont que peu d'intérêt pour la classification, et qu'il faut donc augmenter le poids des mots rares et diminuer celui des mots fréquents dans le corpus (IDF). La première partie fait donc augmenter le rappel, tandis que la deuxième améliore la précision (*Ibid.*). Les manques cités précédemment en ce qui concerne les sacs de mots – l'absence d'informations syntaxiques et sémantiques – persistent dans cette approche.

Les approches par *embedding* tiennent quant à elles compte de phénomènes syntaxiques et sémantiques. Leur principe central est que chaque mot est représenté par un vecteur, et que deux mots qui présentent des similitudes (*pomme* et *poire* par exemple), seront proches dans l'espace vectoriel. Cette technique repose sur l'hypothèse

¹¹⁶ Cet exemple simple est inspiré par celui fourni par Soumya George & Joseph (2014).

¹¹⁷ Si l'on utilise des bigrams, les entrées de notre vocabulaire, pour l'exemple donné, serait ainsi $[1:la\ tarte, 2:tarte\ n', 3:n'\ a, 4:a\ pas, 5:pas\ un, 6:un\ goût, 7:goût\ de, 8:de\ tarte, 9:tarte\ a, 10:a\ un, 11:de\ poire]$.

¹¹⁸ La première partie de l'acronyme, *Term Frequency*, est d'ailleurs un autre nom de l'approche par sacs de mots.

Modèle	Avantages	Limites
Sacs de mots	<ul style="list-style-type: none"> - Facile à calculer - Facile de calculer la similarité entre deux documents - Métrique de base pour extraire les termes les plus descriptifs dans un document - Fonctionne avec des mots inconnus (par exemple, les nouveaux mots dans les langues) 	<ul style="list-style-type: none"> - Ne permet pas de saisir la position dans le texte (syntaxique) - Ne permet pas de saisir le sens du texte (sémantique) - Les mots courants ont un effet sur les résultats (par exemple, "suis", "est", etc.)
TF-IDF	<ul style="list-style-type: none"> - Facile à calculer - Facile de calculer la similarité entre deux documents - Métrique de base pour extraire les termes les plus descriptifs dans un document - Les mots courants n'affectent pas les résultats grâce à l'IDF (par exemple, "suis", "est", etc.) 	<ul style="list-style-type: none"> - Ne permet pas de saisir la position dans le texte (syntaxique) - Ne permet pas de saisir le sens du texte (sémantique)
Word2Vec	<ul style="list-style-type: none"> - Saisit la position des mots dans le texte (syntaxique) - Saisit le sens des mots (sémantique) 	<ul style="list-style-type: none"> - Ne permet pas de saisir le sens du mot à partir du texte (ne parvient pas à saisir la polysémie) - Ne permet pas de saisir des mots du corpus qui ne sont pas dans le vocabulaire du modèle
Représentations de mots conceptualisés	<ul style="list-style-type: none"> - Saisit le sens du mot à partir du texte (prend en compte le contexte, traite la polysémie) 	<ul style="list-style-type: none"> - Consommation de mémoire pour le stockage - Améliore les performances notamment sur les tâches en aval. Le calcul est plus coûteux par rapport à d'autres techniques - Nécessite un autre <i>embedding</i> de mots pour tous les LSTM et les couches de <i>feed forward</i> - Ne permet pas de saisir les mots hors vocabulaire du corpus - Ne fonctionne qu'au niveau de la phrase et du document (ne fonctionne pas au niveau des mots individuels)

TABLEAU 20 : Avantages et inconvénients de différentes approches de vectorisation –
Reproduction partielle et traduction du tableau de Kowsari et al. (2019, pp. 48-49)

distributionnelle (Harris, 1954) qui postule que les mots qui apparaissent dans des contextes similaires ont tendance à partager des significations proches. Elle est notamment implémentée grâce aux modèles *Word2Vec* (Mikolov et al., 2013), dont l'extension *Doc2Vec* permet de vectoriser non pas uniquement des mots mais des documents entiers. Ces approches donnent des résultats très satisfaisants – de l'ordre de 90% – notamment en analyse des sentiments (Rudkowsky et al., 2018; Zharmagambetov & Pak, 2015). Mais, selon la nature de la tâche et le domaine d'application, elles ne

supplémentent pas nécessairement les approches par sacs de mots (Shao et al., 2018; Yogarajan et al., 2020). D'autre part, si ces approches permettent d'incorporer des informations sémantiques absentes des deux autres techniques précédemment citées, elles ne permettent pas d'affronter l'importante question de la polysémie. Autrement dit, *mousse* aura la même représentation vectorielle qu'il s'agisse d'une plante, d'un aliment ou d'un jeune matelot. La solution pour pallier ce problème et aller vers une précision sémantique beaucoup plus importante consiste à prendre en compte le contexte dans lequel se trouvent les mots afin de produire des représentations distinctes pour leurs différentes acceptions. Naissent ainsi les *Représentations de Mots Contextualisés* (*Contextualized Word Representation*), à la base des modèles de langues de type *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin et al., 2018) qui ont, depuis 2018, permis d'obtenir des résultats remarquables dans plusieurs domaines du TAL dont la classification.

Chacune de ces techniques a ses avantages et ses inconvénients, très bien résumés dans un tableau de Kowsari et al. (2019, pp.48-49) que nous reproduisons partiellement¹¹⁹ et traduisons dans le TABLEAU 20.

4.2.2. Types d'algorithmes de classification

Nous présentons ici les types d'algorithmes de classification les plus courants. Précisons que ces algorithmes ne sont pas spécifiques à la classification de texte (ils peuvent par exemple faire du *clustering* ou encore de la classification d'images), mais nous les aborderons à travers ce prisme.

Premièrement, les classifieurs bayésiens naïfs se basent sur le théorème de Bayes (Pearson, 1925) et reposent sur le principe selon lequel l'existence d'une caractéristique – dans le cas qui nous intéresse, généralement, un mot – pour un document appartenant à une classe n'est pas dépendante de celle des autres caractéristiques (Lewis (1998) cité par Adeva et al. (2014)). Relativement simples de déploiement et rapides d'exécution, ces classifieurs rivalisent néanmoins avec d'autres algorithmes plus complexes et sont souvent considérés comme une base de référence pour l'apprentissage supervisé (Xu, 2018).

¹¹⁹ D'autres modèles d'*embedding* que le *Word2Vec* sont également présentés dans le tableau original.

Les arbres de décision ([Morgan & Sonquist, 1963](#)) fonctionnent, comme leur nom l'indique, sur un principe de décision en arborescence ; à partir de la racine, une succession de décisions sont prises par séparation séquentielle des données à chaque nœud de l'arbre, conduisant ainsi à une classification finale qui est une des feuilles ([F.-J. Yang, 2019](#)).

Les classifieurs de type Machine à Vecteur Support (*Support Vector Machine* (SVM)) ([Boser et al., 1992](#)) cherchent, dans les grandes lignes, à identifier, dans un espace vectoriel multidimensionnel dans lequel chaque unité à classifier est un point, l'hyperplan qui séparera au mieux les catégories. Ils sont souvent considérés comme les plus performants par comparaison aux deux premiers types d'algorithmes que nous venons de citer, les rendant ainsi très populaires ([Pranckevičius & Marcinkevičius, 2017](#)).

Les méthodes de type *k plus proches voisins* (*k-nearest neighbors* (KNN)) fonctionnent selon le principe qu'un document est classé dans la même catégorie que les documents qui lui sont le plus similaires (i.e. qui comportent le plus de mots clefs en commun), la similarité étant calculée par distance Euclidienne ([Yang & Chute \(1994\)](#) cités par [Adeva et al. \(2014\)](#)). Là où d'autres méthodes demandent des adaptations pour produire un système multi-classe, les KNN le sont par nature ([Patel & Srivastava \(2014\)](#) et [Sahgal & Ramesh \(2002\)](#), cités par [Kowsari et al. \(2019\)](#)).

La régression logistique est basée sur l'idée de prédire la probabilité de la valeur d'une variable booléenne en fonction de données d'entrées. Moins fréquemment citée et utilisée, elle permet pourtant parfois des performances très satisfaisantes comparativement à d'autres classifieurs ([Pranckevičius & Marcinkevičius, 2017](#)).

Enfin, les réseaux de neurones artificiels reproduisent le principe du fonctionnement cérébral humain ; chaque cellule traite une information à partir d'une donnée d'entrée et transmet une information de sortie à d'autres cellules qui font de même jusqu'à aboutir à une décision. On parle de *deep learning* lorsque ces neurones sont nombreux, organisés en plusieurs couches reliées entre elles avec un certain niveau de complexité, le tout créant un système qui demande une puissance de calcul considérable et qui est capable d'extraire les caractéristiques utiles à la classification ([Abiodun et al., 2018](#)).

4.3. Résultats et facteurs d'influence sur la performance

Les résultats atteints ces dernières années pour les tâches de classification de texte sont bien sûr très variables en fonction de la nature des problématiques abordées, et il est donc difficile d'établir, de façon très générique, ce qu'est un *bon* score sans tenir compte des particularités des données et visées applicatives des différents travaux. Cependant, quelques campagnes d'évaluation récentes peuvent fournir un éclairage sur ce point. La campagne *MEX-A3T* de 2020 (Aragón et al., 2020), pour l'espagnol mexicain, proposait par exemple une tâche de détection de fausses informations sur un corpus de presse (i.e. classification fausse/vraie information) et une autre de détection d'agressivité dans des *tweets* (i.e. classification en agressif/non agressif). Les meilleurs scores obtenus (f-score macro) sont de 0,85 pour les deux tâches¹²⁰. Le *Hate Speech and Offensive Content Identification in Indo-European Languages* (HASOC (2020) (Mandl et al., 2020), dans le cadre du *Forum for Information Retrieval Evaluation* (FIRE 2020)), proposait deux tâches de détection de contenus haineux dans des *tweets* en anglais, hindi et allemand. La première tâche, une classification binaire (haineux vs non offensant), a donné lieu à des performances (f-score macro) de – respectivement pour chaque langue – 0,51, 0,53 et 0,52¹²¹. La deuxième, qui utilisait trois classes (haineux, offensant, et grossier¹²²) n'a permis d'obtenir, dans le même ordre, que des scores de 0,26, 0,33 et 0,29¹²³. Des écarts de résultats sont donc importants, mais des scores très satisfaisants peuvent être atteints ; par exemple, une tâche de détection d'humour de *SemEval 2021* (Meaney et al., 2021) donne lieu à des résultats de l'ordre de $f=0,99$ ¹²⁴ pour la sous-tâche la plus simple (humoristique / non humoristique).

Les importants écarts de scores observés dans la littérature sont, certes, dus à la complexité inhérente à chaque tâche, mais certains facteurs ont une influence générale. La longueur des textes à classer est un premier paramètre à prendre en compte ; l'information contextuelle et la richesse lexicale des unités considérées sont bien évidemment très différentes s'il s'agit de classer un petit segment de texte ou un article entier. Avec l'essor des réseaux sociaux, et notamment de Twitter qui limite le nombre de

¹²⁰ Respectivement : Guzman-Silverio et al. (2020) et Villatoro-Tello et al. (2020).

¹²¹ Respectivement : Mishraa et al. (2020), Raja et al. (2021) et Kumar et al. (2020).

¹²² *Hate, offensive et profane* dans le texte.

¹²³ Respectivement : Ezike & Sivanesan (2020), Kumari (2021) et Sai & Sharma (2020).

¹²⁴ L'équipe gagnante n'a pas souhaité fournir de papier expliquant les détails de son approche.

caractères possibles dans un *tweet*, la dernière décennie a connu un intérêt grandissant pour la classification de textes courts (on entend généralement par *textes courts* des textes de moins de 200 mots (Sangounpao & Muenchaisri, 2019)). Ces textes sont généralement caractérisés par un certain niveau d’informel et de bruit et sont moins thématiquement orientés. Ils permettent d’obtenir des résultats moins satisfaisants – surtout avec l’approche par sacs de mots – à cause du peu d’informations contextuelles et co-occurrenceuses qu’ils contiennent (Flisar & Podgorelec, 2020) ; une possibilité pour pallier ce problème est l’enrichissement du modèle de représentations par l’ajout d’informations sémantiques, par exemple en ayant recours à des ontologies / graphes conceptuels (*Ibid.* ; J. Wang et al., 2017). Malgré le fait que l’on sache que la longueur des textes a un impact certain sur les performances des classifieurs, et malgré une littérature très riche en description de tâches de ce type, très peu d’informations sont données quant à l’impact de la longueur des textes en termes de détérioration des performances, si bien que l’on ne peut établir avec exactitude à partir de quelle taille un texte est considéré comme potentiellement problématique pour une tâche de classification (McCartney et al., 2017)¹²⁵. Le volume de données d’apprentissage a lui aussi un impact déterminant sur les performances des classifieurs, ces derniers étant plus difficilement capables de produire des généralisations à partir d’une petite quantité de données. Les méthodes *deep learning* sont celles qui requièrent le plus de données (Kowsari et al., 2019). Enfin, le nombre de classes a également son importance en termes de résultats, principalement si le corpus est déséquilibré.

A la suite de cette brève présentation de la classification de texte, nous présentons les expérimentations que nous avons menées sur nos données.

5. Tests de classifieurs

Afin de tester le potentiel des classifieurs dans le cadre des présents travaux, nous avons utilisé plusieurs approches de vectorisation et plusieurs algorithmes. Nous exposons dans ce point la méthodologie déployée, les résultats et les conclusions de cette expérimentation.

¹²⁵ Les quelques données fournies dans cette publication de McCartney et al. (2017) afin d’adresser en partie cette question tendent à suggérer que, globalement, à l’échelle d’un *tweet* faisant au maximum 140 caractères, une détérioration importante des résultats apparaît lorsque l’on passe sous le seuil des 20 caractères.

5.1. Méthodologie

Nous présentons ici les corpus, les différents prétraitements, les approches de vectorisation et algorithmes testés, la combinatoire résultante de ces modalités, l'implémentation et le calcul des résultats que nous avons utilisés.

5.1.1. Corpus

Le corpus filtré que nous avons constitué et que nous avons présenté au chapitre précédent a pour avantage de contenir des verbatim représentatifs de chaque Critère. Cependant, il comporte deux inconvénients. Premièrement, il ne rend pas compte de la nature potentiellement multilabel attendue du système final. Deuxièmement, il ne comporte aucune donnée humainement inclassifiable. Or, le système final devra, avant de classifier les verbatim, extraire dans un premier temps ceux qui sont pertinents et donc éliminer ceux qui sont inclassifiables. Nous décidons donc d'utiliser quatre corpus différents :

- *Corpus monolabel – Critères seuls* : il s'agit du corpus filtré dont nous avons présenté la volumétrie dans le [TABLEAU 11](#) (p. 97), et duquel nous avons retiré les verbatim attribués à un Niveau seulement et non à un Critère, et qui contient donc 737 verbatim avec une seule classe chacun.
- *Corpus monolabel – Critères et inclassifiables* : le même corpus auquel nous ajoutons les verbatim qui ont été majoritairement jugés comme inclassifiables lors du filtrage expert (et avaient donc été supprimés). Ces verbatim sont au nombre de 103. Nous obtenons donc un total de 840 verbatim ayant une seule classe, qui peut être un Critère ou la mention *Inclassifiable*.
- *Corpus multilabel – Critères seuls* : afin de construire un corpus multilabel, nous utilisons tous les verbatim qui avaient été évalués lors du filtrage expert et attribuons à chacun toutes les classes qui leur ont été données par au moins un analyste ; nous partons donc du principe que si au moins un analyste humain a fait cet étiquetage, il peut être considéré comme

acceptable de la part de la machine¹²⁶. Nous retirons les verbatim qui n'ont reçu que des étiquettes *Inclassifiable* et supprimons cette classe. Nous obtenons ainsi un corpus de 1 554 verbatim étiquetés en multilabel, sans la classe *Inclassifiable*, et dont chaque verbatim est attribué à au moins une classe.

- *Corpus multilabel – Critères et inclassifiables* : Nous ajoutons au corpus précédent les étiquettes *Inclassifiable* qui avaient été attribuées par les analystes et ajoutons les verbatim, précédemment retirés, qui ont été évalués unanimement comme *Inclassifiables*. Nous avons alors un corpus de 1 661 verbatim étiquetés en multilabel dont *Inclassifiable* est une des classes possibles.

En procédant de la sorte, nous avons la possibilité d'évaluer les performances de classifieurs selon des modalités qui évoluent graduellement du plus simple mais éloigné de la tâche finale (monolabel sans données non pertinentes) au plus complexe et fidèle à l'attendu (multilabel avec des données non pertinentes).

5.1.2. Pré-traitement

Nous testons plusieurs types de prétraitements en combinant trois traitements possibles des mots outils à la possibilité de lemmatiser ou non le texte. Ces trois traitements pour les mots outils sont : leur conservation, leur suppression ou leur remplacement par leur étiquette morphosyntaxique. Nous obtenons donc six configurations possibles, illustrées dans le TABLEAU 21.

Lemmatisation	Traitement des mots outils
Oui	Conservation
Oui	Remplacement par catégorie
Oui	Suppression
Non	Conservation
Non	Remplacement par catégorie
Non	Suppression

TABLEAU 21 : Combinatoire de prétraitements testés pour la classification de texte

¹²⁶ Il s'agit là, bien sûr, d'un postulat fort questionnable, et qui va à l'encontre de la valeur que nous avons attribuée jusque-là à l'accord inter-juge. Cependant, nous ne disposons d'aucun autre moyen d'obtenir des données d'apprentissage précises et volumineuses étiquetées en multilabel et sommes donc contraintes d'adopter cette stratégie, en prenant toute la mesure de ce qu'elle engendre en termes de valeur significative des résultats obtenus.

5.1.3. Méthodes de vectorisation

Nous testons des approches sacs de mots, TF-IDF et d'*embedding*. Pour les deux premières, nous utilisons les fonctions de la librairie Python *Scikit-Learn*¹²⁷ prévues à cet effet et testons comme caractéristiques des unigrams, des bigrams et une combinaison d'unigrams + bigrams. Pour l'*embedding*, nous choisissons d'utiliser le modèle *Doc2Vec* de la librairie *gensim*¹²⁸ pour entraîner un modèle sur nos propres données, et testons également l'application d'un modèle pré-entraîné inclus dans la librairie *spacy*¹²⁹. Nous avons donc quatre modalités de vectorisation possibles.

5.1.4. Algorithmes de classification

Nous choisissons d'expérimenter six algorithmes de classification assez classiques appartenant aux grandes familles que nous avons décrites dans le point 4.2.2 (p. 126) du présent chapitre. Nous excluons les approches *deep learning* en raison de la faible quantité de données dont nous disposons. Ces six algorithmes sont tous implémentés avec la librairie *Scikit-learn* et sont : un classifieur bayésien naïf gaussien, un SVM, un KNN, un *Random Forest* (arbre de décision), de la régression logistique et un perceptron multicouche (classifieur neuronal).

5.1.5. Combinatoire des approches

En résumé, nous avons donc :

- Quatre corpus qui donnent lieu à deux types de tâches (monolabel et multilabel) et deux *sets* de 20 et 21 classes
- Six configurations de prétraitement
- Huit configurations de vectorisation :
 - Sacs de mots avec unigrams, sacs de mots avec bigrams, sacs de mots avec unigrams et bigrams
 - TF-IDF avec unigrams, TF-IDF avec bigrams, TF-IDF avec unigrams et bigrams
 - *Embedding* entraîné sur corpus
 - *Embedding* avec modèle pré-entraîné
- Six algorithmes de classification

¹²⁷ <https://scikit-learn.org/stable/>

¹²⁸ <https://pypi.org/project/gensim/>

¹²⁹ <https://spacy.io/models/fr>

Ces différents paramètres donnent lieu à 1 104 configurations possibles, soit autant de modèles testés¹³⁰.

5.1.6. Implémentation et calcul des résultats

Nous testons donc chacune de ces différentes configurations en mélangeant aléatoirement les corpus et en utilisant 80% pour l'entraînement et 20% pour les tests. Nous évaluons chaque modèle par des mesures de précision, rappel et f-score (macro-moyennes pondérées). Les résultats bruts ainsi produits sont consultables dans les TABLEAUX EN ANNEXE 12 à 14 (pp. 292-299) et commentés dans le point suivant.

5.2. Résultats

		Monolabel		Multilabel	
		Score	Configuration	Score	Configuration
20 Critères	p	0,46	BOW – Random Forest – Non lemmatisé, mots outils convertis, unigrams+bigrams	1	W2V Entraîné sur corpus - SVM - (toutes combinaisons de prétraitement)
	r	0,40	TF-IDF – Régression logistique – Lemmatisé, mots outils conservés, unigrams+bigrams	0,63	W2V Pré-entraîné - Naive Bayes - Mots outils conservés
	f	0,38	TF-IDF – Régression logistique – Lemmatisé, mots outils conservés, unigrams+bigrams	0,36	TF-IDF – Régression logistique - Non lemmatisé, mots outils convertis, unigrams
20 Critères + Inclassifiables	p	0,5	TF-IDF – Perceptron multicouche – Non lemmatisé, mots outils conservés, bigrams	0,93	W2V Entraîné sur corpus - SVM - Lemmatisé, mots outils conservés
	r	0,40	TF-IDF – Régression logistique – Non lemmatisé, mots outils conservés, unigrams+bigrams	0,6	W2V Pré-entraîné - Naive Bayes - Mots outils supprimés
	f	0,39	TF-IDF – Régression logistique – Non lemmatisé, mots outils conservés, unigrams+bigrams	0,39	BOW - Régression logistique - Non lemmatisé, mots outils remplacés, unigrams

TABLEAU 22 : Meilleurs scores par corpus obtenus par ML et configurations correspondantes

Le premier constat que l'on peut dresser par simple observation des résultats est qu'aucune des configurations testées ne semble se démarquer des autres par des scores particulièrement élevés ; toutes configurations confondues, le f-score varie de 0 à 0,39¹³¹, pour des précisions allant de 0,03 à 1 et des rappels de 0 à 0,63. On note cependant que l'approche mêlant TF-IDF et Régression Logistique permet d'obtenir les meilleurs f-scores sur trois des quatre tâches que nous avons testées.

Les meilleurs scores obtenus sont consignés dans le TABLEAU 22¹³². On observe premièrement que les approches *Word2Vec* (W2V) ne permettent d'obtenir aucune des

¹³⁰ Toutes les combinaisons ne sont pas pertinentes. Par exemple, il est inutile de tester le remplacement des mots outils par leur étiquette pour appliquer un modèle d'*embedding* pré-entraîné.

¹³¹ Moyenne et médiane = 0,2, distribution gaussienne.

¹³² Nous utilisons l'acronyme anglais *BOW* (*bag of words*) pour l'approche par sacs de mots.

meilleures performances en monolabel, tandis qu'elles permettent d'obtenir les meilleurs précisions et rappels pour le multilabel. Cependant, ces scores hauts sont associés à des performances très basses sur les autres mesures ; les scores de précision de 1 et 0,93 obtenus sur le multilabel s'accompagnent, respectivement de $r=0$, $f=0$ et $r=0,01$, $f=0,01$. De manière générale, les performances les plus élevées sont obtenues par l'approche TF-IDF, suivie par les sacs de mots et le W2V pré-entraîné – qui sont relativement similaires – et enfin le W2V entraîné sur notre corpus (pour un résumé des scores moyens obtenus par approche, voir le [TABLEAU EN ANNEXE 15](#) (p. 300)).

En outre, s'agissant des algorithmes, nous observons que la Régression Logistique est celui qui permet d'atteindre les meilleurs f-score pour toutes les tâches (cf. [TABLEAU EN ANNEXE 16](#) (p. 300)). De manière tendancielle, le SVM permet d'atteindre les meilleurs scores de précision, et le Naïve Bayes donne de bons rappels pour le multilabel seulement.

D'autre part, toujours en termes d'impact moyen sur les scores, la lemmatisation et le sort réservé aux mots outils ne semblent pas jouer de rôle important ; on observe une très légère tendance à de meilleurs scores lorsque le corpus est lemmatisé (de l'ordre de +0,01 sur chaque moyenne des mesures), et la suppression et conservation des mots outils donnent des mesures moyennes strictement identiques, tandis que leur remplacement par leurs étiquettes morphosyntaxiques fait augmenter la précision de +0,1 et baisser le rappel de -0,01. Concernant la taille des n-grams pour les approches par sacs de mots et TF-IDF, une légère tendance à l'augmentation de la précision est permise par l'emploi conjoint des unigrams et bigrams, tandis que le recours aux unigrams seuls donne sensiblement de meilleurs rappels.

Afin d'observer plus en détails la distribution des résultats entre les classes, nous nous intéressons aux matrices de confusion obtenues par les configurations ayant les meilleurs f-scores, que nous présentons dans les TABLEAUX 23 et 24. Pour ce qui concerne la tâche monolabel, nous constatons principalement une tendance à la sur-classification dans les Critères 1.1, 1.4, 2.3, 2.4 et 3.1, ainsi qu'à la classe Inclassifiable pour le corpus qui la contient. Ces faibles précisions sont corrélées à une surreprésentation de ces Critères dans les données d'apprentissage¹³³. En outre, nous observons des résultats

¹³³ Ils représentent, respectivement 9%, 7%, 10%, 12% et 11% des verbatim du corpus qui ne contient pas les verbatim inclassifiables (pour 5% attendus par Critère pour un corpus équilibré).

		Prédit																				Prédit																					
		1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5	Incl.	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5	
Vrai	1.1	4	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	10	4	0	0	1	0	1	1	3	0	0	1	0	0	0	0	0	0	0	1		
	1.2	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	2	5	0	1	1	0	1	0	2	0	0	2	0	0	0	0	0	0	0	2		
	1.3	2	2	4	0	0	0	0	1	0	1	3	0	0	0	0	0	0	0	0	0	3	0	2	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
	1.4	1	0	0	3	0	0	0	2	1	1	1	1	0	0	0	0	0	0	0	0	1	2	0	4	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	
	1.5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	3	0	0	0	1	1	0	2	0	0	0	0	0	0	0	0	0	0
	2.1	0	1	0	0	0	1	0	1	3	0	2	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2.2	0	0	1	0	0	0	1	1	2	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	2	0	0	0	0	0	0	0	1	0	0	0	0	
	2.3	0	0	2	3	0	1	0	9	5	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	7	1	0	1	0	0	0	1	0	0	0	0	0	
	2.4	0	0	0	2	0	1	0	2	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	8	13	0	0	0	0	0	0	0	0	0	0	0	
	2.5	0	1	1	2	0	0	0	2	2	0	1	0	0	0	0	0	0	0	0	0	2	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	3.1	2	0	0	1	0	0	0	1	1	0	10	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0
	3.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3.3	1	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3.4	0	0	0	1	0	1	0	0	0	0	1	0	0	0	8	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	0	
	3.5	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	4.1	0	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	2	0	0	0	1	0	0	0	0	1	0	0	2	0	1	0	0	0	0	3	0	0	0	0	1	
	4.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4.4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
	4.5	0	1	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4

TABLEAU 23 : Matrices de confusion pour les meilleurs scores en monolabel

		Prédit									
		1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5
Vrai	1.1	non 216 28 oui 47 20	non 266 12 oui 28 5	non 246 13 oui 39 13	non 251 23 oui 25 12	non 270 15 oui 24 2	non 276 3 oui 26 6	non 243 13 oui 49 6	non 207 33 oui 45 26	non 225 28 oui 41 17	non 255 13 oui 35 8
	1.2	non 258 11 oui 35 7	non 305 1 oui 5 0	non 288 1 oui 20 2	non 290 0 oui 18 3	non 292 1 oui 18 0	non 284 1 oui 21 5	non 295 0 oui 16 0	non 298 3 oui 10 0	non 259 14 oui 36 2	non 287 1 oui 17 6
	1.3	non 231 41 oui 37 23	non 284 22 oui 22 4	non 254 19 oui 44 15	non 255 24 oui 35 18	non 294 15 oui 19 4	non 297 14 oui 14 7	non 254 39 oui 30 9	non 205 39 oui 55 33	non 244 24 oui 39 25	non 274 24 oui 26 8
	1.4	non 275 15 oui 25 17	non 321 4 oui 7 0	non 295 13 oui 20 4	non 308 4 oui 15 5	non 316 6 oui 10 0	non 303 7 oui 13 9	non 316 7 oui 9 0	non 321 4 oui 7 0	non 289 11 oui 28 4	non 309 4 oui 10 9
	1.5	non 104 60 oui 56 112	non 294 15 oui 19 4	non 294 15 oui 19 4	non 294 15 oui 19 4	non 294 15 oui 19 4	non 297 14 oui 14 7	non 254 39 oui 30 9	non 205 39 oui 55 33	non 244 24 oui 39 25	non 274 24 oui 26 8
	2.1	non 275 15 oui 25 17	non 321 4 oui 7 0	non 295 13 oui 20 4	non 308 4 oui 15 5	non 316 6 oui 10 0	non 303 7 oui 13 9	non 316 7 oui 9 0	non 321 4 oui 7 0	non 289 11 oui 28 4	non 309 4 oui 10 9
	2.2	non 275 15 oui 25 17	non 321 4 oui 7 0	non 295 13 oui 20 4	non 308 4 oui 15 5	non 316 6 oui 10 0	non 303 7 oui 13 9	non 316 7 oui 9 0	non 321 4 oui 7 0	non 289 11 oui 28 4	non 309 4 oui 10 9
	2.3	non 275 15 oui 25 17	non 321 4 oui 7 0	non 295 13 oui 20 4	non 308 4 oui 15 5	non 316 6 oui 10 0	non 303 7 oui 13 9	non 316 7 oui 9 0	non 321 4 oui 7 0	non 289 11 oui 28 4	non 309 4 oui 10 9
	2.4	non 275 15 oui 25 17	non 321 4 oui 7 0	non 295 13 oui 20 4	non 308 4 oui 15 5	non 316 6 oui 10 0	non 303 7 oui 13 9	non 316 7 oui 9 0	non 321 4 oui 7 0	non 289 11 oui 28 4	non 309 4 oui 10 9
	2.5	non 275 15 oui 25 17	non 321 4 oui 7 0	non 295 13 oui 20 4	non 308 4 oui 15 5	non 316 6 oui 10 0	non 303 7 oui 13 9	non 316 7 oui 9 0	non 321 4 oui 7 0	non 289 11 oui 28 4	non 309 4 oui 10 9
Incl.	non 104 60 oui 56 112	non 294 15 oui 19 4	non 294 15 oui 19 4	non 294 15 oui 19 4	non 294 15 oui 19 4	non 297 14 oui 14 7	non 254 39 oui 30 9	non 205 39 oui 55 33	non 244 24 oui 39 25	non 274 24 oui 26 8	

TABLEAU 24 : Matrices de confusion pour le meilleur score en multilabel

plutôt satisfaisants, en termes de précision, pour les Critères 3.4, 4.1 et 4.5, dont les distributions dans le corpus sont moyennes¹³⁴. Si nous observons les zones dans lesquelles il semble exister le plus de confusion, nous remarquons que ces zones sont assez similaires à celles que nous avons identifiées lors de l'évaluation de l'accord inter-juge par les analystes (cf TABLEAU EN ANNEXE 7, p. 274) ; on retrouve principalement des confusions à l'intérieur des Niveaux 1 et 2, entre ces mêmes Niveaux, et relativement peu en intra et inter Niveaux pour les 3 et 4. Il est intéressant de noter cette similarité entre les difficultés à discriminer rencontrées par l'humain et la machine, mais notre corpus de test étant relativement réduit, cette constatation mériterait d'être confirmée avec un volume de données plus important.

D'autre part, nous avons souhaité mettre en parallèle les résultats de ces tests avec les observations tirées de l'analyse textométrique relatives à la cohésion lexicale que nous avons présentées plus haut. Les FIGURES EN ANNEXE 10 et 11 (p. 301) croisent les f-scores obtenus par classes avec la distribution de ces classes dans le corpus d'apprentissage et la catégorie de vocabulaire attendue sur la base de nos précédentes observations (i.e. stable¹³⁵, variable¹³⁶, stable + variable et incohérents/trop peu de données pour l'estimer). Il en résulte que la corrélation entre vocabulaire stable et f-score élevé n'est pas parfaite, mais l'on observe tout de même des scores tendanciuellement plus hauts pour les Critères pour lesquels nous estimions que le vocabulaire était relativement stable. Un contre-exemple notable est le cas du Critère 4.4, pour lequel nous avons estimé que le vocabulaire était composé d'une partie stable et d'une partie plus variable, et qui obtient le second meilleur f-score pour les deux tâches malgré une faible représentation dans le corpus. Cependant, le nombre de verbatim attribués à ce Critère dans le corpus de test est très faible et cette mesure n'est certainement pas significative. De plus, hormis la classe *Inclassifiable*, qui est un cas particulier, nous observons que les Critères 3.3 et 4.2, pour lesquels nous avons identifié des vocabulaires variables, ont des f-scores de 0. Dans le cas du 4.2 pour le corpus à 21 classes, cela est dû au fait qu'aucun verbatim n'était attribué à ce Critère dans les données de tests, et qu'un seul l'était pour le corpus à 20 classes. Ce n'est pas le cas du 3.3, pour lequel quelques occurrences étaient présentes, et qui donne

¹³⁴ Ils représentent respectivement 6%, 4% et 6% des verbatim du même corpus.

¹³⁵ Relations de synonymie, potentiellement antonymie, et dérivation avec les concepts mobilisés par les Critères.

¹³⁶ Relation d'hyponymie et de connotation avec les concepts mobilisés par les Critères.

lieu à des mesures $f=0$. Cependant, le nombre d'occurrences dans les données de test reste très bas, et cette tendance à l'abaissement du f-score en cas de vocabulaire variable serait à mettre à l'épreuve d'un corpus plus important.

Nous souhaitons observer ce qu'il en est avec une méthode *Word2Vec*, dans la mesure où l'*embedding* est par nature sensible aux similarités sémantiques. Nous produisons donc un graphique similaire aux deux précédents (FIGURE EN ANNEXE 12, p. 302) à partir de la matrice de confusion obtenue avec la configuration *Word2Vec* la plus performante. Il apparaît que la corrélation est cette fois beaucoup plus franche, les cinq meilleurs f-scores étant obtenus sur des Critères dont nous avons estimé que le vocabulaire attendu serait stable. Cela semble, d'une part, confirmer en partie nos déductions au regard des listes de n-grams extraites lors de l'analyse textométrique. D'autre part, cela explique également dans une certaine mesure le fait que les approches *Word2Vec* ne soient pas les plus performantes pour la tâche qui nous préoccupe de par le fait que, précisément, les contenus attendus par les Critères ne sont souvent pas caractérisés par des ensembles de lexies sémantiquement proches. Il demeure cependant possible qu'avec un volume de données plus conséquent, une meilleure cohésion sémantique aurait été obtenue et que les approches par *embedding* permettent des résultats bien meilleurs.

Pour ce qui concerne la classification multilabel, nous constatons que les scores par Critères sont similaires à la tâche monolabel, à quelques exceptions près comme le faible rappel obtenu pour le Critère 4.5 dans le cas du corpus à 20 classes. De manière générale, les scores obtenus en multilabel sont similaires à ceux obtenus en monolabel. Nous notons également que peu de Classes sont attribuées à chaque verbatim (1,14 classes/verbatim pour le corpus à 20 classes et 2,11 classes/verbatim pour le corpus à 21 classes). Cela est légèrement en-dessous de ce que comporte notre corpus d'apprentissage (2,15 classes/verbatim), mais correspond certainement davantage à la réalité de la pratique du codage¹³⁷.

¹³⁷ Pour rappel, notre corpus multilabel a été constitué sur la base des différentes classes attribuées par plusieurs annotateurs lors de la phase de filtrage du corpus. Dans la réalité de la pratique, un seul analyste code généralement les données, et le nombre de verbatim auxquels sont attribués plus d'un Critère sont moins nombreux.

5.3. Conclusions

En conclusion de cette expérimentation menée sur des approches ML de classification, nous pouvons tout d'abord souligner le fait que de telles approches ne permettent pas, sur la base des données dont nous disposons, d'atteindre des performances satisfaisantes en vue de la création d'un système d'autocodage pour la méthode CAUTIC®. Comme il était possible de l'entrevoir, la faible taille de notre corpus et son déséquilibre ont participé à engendrer des performances peu satisfaisantes, mais aussi des difficultés à appuyer certaines de nos conclusions de par le risque de surgénéralisation de phénomènes accidentels. D'autre part, les tâches comportant des données inclassifiables nous ont démontré leur capacité somme toute assez faible à discriminer ce qui est pertinent dans un texte de ce qui ne l'est pas (des f-scores de 0,38 et 0,66 sur la classe *Inclassifiable* sont obtenus par les meilleures configurations). Ce fait n'est absolument pas négligeable dans la mesure où la distinction entre données pertinentes et non pertinentes est la première étape incontournable du codage ; il apparaît dès lors qu'un système basé sur de la seule classification par ML, tout du moins tel qu'entraîné sur notre corpus, ne saurait accomplir cette tâche avec précision, ce qui engendrerait un nombre considérable de bruits et silences par la suite. De façon plus anecdotique, nous n'observons pas de dégradation graduelle des performances selon un continuum de degré de complexité de la tâche ; la plus complexe (multilabel sur 21 classes) obtient une meilleure mesure de f-score que la plus simple (monolabel sur 20 classes).

Le meilleur modèle semble être celui qui utilise une vectorisation en TF-IDF et un algorithme de régression logistique. Il est possible que les autres algorithmes testés permettent des performances supérieures à celles constatées durant cette expérimentation si l'on essaie différents réglages d'hyperparamètres, mais les présents résultats ne nous encouragent pas à poursuivre dans cette voie. D'autre part, nous avons observé que les approches par *embedding* n'étaient pas particulièrement performantes en raison, probablement, de l'instabilité sémantique dans les classes. Des modèles plus récents prenant en compte le contexte – de type BERT –, pourraient peut-être permettre des résultats meilleurs. Cependant, en complément de la présente expérimentation, nous

avons réalisé un test avec le modèle *CamemBERT*¹³⁸ (L. Martin et al., 2019) – avec nos six algorithmes et quatre corpus – et les résultats n’ont pas dépassé ceux que nous avons obtenus avec les autres approches¹³⁹. L’exploration des possibilités de ce type de modèles pourrait bien sûr être approfondie (*finetuning* du modèle, apprentissage sur nos données ou utilisation alternative de *FlauBERT*¹⁴⁰ (Le et al., 2019)). Mais, ces premiers essais ne sont pas encourageants, et nous réservons cette possibilité à un futur dans lequel nous disposerons de davantage de données.

Dans ce chapitre, nous avons exploré les possibilités offertes par des méthodes statistiques, dans une démarche descriptive avec l’analyse textométrique, et à visée applicative avec la classification par ML. Nous avons démontré que les limites pré-identifiées qui résident dans la taille et le déséquilibre du corpus ainsi que dans la complexité sémantique de la tâche ont un impact très fort sur la pertinence du recours à l’emploi de méthodes statistiques. Nous avons également mis en évidence que le déséquilibre du corpus n’est pas seulement d’ordre quantitatif entre les différentes classes, mais est également inhérent à la nature-même des contenus textuels, à travers une forte variabilité de présence de lexies discriminantes parmi les classes. Les « carences en identité lexicale » que nous avons ainsi observées pour certains Critères rendent difficile la caractérisation des contenus textuels qui leur appartiennent sur un axe purement paradigmatique. Pour cette même raison, ces seules informations ne semblent pas suffisantes pour aboutir à un système satisfaisant. Dans le prochain chapitre, nous restons dans une logique d’approche ascendante et tentons de caractériser et exploiter des informations non plus seulement lexico-sémantiques mais aussi syntaxiques.

¹³⁸ <https://camembert-model.fr/> (consulté le 15/10/21).

¹³⁹ Les meilleurs résultats obtenus avec *CamemBERT* sont, en termes de f-scores :

- 0,36 en monolabel sur 20 classes avec l’algorithme de régression logistique
- 0,36 en monolabel sur 21 classes avec l’algorithme de régression logistique
- 0,33 en multilabel sur 20 classes avec l’algorithme Naive Bayes
- 0,39 en multilabel sur 21 classes avec l’algorithme Naive Bayes

¹⁴⁰ <https://github.com/getalp/Flaubert> (consulté le 15/10/21).

CHAPITRE VI : EXPÉRIMENTATION SUR LE RECOURS À UNE RESSOURCE SYNTAXICO-SÉMANTIQUE

Nous souhaitons explorer l'usage d'une ressource sémantique pour accomplir la tâche qui nous est confiée. En effet, nous avons fait face au constat des faibles performances des outils statistiques. D'autre part, même s'ils s'étaient démontrés efficaces, ces outils n'auraient pas su répondre au besoin de transparence attendu pour la solution que nous entendons construire. Nous formulons donc à présent l'hypothèse suivante : l'utilisation d'une ressource lexico-syntaxico-sémantique pourrait permettre une classification sur des bases d'informations plus riches que de la simple statistique lexicale tout en permettant un certain degré d'explicitation des choix d'un potentiel algorithme qui utiliserait ces informations. Nous considérons l'utilisation d'une telle ressource comme un moyen d'enrichir le corpus à partir duquel un algorithme de classification travaillerait. A ce titre, ces travaux se situent bien dans une perspective d'approche ascendante.

Dans le cadre de l'expérimentation que nous allons présenter, nous cherchons par conséquent à :

- a) Enrichir notre corpus par des informations sémantiques issues de ressources préconstruites.
- b) Evaluer le potentiel discriminant de ces informations.
- c) Evaluer, par là-même, le degré de possibilité d'utiliser ces informations comme entrées au système voulu d'autocodage.

Nous continuons donc de chercher des réponses aux interrogations transverses (cf. Partie I, Chapitre III, point 9.2, p. 84) 1) b. et c. relatives à la complexité sémantique de la tâche et à l'utilisation d'un petit corpus étiqueté. Nous abordons également la question 2) b. qui porte sur le besoin de transparence de l'algorithme, dans la mesure où, si une telle approche s'avérait fructueuse, les informations utilisées pour la classification seraient traduisibles en des termes humainement compréhensibles.

Afin d'exposer au lecteur les explorations que nous avons conduites dans cette optique, nous commençons par présenter les ressources existantes pour le français, puis

nous nous arrêtons sur celle que nous avons choisie, Verbønet (Danlos et al., 2016), et présentons ensuite l'expérimentation réalisée avec cette ressource.

6. Les ressources sémantiques pour le français

Parmi les diverses ressources existantes pour le français, nous choisissons de nous concentrer sur celles qui permettent de traiter des questions relatives aux verbes. En effet, nous les avons identifiés, à l'occasion de l'analyse textométrique présentée au chapitre précédent, comme étant la catégorie morphosyntaxique qui avait le plus important potentiel discriminant. Cela n'est d'ailleurs pas surprenant car il est bien connu que les verbes ont un rôle privilégié dans l'accès au sens, en eux-mêmes (Healy & Miller, 1970), mais également à travers leurs structures argumentatives (Bencini & Goldberg, 2000). Nous proposons donc dans ce point un petit inventaire de telles ressources, puis nous abordons la question de leur utilisation dans des tâches d'annotation et de classification.

6.1. Ressources pour le français

Nous exposons dans ce point des ressources qui se concentrent sur le verbe et dont le téléchargement et l'exploitation expérimentale sont accessibles¹⁴¹.

Le **Lexique-grammaire** de Gross (1975) donne naissance, dans les années 1970-1990, à des bases de données informatisées, ce qui constitue une nouveauté d'intérêt pour le TAL (Lapalme & Le Pesant, 2020). Cette ressource est principalement à visée syntaxique et se présente sous forme de tables. Elle encode pour chaque verbe des informations concernant la structure actancielle, les différentes constructions possibles, les restrictions de sélection et les auxiliaires. Le Lexique-grammaire comporte aujourd'hui environ 14 000 entrées¹⁴².

¹⁴¹ D'autres ressources potentiellement riches sont développées par des laboratoires ou entreprises privées qui se réservent le droit exclusif à leurs accès et exploitation (Hadouche & Lapalme, 2010) ; pour des raisons évidentes, nous ne nous y intéresserons pas.

¹⁴² Il est téléchargeable dans sa version augmentée et optimisée pour le TAL (Tolone, 2009, 2011; Tolone et al., 2010) à l'adresse :

<http://infolingu.univ-mlv.fr/DonneesLinguistiques/Lexiques-Grammaires/Telechargement.html> (consulté le 23/10/21).

On doit au travail colossal de Jean Dubois et Françoise Dubois-Charlier la construction de la ressource **Les Verbes Français (LVF)**¹⁴³ (Dubois & Dubois-Charlier, 1997). Cette dernière n'a pas été originellement pensée pour servir au TAL, et a mis un temps certain à être transformée pour servir ce but. Basée sur l'idée d'un lien entre constructions syntaxiques et interprétation sémantique (Levin, 1993), elle propose une organisation des verbes en classes sémantiques corrélées à des structures syntaxiques et organisées hiérarchiquement en trois niveaux (François et al., 2007). Elle fournit un ensemble d'informations syntaxiques et sémantiques très complet sur les quelques 12 000 verbes qu'elle contient : transitivité, forme prépositionnelle ou non, restrictions de sélection, nature des compléments, domaine (par exemple sociologie ou commerce) et propriétés sémantiques (« étiquettes interprétatives »)¹⁴⁴ (Hadouche & Lapalme, 2010).

Le **Dicovalence**¹⁴⁵ de van den Eynde & Mertens (2003) est principalement orienté sur une description syntaxique et a pour originalité d'être basé sur l'approche pronominale (van der Eynde & Blanche-Benveniste, 1978). A ce titre, les descriptions des cadres de valences qu'il propose reposent sur le principe selon lequel les positions actanciennes sont définies par les pronoms qui peuvent les occuper. Sont fournies comme informations dans cette ressource : les cadres de sous-catégorisation qui incluent les fonctions syntaxiques, les réalisations syntagmatiques et les restrictions de sélection, les auxiliaires et la passivation (Hadouche & Lapalme, 2010; Mertens, 2010). Il contient 8 000 entrées pour 3 700 verbes.

FrameNet (Baker et al., 1998) est une ressource pour l'anglais basée sur la sémantique des cadres (*frame semantics*) de Fillmore (1982) qui a été à l'origine du projet. Une version française¹⁴⁶, objet central de la thèse de doctorat de Djemaa (2017), a été réalisée dans le cadre du projet ASFALDA¹⁴⁷. La notion de *frame*¹⁴⁸, qui structure cette ressource, renvoie à « une structure conceptuelle décrivant un type particulier de situation, objet ou événement ainsi que ses participants, parties ou phrases » (Djemaa,

¹⁴³ Aujourd'hui téléchargeable sous plusieurs formats et consultable depuis une interface récemment développée (Lapalme & Le Pesant, 2020); http://rali.iro.umontreal.ca/LVF_DEM/LVF-DEM-req.html?dic=LVF (consulté le 23/10/21).

¹⁴⁴ Par exemple, 'obtenir quelque chose de quelqu'un pour soi-même' pour une acception du verbe *acheter* (exemple de Hadouche & Lapalme (2010)).

¹⁴⁵ <https://www.ortolang.fr/market/lexicons/dicovalence/v1> (consulté le 23/10/21).

¹⁴⁶ <http://asfalda.linguist.univ-paris-diderot.fr/frameIndex.xml> (consulté le 23/10/21).

¹⁴⁷ <https://sites.google.com/site/anrasfalda/> (consulté le 23/10/21).

¹⁴⁸ Nous préférons ici ne pas traduire par *cadre*, par souci de réduction d'ambiguïté.

2017, p.33). Ces derniers sont les éléments du frame (*frame elements*), dont certains ont des rôles *noyaux* (*core*) et d'autres des rôles secondaires (*non-core*). La version française comporte à ce jour environ 100 *frames* et 1 100 entrées lexicales.

6.2. Utilisation de ressources verbales pour l'annotation et la classification

L'utilisation de ressources sémantiques est une possibilité pour des tâches d'annotation/détection et classification ; nous en présentons ici quelques exemples.

Pour de la résolution d'anaphores pronominales sur un corpus de transcriptions d'appels de la RATP, Godbert (2014) utilise les restrictions de sélection fournies par le LVF. Pour le domaine d'application que nous avons présenté dans le second chapitre de cette thèse, l'analyse de données textuelles pour les Sciences Humaines et Sociales, Silberztein (2020) propose d'utiliser cette même ressource, ainsi que le Dictionnaire Electronique des Mots¹⁴⁹, afin de dépasser la possibilité d'analyse par simples formes graphiques que proposent les logiciels¹⁵⁰ communément utilisés. Les tables du Lexique-Grammaire et le Dicovalence demeurent, quant à eux, très peu utilisés dans des travaux à visée applicative (i.e. non théorique ou à visée de constitution de corpus).

Concernant plus spécifiquement l'étiquetage en rôles sémantiques, Béchet et al. (2017) utilisent FrameNet pour étiqueter un corpus encyclopédique historique français dans une démarche visant à l'enrichissement de bases de connaissance pour des moteurs de recherche. La couverture relativement faible de cette ressource à l'heure actuelle tend néanmoins à limiter son utilisation ; Dahm (2021), travaillant sur l'allemand et le français pour lesquels on dispose de ressources FrameNet¹⁵¹, explique préférer utiliser la version anglaise en ayant recours à une traduction, précisément pour ce manque de complétude des ressources francophones et germanophones. De manière générale, si l'annotation en rôles sémantiques permise par de telles ressources a été utilisée pour de nombreuses applications en anglais – Pradet (2015, p. 7) en fournit une liste d'exemples –, elle reste à ce jour peu utilisée pour le français. Un exemple notable est celui de Mekaoui et al. (2020) qui utilisent Verbānet (Danlos et al., 2016) – que nous allons présenter par la suite – pour créer un système d'annotation en rôles sémantiques pour de l'extraction de tâches dans des mails et obtiennent un f-score de l'ordre de 0,90. De manière générale, l'annotation

¹⁴⁹ Autre ressource majeure produite par Jean Dubois et Françoise Dubois-Charlier.

¹⁵⁰ Logiciels de type CAQDAS tels que nous les avons présentés, notamment, dans le Chapitre II.

¹⁵¹ Pour l'allemand, voir <https://gsw.phil.hhu.de/> (consulté le 14/11/21).

en rôles sémantiques pour le français, qu'elle utilise des ressources sémantiques ou un apprentissage sur corpus annoté, peine à égaliser les performances atteintes sur l'anglais (Daza & Frank, 2020) dû au peu de données disponibles.

Suite à cette présentation de quelques ressources pour le français et de leur utilisation, nous nous attardons maintenant sur celle que nous avons choisi d'utiliser dans le cadre des présents travaux de thèse.

7. Présentation de Verbønet

Nous avons retenu la ressource Verbønet (Danlos et al., 2016) comme étant la plus appropriée à notre expérimentation. Ce choix est fondé sur plusieurs critères. Nous avons tout d'abord exclu le Lexique-grammaire, le LVF et le Dicovalence qui ne fournissent pas de description des rôles sémantiques. Dans Framenet, cette notion de rôle sémantique est centrale mais, dans sa version française, la couverture lexicale est assez limitée, et nous estimons que le nombre¹⁵² et la granularité des rôles étaient trop importants par rapport à l'attendu et à la taille de notre corpus¹⁵³. De plus, et surtout, le formalisme de la ressource ne permet pas d'annoter automatiquement et facilement du texte de par l'absence de patrons syntaxiques. Ces exigences sont en revanche satisfaites dans VerbNet et son adaptation en français sur laquelle notre choix s'est porté et que nous présentons dans ce point.

7.1. VerbNet

Verbnet (Kripper Schuler, 2005) se veut être une ressource généraliste à large couverture. Elle se base sur la théorie de Levin (1993) dont le postulat est qu'il existe un lien étroit entre le sens et la valence syntaxique, en ce que des verbes qui partagent une même valence sont sémantiquement proches. Cette ressource est donc pleinement de double nature syntaxique et sémantique.

¹⁵² « La théorie des *frame semantics* [Fillmore, 1982] est une modification de la théorie des cas, et a abouti à FrameNet où chaque situation ou *frame* dispose de ses propres rôles sémantiques, ce qui a créé plusieurs centaines de rôles. Ces rôles partageant parfois le même nom, mais ce sont les liens entre les rôles qu'il faut exploiter pour faire des généralisations [Litkowski, 2014] » (Pradet, 2015, p.31) (la référence à Fillmore (1982) est celle que nous avons incluse dans notre bibliographie, celle à Litkowski (2014) renvoie à : Litkowski, K. (2014). *The FrameNet Frame Element Taxonomy*).

¹⁵³ Le but étant d'enrichir le corpus d'informations discriminantes, nous craignons que la configuration résultant d'un corpus de petite taille, d'une liste de lemmes verbaux limitée et de rôles sémantiques variés ne crée un éparpillement de l'information qu'il aurait par conséquent été difficile d'utiliser pour identifier des constantes au sein des classes. Cette supposition est peut-être erronée, et un test avec FrameNet pourrait être envisagé à la suite de ces travaux de thèse.

Les verbes y sont regroupés en classes hiérarchiques basées sur celles de Levin (1993) et complétées par celles de Korhonen & Briscoe (2004), les sous-classes héritant des propriétés de leur classe parente. Pour chaque classe, elle fournit les *frames* possibles, ainsi qu'une liste de verbes, les rôles thématiques des arguments accompagnés de restrictions de sélection, et des exemples. La ressource compte environ 8 000 verbes pour 270 classes qui utilisent 23 rôles thématiques.

7.2. Verbānet

Verbānet (Danlos et al., 2016)¹⁵⁴ est une ressource sémantique pour le français, construite à partir de VerbNet. Sa création a nécessité une traduction des verbes de la ressource anglaise – par croisement des entrées de LVF, du Lexique-grammaire et de dictionnaires bilingues –, ainsi qu'une réorganisation des classes à partir des *frames* adaptés pour le français. A notre connaissance, elle demeure peu exploitée à ce jour ; nous ne recensons que deux travaux¹⁵⁵ (Mekaoui et al., 2020; Pradet, 2015) dans lesquels elle est utilisée pour réaliser une annotation en rôles sémantiques.

La ressource se présente sous forme de fichiers XML. Nous utiliserons l'exemple de la classe *beg-58.2* pour présenter sa structure. Premièrement, comme illustré dans la FIGURE 8, chaque fichier liste les membres de la classe (ici : *conjur*, *implorer*, *presser*, *prier*, *requérir* et *supplier*). Puis, nous trouvons une description des rôles thématiques pour lesquels les restrictions de sélection sont indiquées. Ces rôles thématiques, organisés hiérarchiquement, suivent ceux définis par Bonial et al. (2011).

```
<MEMBERS>
  <MEMBER name="conjur" />
  <MEMBER name="implorer" />
  <MEMBER name="presser" />
  <MEMBER name="prier" />
  <MEMBER name="requérir" />
  <MEMBER name="supplier" />
</MEMBERS>
```

```
<THEMROLES>
  <THEMROLE type="Agent">
    <SELRESTRS logic="or">
      <SELRESTR Value="+" type="organization"/>
      <SELRESTR Value="+" type="animate"/>
    </SELRESTRS>
  </THEMROLE>
  <THEMROLE type="Recipient">
    <SELRESTRS logic="or">
      <SELRESTR Value="+" type="organization"/>
      <SELRESTR Value="+" type="animate"/>
    </SELRESTRS>
  </THEMROLE>
  <THEMROLE type="Topic"/>
</THEMROLES>
```

FIGURE 8 : Membres et rôles thématiques de la classe *beg-58.2* de Verbānet

¹⁵⁴ La description de la ressource que nous présentons dans ce point est basée sur cette publication et, sauf indication contraire, les différentes précisions que nous donnons à son propos dans ce chapitre en sont issues (ainsi que sur l'observation des fichiers téléchargés).

¹⁵⁵ Dont le plus récent n'avait pas été publié au moment où nous avons réalisé l'expérimentation que nous présentons ici.

Puis, sont donnés les *frames*, dont un exemple est présenté dans la FIGURE 9. Deux remarques peuvent être formulées à ce propos. Tout d'abord, notons que de nombreux cas d'alternance syntaxique sont traités, mais l'alternance passive ne l'est jamais. Deuxièmement, on trouve autant de *frames* que de réalisations syntaxiques possibles des rôles syntaxiques. Autrement dit, une classe qui peut avoir pour complément d'objet un groupe nominal ou un groupe infinitif comportera les deux *frames* correspondant à ces deux possibilités. Cela est uniquement vrai dans le cas des compléments ; les sujets sont seulement codés dans leur réalisation nominale, même si des formes phrastiques ou infinitivales sont en réalité possibles. D'autre part, la possibilité qu'un complément soit omis n'est pas encodé dans la ressource ; pour le verbe *parler* on trouvera le *frame* correspondant à « Susan a parlé à Marie du problème. » mais pas celui correspondant à « Susan a parlé à Marie. ». Les possibilités de différents ordres de compléments ne sont pas non plus encodées ; « Susan a parlé du problème à Marie. » ne donne pas lieu à un *frame* différent de celui de « Susan a parlé à Marie du problème. ». Sur ces deux derniers points, les auteurs formulent le conseil, pour un éventuel système de TAL utilisant Verbænet, de considérer que toutes les sous-structures et toutes les permutations sont possibles.

```

<FRAMES>
  <FRAME>
    <DESCRIPTION primary="NP V NP de V-inf" syntax="Agent V Recipient {de} Topic<VRecipient-inf>"/>
    <EXAMPLES>
      <EXAMPLE>Luc a prié Léa de faire la vaisselle.</EXAMPLE>
    </EXAMPLES>
    <SYNTAX>
      <NP value="Agent">
        <SYNRESTRS/>
      </NP>
      <VERB/>
      <NP value="Recipient">
        <SYNRESTRS/>
      </NP>
      <PREP>
        <SELRESTRS>
          <SELRESTR Value="de"/>
        </SELRESTRS>
      </PREP>
      <VINF emptysubjectrole="Recipient" value="Topic"/>
    </SYNTAX>
    <SEMANTICS>supplicate(during(E), Agent, Recipient, Topic)</SEMANTICS>
  </FRAME>
</FRAMES>

```

FIGURE 9 : Exemple de frame de la classe *beg*

Pour chacun de ces *frames*, nous trouvons une description des champs surfaciques (pour *parler* : *NP V NP de V-inf - Agent V Recipient {de} Topic<VRecipient-inf>*). Quelques particularités sont à noter dans cette partie. Premièrement, le sujet est encodé *IL* lorsqu'il s'agit d'une construction impersonnelle (exemple a.). Les clitiques apparaissent parfois et

sont notés *LUI* (qui a pour réalisations lexicales possibles *me, te, lui, nous, vous, leur* et *se*) ; cette indication est uniquement présente pour indiquer un cas d'alternance particulier qui implique la clitisation du possesseur d'une partie du corps (exemple b.). Les restrictions de lemme pour les prépositions sont fournies (exemple c.). Les prépositions locatives ne sont pas lemmatisées (exemple d.) mais il est précisé, lorsque c'est le cas, qu'elles introduisent une destination (exemple e.).

- a. *IL V* (« **Il** pleut. »)
- b. *Agent LUI+Paritent V Attribute* (« Luc **lui** a agrippé le bras. », alternance syntaxique pour « Luc a agrippé **le bras de Max.** »).
- c. *Stimulus V Experiencer {par/avec} Attribute<+genitive (Stimulus)* (« Le clown amuse les enfants **par/avec** ses grimaces. »)
- d. *Theme V {{+loc}} Location* (« Des abeilles grouillent **dans** le jardin / **sous** l'arbre / **derrière** la maison / etc. »)
- e. *Theme V {de} Initial_Location {{Loc-Dest}} Destination* (« Luc est venu de Gap **à** Dax. »)

Nous trouvons ensuite un exemple (« Luc a prié Léa de faire la vaisselle. »), puis un développement du champ syntaxique. Enfin, pour le champ sémantique, nous trouvons une formule qui indique sous forme de prédicats sémantiques les actions et états des participants selon les différents moments du procès décrit par le *frame* (<SEMANTICS>*supplicate(during(E), Agent, Recipient, Topic)*</SEMANTICS>).

8. Expérimentation

L'objectif de l'expérimentation est de projeter la ressource sémantique Verbønet sur notre corpus, et ce dans deux optiques. La première est d'observer s'il existe des corrélations d'intérêt entre les classes sémantiques de Verbønet et les classes que constituent les Critères CAUTIC®. La seconde est d'obtenir un corpus annoté sémantiquement sur lequel réaliser un apprentissage. Dans ce point, nous présentons la méthodologie utilisée, les résultats et les conclusions de cette expérimentation.

8.1. Méthodologie

Nous n'utilisons dans cette expérimentation que le corpus filtré dont nous avons décrit le processus de constitution dans le Chapitre IV (i.e. celui qui contient seulement les verbatim ayant passé le test de l'accord inter-juge) ; les trois autres corpus présentés

dans le Chapitre V n'ayant été construits que pour les besoins de l'expérimentation qui y était décrite, le corpus filtré redevient notre unique corpus de référence. Nous décrivons ici le principe général de l'annotation que nous avons réalisée, puis nous détaillons l'étiquetage préalable du corpus, la projection de la ressource Verbønet sur ce dernier, ainsi que le procédé d'extraction et d'analyse des informations sémantiques ainsi obtenues.

8.1.1. Principes généraux

Pradet (2015, pp. 34-35) rappelle que l'annotation en rôles sémantiques peut reposer sur des approches supervisées, semi-supervisées, non supervisées et fondées sur la connaissance (i.e., à partir de la ressource sémantique seulement, sans corpus d'apprentissage annoté). Il formule une proposition méthodologique pour annoter un corpus à partir de VerbNet (pp. 67-83) qui s'inscrit dans le cadre de la dernière de ces approches. Notre méthodologie, relativement simple, s'y inscrit également et consiste à réaliser un étiquetage morphosyntaxique du corpus, sur lequel nous superposons l'étiquetage en rôles sémantiques à partir de la correspondance entre les *frames* de Verbønet et les catégories morphosyntaxiques qui forment le contexte du verbe dans le corpus. Les deux travaux concernant Verbønet précédemment cités, ceux de Pradet (2015) et Mekaoui et al. (2020), utilisent tous deux une approche fondée sur la connaissance¹⁵⁶, mais les modalités d'étiquetage du corpus et d'identification des actants qui en découlent sont différents ; le premier a recours à un repérage linéaire des arguments, tandis que les seconds utilisent une annotation en dépendances. C'est la première de ces approches que nous utiliserons. Nous cherchons, par conséquent, premièrement, à effectuer un étiquetage en constituants puis à chercher, pour chaque verbe présent dans le corpus, s'il est présent dans Verbønet et si son entourage correspond à un des *frames* qui lui est associé. Si tel est le cas, nous étiquetons sa classe et les rôles sémantiques des actants qui l'accompagnent.

¹⁵⁶ Cela est en partie vrai seulement concernant les travaux de Mekaoui et al. (2020), car ils complètent cette approche par une seconde qui ne mobilise pas de ressource. D'autre part, la méthode proposée par Pradet (2015) a en réalité été appliquée avec la version anglaise VerbNet (et inspirée des précédents travaux de Swier & Stevenson (2004, 2005)) mais s'inscrit dans l'optique plus général de travaux qui visent à une annotation en rôles thématiques pour le français dans laquelle Verbønet est centrale.

8.1.2. Étiquetage préalable du corpus

Notre corpus est au format XML. Un premier étiquetage morphosyntaxique de base est réalisé avec *TreeTagger* (Schmid, 1994) auquel nous combinons des informations morphologiques à partir de la ressource Lexique (New et al., 2001). Nous réalisons ensuite un étiquetage en couches successives dont l'objectif est de parvenir à une segmentation en constituants qui soit adaptée aux patrons de Verbønet (i.e. d'ajouter dans le corpus des balises de type <NP>¹⁵⁷, <ADJ>, etc. qui correspondent à celles renseignées dans la ressource). Nous ne détaillons pas ces différentes couches mais présentons ci-dessous les groupes qu'elles permettent d'identifier.

Les **groupes verbaux** sont repérés par des patrons qui permettent d'identifier les constructions pronominales et/ou composées et/ou négatives (i.e., les adverbes de négation sont inclus dans le groupe). Nous obtenons donc par exemple un étiquetage du groupe verbal *ne se sont pas connectées* dans la phrase « *elles [ne se sont pas connectées] au bon moment* » qui a pour tête le lemme *connecter*. Nous traitons indistinctement les verbes intrinsèquement et accidentellement pronominaux. Nous réalisons également un étiquetage qui prend en compte les verbes auxiliaires, selon certaines conditions. En effet, Verbønet n'encode aucune construction verbale polylexicale, y compris les constructions à verbe auxiliaire, mais contient des verbes qui peuvent être auxiliaires. Nous partons d'une liste de tels verbes donnée par L'Homme et al. (2019, pp. 16-18), qui répertorie des constructions *V Vinf* (*aller Vinf, devoir Vinf*), *V à/de Vinf* (*continuer à Vinf, finir de Vinf*) et d'autres structures spécifiques impliquant des groupes prépositionnels, des adjectifs et des noms (*avoir pour effet de Vinf, être le premier à Vinf, être sur le point de Vinf, susceptible de Vinf*, etc.). Nous éliminons ceux dans lesquels le contenu sémantique de l'auxiliaire doit *a priori* être identifié séparément du verbe dans le contexte qui nous intéresse ; c'est par exemple le cas de *rendre facile/difficile de Vinf, être capable/incapable de Vinf* ou *empêcher de Vinf*, car les notions de 'facilité', de 'capacité' ou d'"empêchement (de faire quelque chose)' sont des éléments d'intérêt qui doivent être identifiés en tant que tels dans le cadre de l'analyse CAUTIC®¹⁵⁸. De cette liste, nous ne conservons que les auxiliaires qui

¹⁵⁷ NP renvoie à *Noun Phrase* (groupe nominal).

¹⁵⁸ i.e. *cela rend facile de manœuvrer, je suis incapable de manœuvrer* et *cela va m'empêcher de manœuvrer* sont des syntagmes qui sont fortement susceptibles d'être classés dans des Critères différents.

ne sont pas dans Verbønet ou bien qui y sont présents mais dont aucun des *frames* possibles ne correspond à une construction de verbe auxiliaire.

Les **groupes nominaux** simples sont repérés, lors d'une première couche d'étiquetage, par des patrons syntaxiques (*[DET ADJ NOM ADV ADJ]*, *[DET ADV ADJ NOM]*, etc.). Sont étiquetés comme groupes nominaux les pronoms *je, tu, il, ce, on, elle, nous, vous, ça, cela* et *ceci*, afin de pouvoir être repérés comme sujets selon les *frames* de Verbønet. Une deuxième couche permet d'étiqueter des groupes nominaux plus complexes, sur la base de ceux précédemment identifiés (*[ni NP, ni NP de VINF]*, *[NP, NP, etc.]*, *[NP de VINF NP]*, etc.). Afin de distinguer les cas où *de NP* doit être étiqueté comme un groupe nominal (*il met [de la musique]* ou *on va récupérer [de la place]*) des cas où la préposition *de* fait partie du schéma de complémentation du verbe (*il s'approche [de] [la sortie]*, *il diminue [d'] [un cran]*), nous utilisons LVF¹⁵⁹ ; un potentiel groupe nominal *de NP* est identifié comme tel seulement s'il ne suit pas un verbe répertorié par cette ressource comme ayant potentiellement un complément introduit par *de*¹⁶⁰. Enfin, afin de couvrir les cas de relativisation du sujet – que Verbønet n'encode pas – nous étiquetons comme groupe nominal les constructions *NP qui* qui précèdent un verbe ; on obtiendra ainsi le groupe *il peut y avoir [des trucs qui] tombent dedans avec* pour tête le lemme *truc*.

Les **locutions adverbiales, prépositionnelles et conjonctives** sont repérées et étiquetées grâce au Dictionnaire Electronique des Mots Composés (DELAC) (Silberztein, 1990). Cette ressource est très riche et permet d'identifier de nombreuses expressions polylexicales telles que *plus ou moins, sous réserve de, à bras le corps*, etc. Elle comporte néanmoins des expressions susceptibles de créer du bruit, que nous supprimons au fur et à mesure que nous les rencontrons dans le corpus. Par exemple, il est davantage probable de rencontrer des occurrences compositionnelles de *si on veut* et de *sur ce* dans notre corpus, et nous les supprimons donc. De la même manière, lorsque nous rencontrons une expression dans notre corpus qui n'est pas présente dans la ressource, nous l'y ajoutons.

¹⁵⁹ Nous aurions pu utiliser les *frames* de Verbønet, mais LVF est plus complet.

¹⁶⁰ Cela signifie que nous ne pouvons identifier les cas où *de* introduit un complément attendu par le verbe mais qui suit un autre complément nominal ; *il diminue le son d'un cran* sera étiqueté *il diminue [le son d'un cran]* (avec le patron *[NP de NP]*). Cela va à l'encontre de la recherche de correspondance avec les *frames* de Verbønet. Cependant, une vérification manuelle de la centaine d'occurrences ainsi étiquetées avec le patron montre que ce choix s'avère tout de même judicieux, car il ne crée que très peu d'erreurs de ce type par rapport aux groupes nominaux correctement identifiés ainsi. D'autre part, suivant les conseils de Danlos et al. (2016), nous ferons le choix – nous y reviendrons – de considérer que chaque actant peut être omis. Donc, si le *frame* *[NP V NP de NP]* n'était pas identifié ainsi pour *il diminue le son d'un cran*, le *frame* *[NP V NP]* le serait, et *diminuer* serait alors bien étiqueté comme appartenant à la bonne classe.

Les **groupes adjectivaux** isolés (i.e. ne faisant pas partie d'un groupe nominal) sont identifiés et étiquetés à l'aide de patrons, tout comme les **groupes adverbiaux** (qui ne sont pour leur part ni compris dans un groupe nominal ni dans un groupe adjectival) non identifiés dans DELAC car non figés (*pratiquement rien, assez régulièrement, etc.*). Nous étiquetons également comme groupes adverbiaux les syntagmes composés d'une locution prépositionnelle terminant par *de* suivie d'un groupe nominal, comme dans « *Si [en plus de ces factures], pouvait montrer que [...]* », « *Si j'installe, [de part et d'autre de la paroi], je peux [...]* » ou encore « *Ils peuvent réagir [en fonction de la typologie des bâtiments].* ». Ce choix crée très peu de bruit par rapport au nombre de groupes correctement identifiés.

Les **propositions subjonctives ou indicatives introduites par *que*** sont repérées par patrons et regroupées sous la même étiquette (<PSUBJIND>) ; Verbānet fait la distinction entre les deux modes, cependant, la désambiguïsation étant difficile dans de nombreux cas, nous choisissons de les regrouper en une seule catégorie.

Un exemple de l'étiquetage ainsi obtenu est présenté dans la [FIGURE EN ANNEXE 13](#) (pp. 303-304).

8.1.3. Projection de Verbānet

La projection de Verbānet consiste donc tout simplement à repérer chaque verbe, le considérer comme déclencheur potentiel s'il fait partie de la ressource, et chercher si les constituants qui l'entourent correspondent à un des *frames* possibles. Cette recherche se fait linéairement et en ignorant les ponctuations qui ne marquent pas une fin de phrase. Autrement dit, si un verbe déclencheur est présent dans le corpus dans une suite telle que *NP, V de NP* et qu'un des *frames* possibles pour ce verbe est *[NP V de NP]*, cette suite est étiquetée. Le repérage se fait donc sur la base des étiquettes morphosyntaxiques des constituants, et de leur lemme lorsque celui-ci est précisé dans Verbānet. Sur ce point, concernant les prépositions locatives – qui ne sont pas lexicalisées dans la ressource – nous utilisons celles qui sont répertoriées dans la ressource PrepLex ([Fort & Guillaume, 2007](#)) comme liste de lemmes possibles.

D'autre part, suivant les conseils de [Danlos et al. \(2016\)](#) que nous avons précédemment évoqués, nous considérons comme possibles toutes les permutations et sous-structures de chaque *frame* ; le *frame* *[NP V NP à NP]* permettra de repérer dans le corpus *[NP V NP]*, *[NP V à NP]* et *[NP V à NP NP]*. Enfin, afin de repérer les cas de

relativisation des compléments d'objet que Verbønet n'encode pas, nous faisons en sorte que notre système inclue la possibilité que chaque *NP* positionné après le verbe dans les *frames* puisse précéder le sujet et être suivi de *que* ; autrement dit, à partir d'un *frame* donné [*NP(Agent) V NP(Patient)*] nous générons comme possible *frame* alternatif [*NP(Patient) que NP(Agent) V*]¹⁶¹.

L'objectif de notre étiquetage n'étant pas d'obtenir la précision la plus élevée possible, mais uniquement d'identifier des récurrences de classes et de patrons dans notre corpus, nous ne traitons pas certaines problématiques comme l'alternance passive, qui n'est pas encodée dans Verbønet et nécessite donc d'ajouter des patrons et/ou règles spécifiques. Les ambiguïtés de classe (i.e. deux classes possibles pour le même *frame*) sont simplement traitées en annotant les deux classes possibles. Nous n'appliquons pas non plus les restrictions de sélections purement sémantiques (*animé, animal*, etc.) mais nous nous contentons des restrictions de lemmes (prépositions ou adverbes lexicalisés dans Verbønet comme *avec, si* ou *comment*) et de morphologie (pluriel).

Un extrait du corpus annoté selon ces modalités est présenté dans la [FIGURE EN ANNEXE 14](#) (p. 304-306).

8.1.4. Extraction des résultats

Sur ce corpus ainsi annoté, nous extrayons tous les verbes associés aux Critères des verbatim auxquels ils appartiennent. Pour ceux qui n'ont pas été annotés, nous renseignons s'ils étaient ou non présents dans Verbønet. Pour ceux ayant donné lieu à une annotation, nous extrayons leur classe et leur *frame* (tête de chaque groupe et rôle sémantique associé). Nous obtenons ainsi un inventaire de tous les verbes par Critère et, le cas échéant, de leur classe et *frame*. Nous pouvons donc chercher sur cette base des récurrences d'éléments potentiellement discriminants.

8.2. Résultats

Nous présentons dans ce point les résultats obtenus en termes de couverture et précision, puis ceux relatifs à la répartition des classes dans les Critères en vue d'identifier leur potentiel discriminant.

¹⁶¹ Par exemple : « Luc a associé les deux compagnies. » → « Les deux compagnies que Luc a associées. ».

8.2.1. Couverture et précision

	Occurrences	Lemmes
Frames trouvés	1337	219
Verbes sans frames :	3008	403
- dont présents dans Verbænet	1046	233
- dont absents de Verbænet	1962	170

TABLEAU 25 : Couverture de l'annotation avec Verbænet

Comme indiqué dans le TABLEAU 25, notre procédé nous permet d'annoter 1 337 occurrences verbales, pour 219 lemmes différents. 3 008 occurrences ne sont pas annotées, dont environ les deux tiers sont tout simplement des verbes absents de Verbænet. Dans cette catégorie, environ 60% de ces occurrences non annotées concernent le verbe *être*, 8% le verbe *falloir* et 5% le verbe *pouvoir*. Parmi les 27% d'occurrences restantes, non trouvons quelques verbes qui sont présents dans la ressource mais non sous leur forme pronominale utilisée dans le corpus – cela ne concerne que 28 occurrences –, des cas de différences de graphie entre le corpus et Verbænet (*abimer* vs *abîmer*) et des fautes de saisies identifiées à tort comme des verbes par *TreeTagger* (**apèrs*). Les cas que nous venons de citer éliminés, ce sont 97 verbes (256 occurrences) qui sont de réelles absences des verbes dans la ressource, dont certains appartiennent à des sociolectes techniques (*paramétrer*, *configurer*) mais une majorité sont relativement communs (*anticiper*, *décaler*, *généraliser*, *occuper*, etc.).

Afin de faire un état des lieux des raisons qui entraînent le fait que des verbes, pourtant présents dans Verbænet, ne sont pas annotés, nous observons une cinquantaine d'occurrences choisies aléatoirement. La raison la plus fréquente (20%) est que le verbe n'a pas de sujet ou pas de complément¹⁶² dans le corpus (« *c'était avant d'avoir vu le proto* », « *Je ne sais pas.* »). La seconde raison la plus courante (16%) est une erreur d'étiquetage de *TreeTagger* qui a identifié un verbe à tort – principalement des participes passés employés comme adjectifs – ou bien créé des erreurs dans l'étiquetage des actants qui ont empêché leur repérage (**« quand la machine [est [arrêtée]VERBE]VERBE »*, **« ça ne fait pas [mal]ADV »*). Enfin, à parts égales (14%), nous trouvons des cas où le complément est pronominalisé – cas non prévu par Verbænet et que nous n'avons pas traité par règles – (« *ça lui plaît bien* ») ainsi que des absences normales d'annotation dues au fait que le(s)

¹⁶² Concernant l'absence de complément, hormis les cas prévus par Verbænet, nous avons choisi d'imposer la présence d'au moins un complément pour que le *frame* soit détecté, afin de limiter l'ambiguïté.

complément(s) présent(s) dans le texte n'est/ne sont pas encodé(s) dans Verbønet (« *je ne sais pas si on peut le faire soi-même* » : aucun *frame* associé au verbe *savoir* n'a pour complément une proposition introduite par *si*). Les autres raisons sont plus anecdotiques, et l'alternance passive ne concerne que 8% des cas.

Enfin, nous souhaitons évaluer la précision des annotations qui ont été réalisées. Pour cela, nous observons une centaine de *frames* étiquetés sélectionnés au hasard et évaluons plusieurs dimensions.

Tout d'abord, nous nous intéressons à la pertinence de l'identification syntaxique des actants. Par exemple, dans « *[on]*_{Theme} *[suit]*_{VERBE}, *[on]*_{Co-Theme} *relève* » le second *on* est identifié comme étant complément du verbe *suivre*, et il s'agit donc d'une erreur. Nous obtenons sur ce point une précision de 0,70, mais, ces erreurs présentant des caractéristiques assez récurrentes, nous pourrions relativement facilement les éviter en ajoutant des règles spécifiques ; à partir de l'exemple présenté, nous pourrions imposer qu'aucune virgule ne sépare le verbe de son complément.

Ensuite, nous cherchons à savoir si la/les classe(s) attribuée(s) est/sont sémantiquement pertinente(s). Nous nous servons pour cela des exemples et prédicats sémantiques fournis dans Verbønet. Par exemple, dans *ça ne sert à rien*, la classe *fulfilling-13.4.1* est attribuée au verbe *servir*. Cette classe – qui comporte également les verbes *fournir* et *livrer* – a pour phrase d'exemple correspondant à ce *frame* « Luc fournit du lait à Max » et pour formule sémantique *has_possession(start(E), Agent, Theme) has_possession(end(E), Recipient, Theme) transfer(during(E), Theme) cause(Agent, E)*. Il ne s'agit donc pas du sens lié à l'utilité présent dans cette phrase, et nous considérons donc cette attribution de classe comme fausse. Nombreuses sont les erreurs de classes qui sont dues à la non-compositionnalité. En effet, les expressions polylexicales verbales (semi-)figées / constructions à verbe support ne sont pas encodées dans Verbønet, et le sens attribué au verbe seul n'est donc pas correct. Ce problème est notamment présent avec les verbes *avoir* et *faire* qui sont fréquents et très couramment utilisés comme verbes supports. Par exemple, les classes *own-100* et/ou *hire-13.5.3*, qui renvoient respectivement aux notions de 'possession' et d'"acquisition"/recrutement', sont associées au verbe *avoir* dans *il a l'air* ou encore *j'ai l'impression*. Nous considérons ces assignations de classe comme fausses. Au total, seules 41% des classes étiquetées permettent d'identifier correctement le sens présent dans la phrase.

L'évaluation de la pertinence des rôles thématiques assignés découle directement des deux autres évaluations ; si les actants sont identifiés syntaxiquement correctement, et si la classe du verbe est adéquate, alors les rôles thématiques sont corrects.

8.2.2. Répartition des classes

Afin d'étudier la répartition des classes dans les différents Critères, nous commençons par ignorer toutes celles annotées à partir des verbes *avoir* et *faire* afin de limiter l'imprécision de l'étiquetage ; ce choix est motivé principalement par la fréquence de leur fonction de verbe support qui, comme nous l'avons expliqué, crée beaucoup de bruit, mais également de par leur importante polysémie dont Verbænet ne rend que très peu compte. Il nous reste ainsi 931 occurrences de verbes annotés.

Il en résulte une prédominance importante des classes relatives à l'expression d'opinions et à l'interaction telles que *consider-1-29.9* et *transfer_mesg-37.1.1* qui cumulent à elles deux 21% des annotations. La présence de ces classes est due principalement à la nature du corpus et n'a pas de potentiel informatif intéressant pour la discrimination entre Critères ; il s'agit simplement d'occurrences nombreuses et prévisibles dans un tel corpus telles que *je crois que, je pense que, je suppose que, je vous dis* ou encore *ce que vous décrivez ici*. Nous les éliminons également. Nous supprimons ensuite les classes pour lesquelles un nombre très bas (<5) d'annotations ont été effectuées. Nous obtenons un total de 41 classes dont nous observons la répartition entre les différents Critères.

Cette répartition, illustrée dans les FIGURES EN ANNEXE 15 à 19 (pp. 307-311), est relativement pauvre en termes d'apports informatifs utiles à la discrimination entre Critères. Une seule classe est présente dans un seul Critère, la classe *fit-54-3* pour le Critère 4.5 – « *Quel prix l'utilisateur consent-il à payer pour le concept ?* », et cela est dû au verbe *valoir*. La classe *avoid-52* est exclusivement présente dans des Critères du Niveau 2 (relatif aux pratiques), ce qui était moins prévisible et est dû aux comparaisons du type « *ça nous évitera de nous salir les mains* ». La prédominance du Niveau 1, et notamment du Critère 1.3 – « *L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?* », de la catégorie *exist-47.7* est quant à elle plus prévisible. Ces trois constats sont les seules informations significatives que nous identifions.

D'autre part, ces résultats restent impactés de manière notable par les diverses erreurs de désambiguïisation, et notamment par la non prise en compte de la

polylexicalité. Par exemple, la classe largement majoritaire, *want-32.1-2*, comporte en réalité 33% d'occurrences de l'expression *vouloir dire*. Cela crée non seulement du bruit mais également de nombreux silences sur des prédicats *a priori* d'intérêt ; toutes les occurrences des expressions polylexicales verbales sont soit non étiquetées soit mal interprétées. A ce silence s'ajoute celui des verbes absents de la ressource. Ainsi, pour un Critère identifié précédemment comme ayant une « identité lexicale » relativement marquée et stable¹⁶³, le 2.4 – « *Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?* », nous constatons que l'utilisation de Verbānet se révèle bien insuffisante pour capturer cette identité ; les verbes *résoudre* et *faciliter* sont absents de la ressource et *poser problème* n'est étiqueté que sur la base du verbe qui est associé à des classes relatives à une dimension spatiale. En outre, certains verbes moyennement ou très polysémiques étant présents dans la ressource mais répartis entre peu de classes, nous obtenons des étiquetages presque systématiquement faux pour des verbes potentiellement d'intérêt, ce qui provoque des associations entre classes sémantiques de Verbānet et Critères aussi incorrectes que surprenantes ; le verbe *perdre*, par exemple, n'est présent que dans la classe *substance-emission-43.4*¹⁶⁴, qui est alors majoritairement associée au Critère 2.4 susmentionné.

En somme, si la précision de l'étiquetage pourrait être améliorée, il n'en demeure pas moins qu'il existe de nombreuses inadéquations entre la classification proposée par Verbānet et les signifiés d'intérêt dans le cadre de la tâche qui nous incombe. Cette constatation nous incite à ne pas pousser plus loin nos investigations qui auraient pu par exemple consister en une analyse des *frames* entiers (i.e. quels éléments lexicaux occupent quels rôles avec quelles classes dans quels Critères), et en l'utilisation de la ressource pour enrichir le texte en vue de l'utilisation d'un classifieur¹⁶⁵.

8.3. Conclusions

Cette expérimentation montre le fait que la seule ressource sémantique pour le français permettant une annotation en rôles sémantiques sans apprentissage sur corpus n'est pas adaptée à la poursuite de notre objectif en ce que les informations qu'elle permet

¹⁶³ Cf. Chapitre V, point 3.2.2 pp. 111-119.

¹⁶⁴ Par exemple « La voiture perd de l'huile ».

¹⁶⁵ Quelques tests ont d'ailleurs été menés dans ce sens et les performances obtenues étaient en dessous de celles obtenues sur le corpus non enrichi.

d'obtenir n'ont pas de pouvoir discriminant entre nos différents Critères. Une des premières raisons à cela est qu'une part importante des annotations de Verbønet renseignent sur la nature du corpus (expression d'opinions et interaction) et non sur le contenu des opinions exprimées, ce qui ne nous est d'aucun intérêt. De plus, la couverture de la ressource, en termes de nombre de verbes et d'encodage de la polysémie de ceux qu'elle contient, est probablement insuffisante aujourd'hui pour réaliser une analyse sémantique complexe ; Mekaoui et al. (2020), pour parvenir à l'exploiter pleinement, ont d'ailleurs dû l'enrichir et compléter leur système par un module qui ne l'utilisait pas. D'autre part, la précision obtenue par notre méthode n'est pas très élevée mais elle pourrait être augmentée par des ajouts de règles assez simples. L'absence de prise en compte de la possible polylexicalité des prédicats verbaux par Verbønet est en revanche un problème moins facilement résoluble – non qu'il ne soit impossible à adresser¹⁶⁶ – et qui impacte directement la précision de l'étiquetage mais aussi l'utilité-même du recours à la ressource pour une analyse sémantique telle que celle que nous souhaitons mener. Ces différents problèmes pourraient donc être traités mais les résultats constatés en termes de potentiel discriminant des annotations issues de Verbønet ne nous encouragent pas à poursuivre cette voie. Nous considérons donc que les différents défis pour lesquels nous avons envisagé cette solution (traiter un certain degré de complexité sémantique, sur un corpus de petite taille, en utilisant des informations humainement interprétables pour obtenir un outil final transparent) ne sont pas relevés.

Les divers tests que nous avons réalisés dans une logique ascendante et présentés dans cette seconde partie indiquent que cette approche, la plus fréquente en TAL, ne permet pas d'obtenir des résultats satisfaisants à partir d'un petit corpus déséquilibré, pour une vingtaine de classes cibles et pour une tâche qui demande un certain degré d'interprétation. Nous choisissons donc d'envisager une approche différente reposant sur une logique descendante et que nous présentons dans la prochaine partie de cette thèse.

¹⁶⁶ Mekaoui et al. (2020) expliquent par exemple utiliser JeuxDeMots (Lafourcade & Joubert, 2008) pour repérer les locutions verbales dans le texte. Ils n'expliquent pas précisément comment ces locutions sont ensuite reliées aux prédicats de Verbønet, mais l'on peut tout à fait imaginer l'utilisation des relations de synonymie fournies par cette ressource entre des locutions et des mots simples pour « ramener » les locutions à des verbes présents dans Verbønet, bien que cela poserait de nombreux problèmes de différences syntaxiques. Par exemple, *avoir envie* a pour synonyme possible dans JeuxDeMots le verbe *souhaiter*. On pourrait donc effectuer une recherche de *frames* sur la base du verbe déclencheur *souhaiter* pour chaque occurrence de *avoir envie* rencontrée dans le corpus. Cependant, étant donné que – à juste titre – *NP V de NP* n'est pas un des *frames* possibles du verbe *souhaiter* (**elle souhaite de chocolat*), toutes les occurrences de *avoir envie de NP* ne seraient pas étiquetées.

PARTIE III : APPROCHE

DESCENDANTE

CHAPITRE VII : MODÉLISATION EXPERTE DES CRITÈRES CAUTIC®

Note pour le lecteur : cette partie traitant de la modélisation par ontologie, nous mentionnerons souvent le mot concept. Nous rappelons donc à toutes fins utiles de clarification que nous utilisons la majuscule (Concept) lorsque nous nous référons au produit ou service innovant testé dans le cadre des études CAUTIC®, et la minuscule (concept) lorsqu'il s'agit d'une autre acception du mot.

Cette troisième partie de thèse aborde la tentative de résolution du problème qui nous est posé par une approche descendante, dans le sens où nous partons d'un modèle expert de la méthode sous forme d'ontologie pour aller vers le texte par projection de ce modèle. Les approches traditionnelles ascendantes s'étant montrées infructueuses, nous faisons l'hypothèse que la part de connaissances nécessaires à une automatisation de l'analyse peut être obtenue par l'élicitation des spécificités de la méthode plutôt que par la seule émergence de régularités dans le corpus. Nous entendons cependant éviter l'écueil d'une représentation artificielle décorrélée de la réalité des données, et veillerons donc à positionner notre regard au croisement de la modélisation théorique et de la réalité linguistique que nous pouvons observer dans notre corpus. Cette approche tend à imiter l'apprentissage humain de l'analyse qui passe par une appropriation des connaissances théoriques et méthodologiques qui viennent se heurter constamment aux données réelles afin d'être consolidées.

Les trois prochains chapitres présentent donc cette approche et les expérimentations menées dans ce cadre. Dans celui-ci, nous introduisons les modalités de réalisation d'une modélisation experte de la méthode CAUTIC®. Dans le prochain, le Chapitre VIII, nous présentons la confrontation du modèle aux données, et par là même son enrichissement par sa projection sur le corpus. Enfin, dans le Chapitre IX, nous concluons cette thèse par une proposition de système opérationnel final qui vient réconcilier les deux grandes approches, ascendante et descendante, qui ont structuré nos travaux.

Le présent chapitre expose la manière dont nous avons modélisé la méthode CAUTIC®. Cette modélisation ayant pris la forme d'une ontologie, nous commençons par définir cet objet et ses possibilités. Puis, nous décrivons le processus que nous avons suivi pour réaliser notre modélisation.

1. Les ontologies comme outils d'analyse sémantique

Les ontologies, dans le sens informatique du terme, représentent un moyen particulier d'apporter à la machine les connaissances du monde nécessaires à la réalisation de tâches qui les utilisent. Il s'agit d'un domaine de recherche et applicatif à part entière dont ces quelques pages ne sauraient rendre compte de l'étendue, de la complexité et de la multiplicité de conceptions et de modèles existants. Aussi, nous présentons dans ce point un aperçu très synthétique de ce que sont les ontologies. Nous omettons ainsi de nombreuses dimensions¹⁶⁷ et conceptions alternatives de ce que peuvent être les ontologies qui ne sont pas particulièrement pertinentes pour la suite ; c'est donc un portrait simplifié que nous dressons ici. Nous commençons pour cela par formuler quelques définitions autour de cet objet, avant de présenter les différentes façons de les construire puis de donner des exemples de leur utilisation pour l'analyse textuelle.

1.1. Définitions

Si le terme *ontologie* appartient originellement au domaine de la philosophie depuis l'antiquité, son acception informatique commence à apparaître sporadiquement dans les années 1980 dans le milieu de l'Intelligence Artificielle ; il est alors employé pour désigner aussi bien une théorie d'un monde modélisé qu'un composant de systèmes à base de connaissances. Au début des années 1990 le consortium *Knowledge Sharing Effort* (KSE), soutenu par la *Defense Advanced Research Projects Agency* (DARPA) entérine le terme. Ce consortium a émergé du constat de l'apparition de nombreux Systèmes Experts dans des domaines variés et dont la création nécessitait de construire une nouvelle base de connaissances *ex nihilo* pour chaque système, faute de méthodes et moyens pour les partager et les réutiliser (Patil et al., 1998). Thomas Gruber, qui participe au KSE, publie

¹⁶⁷ Nous pensons par exemple à tout ce qui a trait au raisonnement logique/pouvoir d'inférence qui s'appuie sur les différents axiomes posés par une ontologie donnée (voir par exemple [Bienvenu et al. \(2020\)](#)).

en 1995 un article dans lequel il propose une définition de l'ontologie qui fait aujourd'hui encore autorité :

« Un ensemble de connaissances représentées formellement est basé sur une conceptualisation : les objets, les concepts et les autres entités qui sont présumés exister dans un domaine d'intérêt et les relations qui existent entre eux (Genesereth & Nilsson, 1987). Une conceptualisation est une vue abstraite et simplifiée du monde que nous souhaitons représenter dans un but précis. [...] **Une ontologie est une spécification explicite d'une conceptualisation.** »¹⁶⁸ (Gruber, 1995, p. 908)

Si cette définition fait consensus dans la communauté¹⁶⁹ pour son élégante précision, la notion d'ontologie ainsi formulée englobe une variété importante d'objets qui diffèrent en termes d'applications visées (extraction d'informations, traduction, bibliothèques digitales, etc.), types de contenu (ontologies de haut-niveau, de domaine, de tâche et d'application), langages informatiques utilisés (OML, RDF, OWL, etc.) et formalismes (basés sur la logique, réseaux sémantiques, graphes conceptuels, etc.) (Slimani, 2015).

Le grand principe commun qui unit les multiples incarnations des ontologies dans toute cette variété est que ces dernières représentent des *concepts* (également nommés *classes*) liés entre eux par des *relations* (également nommés *propriétés*). Les relations qui en constituent la « colonne vertébrale »¹⁷⁰ sont d'ordre taxonomique (Guarino et al., 2009, p.2) mais des relations associatives d'une autre nature viennent également lier les concepts entre eux. Par exemple, une ontologie sur l'art pourrait organiser hiérarchiquement les concepts ARTISTE et PEINTRE ainsi que ŒUVRE et TABLEAU par des relations hiérarchiques telles que *est un type de*, tandis qu'une relation *crée* servirait à lier PEINTRE et TABLEAU. L'ontologie peut alors être décrite comme un ensemble de *triplets*,

¹⁶⁸ Nous traduisons : "A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them (Genesereth & Nilsson, 1987). A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. [...] An ontology is an explicit specification of a conceptualization."

La référence citée renvoie à : Genesereth, M. R. (1992). An Agent-Based Framework for Software Interoperability. *Proceedings of the DARPA Software Technology Conference*. Meridian Corporation, Arlington VA, 395-366.

¹⁶⁹ Bien que critiquée par certains comme Neuhaus (2018) qui considère que cette définition viole le principe selon lequel une définition doit être plus claire que le terme défini.

¹⁷⁰ Nous traduisons *backbone*.

c'est-à-dire de classes reliées par des propriétés (*Tableau_estUnTypeDe_Oeuvre*, *Peintre_estUnTypeDe_Artiste*, *Artiste_crée_Tableau*). Notons que ces propriétés sont orientées, et que l'on désigne généralement le premier argument (*Tableau* dans le premier exemple) comme *domain*, tandis que le second (*Œuvre* dans le premier exemple) comme *range*. D'autre part, la notion d'*instance* renvoie à un individu particulier d'une classe ; dans notre exemple *Botticelli* serait une instance de PEINTRE et *Le Printemps* une instance TABLEAU. Une ontologie peuplée d'instances est appelée une *base de connaissances*, même s'il existe une certaine confusion terminologique dans l'usage, étant donné qu'il est souvent difficile de positionner la frontière entre ces deux objets (Noy & McGuinness, 2001).

1.2. Méthodes d'ingénierie ontologique

L'ingénierie ontologique est un domaine qui repose sur des « choix de représentation qui permettent de saisir les distinctions pertinentes d'un domaine au plus haut niveau d'abstraction tout en étant aussi clair que possible sur la signification des termes »¹⁷¹ (Gruber, 2009, p. 1965). Si la tâche de constitution d'une ontologie n'est pas chose aisée et qu'il n'existe pas *une* méthodologie de référence, plusieurs propositions ont été formulées à cet effet et quelques grandes lignes directrices sont partagées par la communauté ; l'une d'entre elle est notamment le fait que l'application visée pour l'ontologie est l'élément le plus déterminant des choix de conception (i.e. une ontologie est toujours conçue par rapport à un objectif applicatif). Parmi les propositions de méthodologies les plus fréquemment citées, Fernández-López et al. (1997) (méthode *Methontology*) souhaitent ancrer la création d'ontologies dans un processus d'ingénierie opposé aux pratiques « artisanales »¹⁷² qui sont souvent mobilisées. Ils définissent que le processus de développement d'une ontologie doit inclure des phases : de planification des tâches à réaliser, de spécification des objectifs, d'élicitation des connaissances, de conceptualisation, de formalisation, d'intégration d'ontologies existantes, d'implémentation en langage formel, d'évaluation, de documentation et, enfin, de maintenance. Noy & McGuinness (2001) (méthode *Ontology Development 101*) proposent quant à elles de commencer par déterminer le domaine et l'étendue de l'ontologie – ce qui

¹⁷¹ Nous traduisons : “representational choices that capture the relevant distinctions of a domain at the highest level of abstraction while still being as clear as possible about the meanings of terms”.

¹⁷² Nous traduisons *craft* (p. 33).

peut notamment être réalisé en dressant une liste des *questions de compétences* (Gruninger & Fox, 1995) à laquelle l'ontologie visée devra être en mesure d'apporter des réponses –, puis d'explorer les possibilités de réutilisation d'ontologies existantes, de lister les termes importants de l'ontologie et enfin de créer les différents éléments de l'ontologie. D'autres suggestions de méthodologies ont été proposées, et nous renvoyons le lecteur au Chapitre III de Gómez-Pérez et al. (2004, pp. 107-198) pour en découvrir davantage.

La majeure partie des ontologies existantes aujourd'hui partagées et utilisées par la communauté ont été développées manuellement ; cela a pour conséquence de susciter un phénomène important de *goulot d'étranglement* dans l'acquisition des connaissances¹⁷³, problème auquel le TAL constitue une des solutions possibles¹⁷⁴ (Somodevilla et al., 2018). Les travaux qui s'inscrivent dans le domaine que l'on nomme *apprentissage des ontologies* visent à proposer des solutions d'automatisation pour alléger les processus de création et d'amélioration/maintenance en termes de ressources et de temps. Elles présentent également l'avantage de pouvoir limiter le recours à la nécessaire consultation d'experts pour la phase de conceptualisation, ces derniers étant souvent fort occupés (Mondary, 2011, p.1). Somodevilla et al. (2018) distinguent quatre types fondamentaux d'apprentissages d'ontologies : à partir de textes, par *linked data mining*, par apprentissage de concepts par les logiques de description¹⁷⁵, et par *crowdsourcing*. Asim et al. (2018) proposent quant à eux de catégoriser les techniques d'apprentissage d'ontologies selon qu'elles soient de nature linguistique, statistique ou logique. La diversité des approches est donc considérable, même si certains grands procédés sont récurrents comme, dans le cas de l'apprentissage à partir de textes, l'utilisation de patrons pour extraire des concepts et/ou relation dans les textes¹⁷⁶ (par exemple Pantel &

¹⁷³ Nous traduisons *knowledge acquisition bottleneck* (p. 37).

¹⁷⁴ Goncharova (2017, p. 34) donne une liste de techniques du TAL qui ont été utilisées dans ce but, comme par exemple l'extraction thématique, la reconnaissance des entités nommées ou encore l'extraction d'événements.

¹⁷⁵ Les *logiques de description* sont un ensemble de langages utilisés pour la représentation des connaissances et qui sont naturellement liés aux ontologies qu'ils permettent de décrire. Cette approche « par apprentissage de concepts par la Logique de description » consiste en l'apprentissage des schémas axiomatiques et instances d'ontologies existantes.

¹⁷⁶ Ces travaux s'inscrivent une tradition initiée par Hearst (1992) qui propose un modèle de patrons destinés à extraire des relations d'hyponymie ; par exemple le patron NP_0 such as $\{NP_1, NP_2 \dots, (and/or)\} NP_n$ (p.1) permet de déterminer que toutes les unités lexicales $NP_1 \dots NP_n$ sont des hyponymes de NP_0 .

Pennacchiotti (2006) ou Ghadfi et al. (2014)) ou encore l'application du modèle d'*analyse formelle des concepts*¹⁷⁷ (par exemple Mondary (2011) ou Bendaoud et al. (2010)).

1.3. Les ontologies pour l'analyse sémantique textuelle

Les ontologies sont utilisées à plusieurs fins et dans plusieurs domaines tels que le Web Sémantique, la recherche d'informations ou l'ingénierie logicielle. Nous nous intéressons ici à leur usage pour de l'analyse sémantique textuelle.

De nombreux systèmes d'analyse textuelle reposent sur des ontologies de type réseaux sémantiques qui organisent le lexique en termes de relations lexicales (synonymie, hyperonymie, etc.) et dont WordNet (Miller, 1995) est le représentant le plus communément employé. D'autres types d'ontologies permettent également de réaliser différents types de traitements sémantiques. Ma et al. (2009) proposent une architecture d'« ontologies étendues » pour l'annotation de texte qui repose sur un système de règles qui permettent de distinguer ce qui relève de l'annotation ontologique (des concepts) de ce qui relève du peuplement (des instances). Salguero & Espinilla (2018) utilisent des règles basées sur une ontologie qui représente un document comme une suite séquentielle – de phrases, de mots – afin de proposer une analyse visant à la détection de langage discriminant. Kontopoulos et al. (2013) proposent l'utilisation d'ontologies de domaine comme alternative à l'utilisation de classifieurs pour l'analyse des sentiments sur des *tweets*. Alkhatib et al. (2018) utilisent une approche basée sur une ontologie qui tient lieu de classifieur multilabel sans entraînement et dont l'originalité réside dans le fait que l'ontologie n'est pas créée en amont mais générée à partir du texte. Enfin, Halawi et al. (2018) proposent un système basé sur des ontologies pour détecter des *spams* dans les *tweets* sur la base de leur contenu.

Ces quelques exemples donnent un aperçu des possibilités variées qu'offrent de telles approches qui sont souvent utilisées comme alternatives aux méthodes statistiques dont on cherche ainsi à combler les carences. En outre, le recours à des ontologies peut être un choix qui entre en complémentarité avec le ML ; elles peuvent notamment aider à sélectionner les caractéristiques données en entrées aux algorithmes (Sikelis et al., 2021),

¹⁷⁷ L'analyse formelle des concepts (Ganter & Wille, 1999) est un modèle mathématique qui permet de créer une hiérarchie de concepts à partir de données.

ou encore à contrer l'effet « boîte noire » de ces derniers en tentant d'expliquer leurs choix (Croce et al., 2020).

Après cette présentation des ontologies, nous expliquons le processus que nous avons mis en place pour réaliser celle qui servira à construire notre système d'autocodage.

2. Processus de modélisation experte

Nous expliquons dans ce point notre processus de collecte d'informations expertes sur la méthode CAUTIC®, puis la manière dont nous avons utilisé ces informations pour construire une ontologie, que nous décrivons.

2.1. Collecte d'informations et données stéréotypiques expertes

La première étape de notre processus de modélisation consiste à récolter une description experte du processus d'analyse humaine CAUTIC®. Nous expliquons ici la méthodologie de collecte et d'analyse suivie puis précisons les informations que nous avons ainsi obtenues.

2.1.1. Méthodologie

Nous choisissons de réunir les analystes de l'entreprise et leur proposons de passer en revue les 20 Critères de la méthode CAUTIC®, en leur demandant à chaque fois d'explicitier ce que signifie ce Critère et de donner plusieurs exemples de verbatim ou d'indices linguistiques qui les pousseraient à classer une unité textuelle de telle ou telle manière. Nous assumons lors de cette réunion une position d'animatrice naïve malgré notre propre maîtrise de l'analyse, afin de préserver notre positionnement neutre dans le cadre de ces travaux.

Nous transcrivons ensuite l'intégralité de cette réunion puis l'analysons en extrayant pour chaque Critère des exemples de verbatim stéréotypiques complets ou semi-lexicalisés (« *j'ai pas compris [moment de la présentation]* »), des indices linguistiques courts (mot ou court syntagme) et des descriptions de type d'éléments recherchés qui ne sont pas des exemples (*Éléments en lien avec la sécurité des personnes et des données*). Ces extractions sont consignées dans le [TABLEAU EN ANNEXE 17](#) (pp. 312-316). Nous réalisons ensuite une analyse de ces données dont nous décrivons ci-dessous les points saillants des résultats.

2.1.2. Points saillants des résultats de l'analyse

Nous mettons ici en exergue quelques points particuliers qui ont été soulevés pendant la réunion de concertation et qui concernent : 1) les champs conceptuels couverts par certains Critères complexes, 2) les liens entre les Critères CAUTIC® et la structuration de la présentation du Concept aux participants.

2.1.2.1. Champs conceptuels couverts par certains Critères complexes

Les commentaires des analystes Ixiade nous permettent d'avoir une idée plus précise des champs conceptuels couverts par chacun des Critères. Nous présentons ici les cas problématiques, pour lesquels la simple lecture des intitulés des Critères mérite d'être éclaircie par les précisions fournies par les analyses. Nous proposons quelques extraits illustratifs de la transcription de la réunion auxquels nous ferons référence dans notre commentaire.

- a. « De toute façon, ce Critère-là [1.1] est plus un Critère de synthèse, plus qu'un Critère de... plus qu'un Critère CAUTIC® »
- b. « [le 1.1,] il y a un bout du 1.2 dedans »
- c. « L'intérêt de certaines fonctions, donc ça va être spécifique sur des choses qui permettent des actions, etc. Alors que dans le premier, c'est : « est-ce qu'on le trouve intéressant *a priori* ? » »
- d. « Il y a deux sous-Critères dans ce Critère. »
- e. « C'est un peu compliqué parce que tu as l'intérêt et la compréhension. »
- f. « A condition qu'elle intervienne au début d'un entretien et pas n'importe où. »
- g. « Bah ça dépend en fait si le 1.1 tu l'abordes du coup de manière chronologique, ou si c'est la synthèse. Sur l'aspect compréhension, personnellement, pour moi, c'est la synthèse, et sur l'aspect intérêt, là, on est plus [+] sur la chronologie puisqu'en plus on nous dit dans la question « intéressant *a priori* ». »

Tout d'abord, le Critère 1.1 – « *L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?* » est sujet à débat entre les experts et ce pour plusieurs raisons. Premièrement, il s'agit d'un Critère de « synthèse » de la compréhension du Concept par l'interviewé, ce qui provoque le fait que les limites conceptuelles entre ce Critère et les

autres du même Niveau sont parfois floues (a., b.). Certains indices permettent cependant de bien définir le type de contenu qui va être classé dans ce Critère par rapport au 1.2 (celui avec lequel on trouve le plus d'ambiguïté) : si l'intérêt exprimé par le participant porte sur un élément précis du Concept / une fonction, le verbatim sera classifié en 1.2, tandis que s'il porte sur le Concept dans son ensemble, il sera classifié dans le 1.1 (c.). De plus, ce Critère va parfois servir à classifier des verbatim qui portent sur la compréhension et que l'on ne peut classifier dans aucun des autres Critères du Niveau 1. Deuxièmement, il englobe deux notions : la compréhension et l'intérêt (d., e.). Enfin, la notion d'*a priori* dans l'intitulé du Critère pose problème dans le sens où cet *a priori* n'est souvent pas interprété comme tel à travers une quelconque verbalisation de l'interviewé mais plutôt selon un critère chronologique : il s'agit souvent de verbatim qui sont exprimés immédiatement après qu'on lui a présenté le Concept (f.). Ce point ne fait cependant pas consensus entre les membres de l'équipe. Une hypothèse soulevée est que ce qui a trait à la compréhension relève de la synthèse, et ce qui porte sur l'intérêt est dépendant de la chronologie (g.).

D'autre part, les Critères 1.3 – « *L'utilisateur peut-il raccrocher des techniques mobilisées par le concept à des techniques existantes ?* », et 2.2 – « *Le concept compléterait-il sans les concurrencer les habitudes existantes ?* », appellent tous deux à rechercher des comparaisons avec l'existant. La différence réside dans la distinction entre les éléments qui suscitent la comparaison. Dans le Critère 1.3, le déclencheur est à chercher dans les connaissances du participant et son analyse des caractéristiques du Concept. On se situe bien au niveau cognitif. Pour le Critère 2.2, la comparaison est déclenchée par une pratique proposée. Il n'en reste pas moins que cette distinction reste délicate à identifier dans de nombreux contextes (h.) ; il sera difficile de classifier les syntagmes de type *ça existe déjà* si le contexte ne spécifie pas suffisamment la nature du coréférentiel de *ça*.

- h. « Dans cette idée de cercles concentriques et d'ouverture, là, c'est vraiment, encore une fois, du technique, du cognitif et presque du spontané : « *ça me fait penser à ci, à ça* » parce que *ça* a les mêmes caractéristiques et non pas « *ça me fait penser à ça* » parce que « *avec ça, je fais la même chose* ». Donc, c'est pas tout à fait pareil. C'est... la nuance, là elle va être difficile à trouver. »

Le Critère 1.5 – « *L'usage du concept est-il envisageable comme un système d'usage courant ?* », conduit à rechercher des indices de fréquence et des mentions de contextes d'utilisation. Un système de TAL qui chercherait de tels indices obtiendrait certainement des résultats satisfaisants, mais si l'on souhaite être précis, il faudrait que la mention de ces fréquences et contextes soit motivée par un degré de compréhension, c'est-à-dire qu'il faudrait que ce soit, par exemple, la simplicité du Concept qui provoque son utilisation courante (i.). Cependant, les probabilités que cette chaîne causale soit explicitement verbalisée par le participant sont plutôt faibles, et il est raisonnable de décider de se limiter à la recherche des indices de fréquences et contextes seulement. C'est d'ailleurs ce qui est souvent pratiqué aujourd'hui par les analystes. De plus, se limiter à une telle recherche permet d'éviter les chevauchements avec les verbatim du Critère 2.1 – « *Le concept est-il simple à utiliser ?* ».

- i. « Le Critère s'est un peu transformé. Parce qu'en fait, ce Critère 1.5, normalement, il est sur le principe technique et sur la compréhension du principe technique. C'est-à-dire que le principe technique est perçu suffisamment simple pour que la personne puisse envisager, à chaque fois que l'occasion se présente, de l'utiliser. »

Les verbatim classés dans les Critères 2.2 – « *Le concept compléterait-il sans les concurrencer les habitudes existantes ?* » et 2.3 – « *La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?* » sont souvent difficiles à distinguer les uns des autres. Dans les grandes lignes, il s'agit, dans le premier cas, de juger de l'existence ou non de pratiques similaires à celles proposées par l'innovation et, pour le deuxième, d'évaluer ces nouvelles pratiques par rapport aux anciennes, en complétant ainsi la réponse précédente (j.). Cette distinction cependant est souvent difficile à réaliser face aux données et représente une difficulté récurrente dans l'analyse.

- j. « 2.3, c'est le pourquoi du 2.2. [...] C'est presque une question de : « est-ce que ça s'oppose, ou est-ce que ça s'insère ? ». Et ensuite c'est : « en quoi ça s'oppose et en quoi ça s'insère ? » Et donc, « qu'est-ce que ça amène ? ». »

Le Critère 3.4 – « *Le concept suscite-t-il un imaginaire d'appropriation ?* », n'inspire que peu d'exemples aux experts. Ils citent notamment les noms d'animaux car ce Critère a souvent été traité sous forme de portrait chinois durant la collecte ; on posait aux participants la question « Si le concept était un animal, lequel serait-il ? ». Ce procédé de collecte est de moins en moins utilisé. La notion d'évaluation de l'activation ou non d'un « imaginaire d'appropriation » par le Concept chez les participants est difficilement saisissable à travers des unités textuelles courtes seules et les analystes sont très sceptiques quant à la possibilité que le TAL y parvienne (k.).

- k. « A mon avis, on est sans doute hors champ de l'analyse TAL. Parce que là, bah là, faut... ça demande de faire appel à des corpus de connaissances pour pouvoir interpréter des choses. Donc c'est pas parce que les gens vont s'exprimer autour d'un serpent ou d'un castor qu'on va faire quelque chose de ça quoi. »

La frontière entre les Critères 4.2 – « *Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?* » et 4.3 – « *Le concept est-il adapté à la place qu'occupent les professionnels dans la filière économique / aux relations familiales de l'utilisateur et à leur évolution ?* » est souvent difficile à délimiter. Un possible élément de distinction est le fait que le Critère 4.2 concerne un cercle plus restreint (collègues, entreprise, famille), tandis que le 4.3 est un peu plus large en ce qu'il aborde davantage des domaines d'activité ou des métiers en général (l.).

- l. « On a le sentiment que le 4.2 c'est plus dans un écosystème de travail restreint, tu vois, clients, fournisseurs, machin, et le 4.3 on est plus... entre autres types d'activités, quoi, c'est beaucoup plus large. »

2.1.2.2. Liens entre certains Critères et description du Concept

Cette réunion a également été l'occasion de rappeler qu'il existe un rapport entre certains Critères et la manière dont le Concept est décrit aux participants. En effet, la présentation qui est faite du Concept aux participants, comme nous l'avons précédemment introduit (Partie I, Chapitre I, point 2.3.1, pp. 14-15), est toujours

structurée autour de sept questions¹⁷⁸. Or, certains Critères sont directement liés à ces dernières. Par exemple, le Critère 1.2 – « *Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?* » fait directement écho à la question « Comment ça marche ? », et l'ensemble des Critères du Niveau 2 tendent à répondre à « Comment ça s'utilise ? ».

Cette phase nous a permis d'avoir une description précise des éléments que nous allons chercher dans le texte ainsi que de nombreux exemples et points de réflexion à partir desquels nous avons commencé à construire une modélisation.

2.2. Constitution de l'ontologie et peuplement préliminaire

Dans ce point, nous décrivons la manière dont nous avons construit notre modèle et le décrivons. Pour cela, nous expliquons tout d'abord comment nous avons conçu notre ontologie à partir des informations collectées auprès des experts, puis précisons comment nous envisageons de créer un système de règles à partir de cette dernière. Nous proposons ensuite une description des classes de l'ontologie résultante puis de la manière dont nous leur associons un lexique.

2.2.1. Constitution de l'ontologie

Notre choix de modélisation s'est porté sur une représentation conceptuelle par ontologie assortie à des règles de classification qui utilisent des patrons de catégories lexico-sémantiques *ad hoc*. Nous précisons la nature de l'ontologie créée puis expliquons la démarche suivie pour la réaliser à travers une description méthodologique générale puis en expliquant comment nous effectuons une augmentation progressive de l'ontologie à partir des données d'amorce.

2.2.1.1. Nature de l'ontologie

Avant de détailler ce que contient cette ontologie, précisons tout d'abord sa nature. La difficulté inhérente à la création d'une ontologie est de définir, étant donné un but – ici, l'analyse sémantique selon CAUTIC® – quelle sera la conceptualisation la plus adaptée. En d'autres termes, il s'agit de répondre à la question : « Pour parvenir à tel objectif, que doit représenter l'ontologie ? ». Nous faisons le choix de nous rapprocher au plus près de la

¹⁷⁸ Pour rappel, ces sept questions sont : « Dans quel contexte le Concept s'inscrit-il ? », « Qu'est-ce que c'est ? », « A qui ça s'adresse ? », « A quoi ça sert ? », « Comment ça s'utilise ? », « Comment ça marche ? » et « Quels sont les bénéfices ? ».

situation dans laquelle se trouve l'analyste humain confronté au texte à traiter ; ce dernier doit prendre en compte plusieurs éléments qui influencent l'analyse tels que le Concept, sa présentation, le participant en tant qu'individu et sa parole en tant qu'expression de ses opinions et ressentis. Il nous a donc semblé opportun de chercher à faire en sorte que l'ontologie soit une modélisation du paradigme de l'entretien dans toute sa complexité. Nous entendons ici par « entretien » toute situation dans laquelle un intervieweur questionne un interviewé quant à un Concept innovant, indépendamment du mode de collecte (en face à face ou en ligne, en groupe ou individuellement). D'autre part, nous préférons le terme de *paradigme* de l'entretien à celui de *situation* que nous trouvons trop restrictif et peu approprié en ce qu'il met l'accent sur l'occurrence de l'entretien en elle-même seulement et non sur tous les éléments qui gravitent autour de cet événement comme l'environnement familial de l'interviewé ou la prise en compte d'un marché concurrentiel pour le Concept. Nous pouvons donc définir notre ontologie comme suit : il s'agit de la conceptualisation du paradigme d'entretien qualitatif d'acceptabilité, construite à visée d'analyse sémantique portant sur les objets et événements particuliers de ce paradigme.

2.2.1.2. Démarche générale et structure de base

D'un point de vue informatique, nous utilisons le logiciel Protégé¹⁷⁹ qui est un éditeur d'ontologies très répandu, et utilisons le format OWL. En plus d'une manipulation aisée, il permet une visualisation en forme de graphe qui est pratique pour le partage d'informations à des membres de l'équipe non-initiés au domaine.

En ce qui concerne le processus, nous nous basons sur notre analyse de la réunion de concertation précédemment commentée dans ce chapitre pour dégager les concepts et relations que notre modélisation doit intégrer. Nous tirons donc ces concepts et relations des explications fournies par les experts et les verbatim stéréotypiques et indices linguistiques. Précisons que, étant donné qu'il est bien connu qu'un contenu linguistique imaginé correspond rarement à un contenu linguistique réel, les éléments lexicalisés fournis par les experts n'ont aucunement vocation à se substituer aux données produites lors d'études réelles ; il ne s'agit pas de se baser sur un contenu linguistique artificiel, mais seulement d'utiliser ces éléments comme des amorces permettant d'identifier les

¹⁷⁹ <https://protege.stanford.edu/> (consulté le 09/12/21).

concepts recherchés par les experts dans les textes qu'ils analysent. Cette amorce cèdera d'ailleurs rapidement le pas à un appui sur les données réelles du corpus.

Nous commençons par introduire les classes et propriétés de base de l'ontologie, telles que *Concept*, *User*¹⁸⁰, *Interviewer* et *Presentation*¹⁸¹ pour les classes et *showsPresentation* ou *focusesOnConcept* pour les propriétés (qui relie respectivement *Interviewer* à *Presentation* et *Presentation* à *Concept*). Notons que nous utiliserons indifféremment *utilisateur*, *interviewé* et *participant* dans la suite de ce texte pour nous référer à la personne qui est interrogée dans le cadre de l'entretien et qui est associée à la classe *User*.

En outre, nous avons pris le parti de créer la classe *ReactionToConcept* qui représente le fait que l'interviewé émette une réaction verbale au Concept. En procédant ainsi, nous considérons l'expression d'une opinion comme un événement de l'entretien, dont la nature sera précisée par la sous-classe. Par exemple, *Understanding*, *Question* ou *NewPossibilityPerception* seront des sous-classes de *ReactionToConcept* qui feront respectivement référence au fait que le participant : 1) s'exprime sur sa compréhension du Concept, 2) pose une question, ou bien 3) fasse un commentaire qui indique le fait qu'il perçoive que le Concept est porteur d'une nouvelle possibilité (« *ça me permettrait d'avoir un relevé de mon parcours* »). Ce parti pris nous amène donc à considérer l'expression d'une réaction au Concept comme une entité à part entière, entité à travers laquelle on accède à l'information que l'on cherche à acquérir, i.e. l'opinion du participant. Il s'agit là d'une sorte de contorsion conceptuelle qui permet de traiter dans la pratique deux niveaux d'informations différents – ce qui relève du factuel et ce qui relève de la perception – de la même façon, tout en les distinguant facilement. De plus, la classe *ReactionToConcept* est reliée par la relation *isAbout* à tout un ensemble de classes de l'ontologie ; par exemple, pour *ThreatPerception* qui est une sous-classe de *ReactionToConcept* indiquant la perception d'une potentielle menace apportée par le Concept, le triplet *ThreatPerception_isAbout_UserJob* permet de représenter l'événement : « l'utilisateur exprime le fait qu'il perçoit une menace apportée par le Concept qui concernerait son métier ». Nous illustrons dans la FIGURE 10 quelques classes reliées à celle de *User* dont la classe *ReactionToConcept*. Les classes y sont représentés dans les

¹⁸⁰ L'interviewé, utilisateur potentiel du Concept.

¹⁸¹ La présentation du Concept – vidéo ou diapositives – montrée à l'interviewé.

formes ovales en orange. Les propriétés sont indiquées par des flèches bleues ; celles qui ne sont pas étiquetées correspondent à la relation taxonomique *subclassOf*. Enfin, en bleu foncé, sous les classes, sont donnés des exemples d'unités lexicales susceptibles de dénoter ces classes dans le texte. Nous présentons également, selon le même modèle de représentation, quelques classes qui gravitent autour de la classe *Concept* dans la [FIGURE EN ANNEXE 20](#) (p. 317).

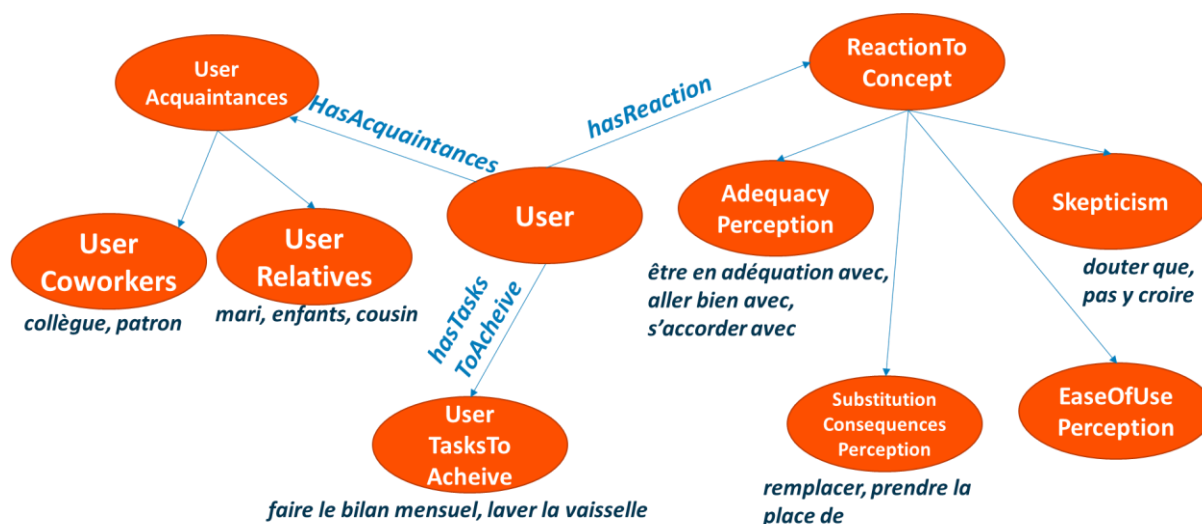


FIGURE 10 : Exemples de classes gravitant autour de la classe User

Avant de développer la méthode utilisée pour augmenter cette ontologie, nous souhaitons ouvrir ici une parenthèse sur sa granularité. Notre modélisation vise à identifier plusieurs éléments de l'opinion exprimée par l'interviewé et l'on comprend que, l'objet de nos travaux étant un système de classification, il s'agira donc de déterminer quelle opinion doit être classée dans quel Critère (en l'occurrence, l'exemple que nous venons de présenter devra déclencher une classification dans le Critère 3.2 – « Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ? »). Cependant, hors du contexte de la tâche de classification, il est à noter que la simple extraction de cette information dans le texte – le fait que l'interviewé exprime la perception d'une menace pesant sur son métier – est intéressante en elle-même pour l'analyste. Le système imaginé, si la tâche de classification venait à ne pas être aussi satisfaisante qu'escompté, devrait permettre *a minima* à l'analyste de repérer dans le texte l'expression de telles opinions. D'autre part, il est envisageable que, sur la base de cette même modélisation, on puisse construire d'autres systèmes d'analyse, en changeant les règles de classification associées aux relations identifiées, en retranchant une partie de l'ontologie ou encore en

l'augmentant. Pour garantir cette possibilité d'exploration du texte par type d'opinion et cette ouverture vers d'autres systèmes de classification, nous faisons le choix d'adopter un niveau de granularité plus fin que nécessaire. Par exemple, la classification en Critères CAUTIC® nécessite d'identifier dans le texte les mentions des fonctions du Concept. En revanche, le fait que cette fonction soit une fonction de *feedback*, d'enregistrement ou encore de chargement de données n'a aucune importance pour classer le verbatim. Nous choisissons tout de même de modéliser ces différents types de fonctions, afin de laisser la possibilité pour l'analyste, en utilisant le même modèle, d'effectuer une recherche sur les réactions liées à un certain type de fonctions du Concept seulement s'il le souhaite. Autre exemple, la classification de verbatim dans le Critère 1.5 – « *L'usage du concept est-il envisageable comme un système d'usage courant ?* » nécessite souvent de pouvoir identifier la mention d'une fréquence. Cependant, le fait que cette fréquence soit élevée (*tous les jours*) ou faible (*tous les 36 du mois*) ne change rien à la classification ; nous décidons cependant d'inclure cette distinction dans la modélisation (*HighFrequency* vs. *LowFrequency*), car d'autres méthodes d'analyse pourraient avoir besoin de cette information qui serait à nouveau accessible sans avoir à modifier le modèle.

2.2.1.3. Augmentation à partir des données d'amorce stéréotypiques

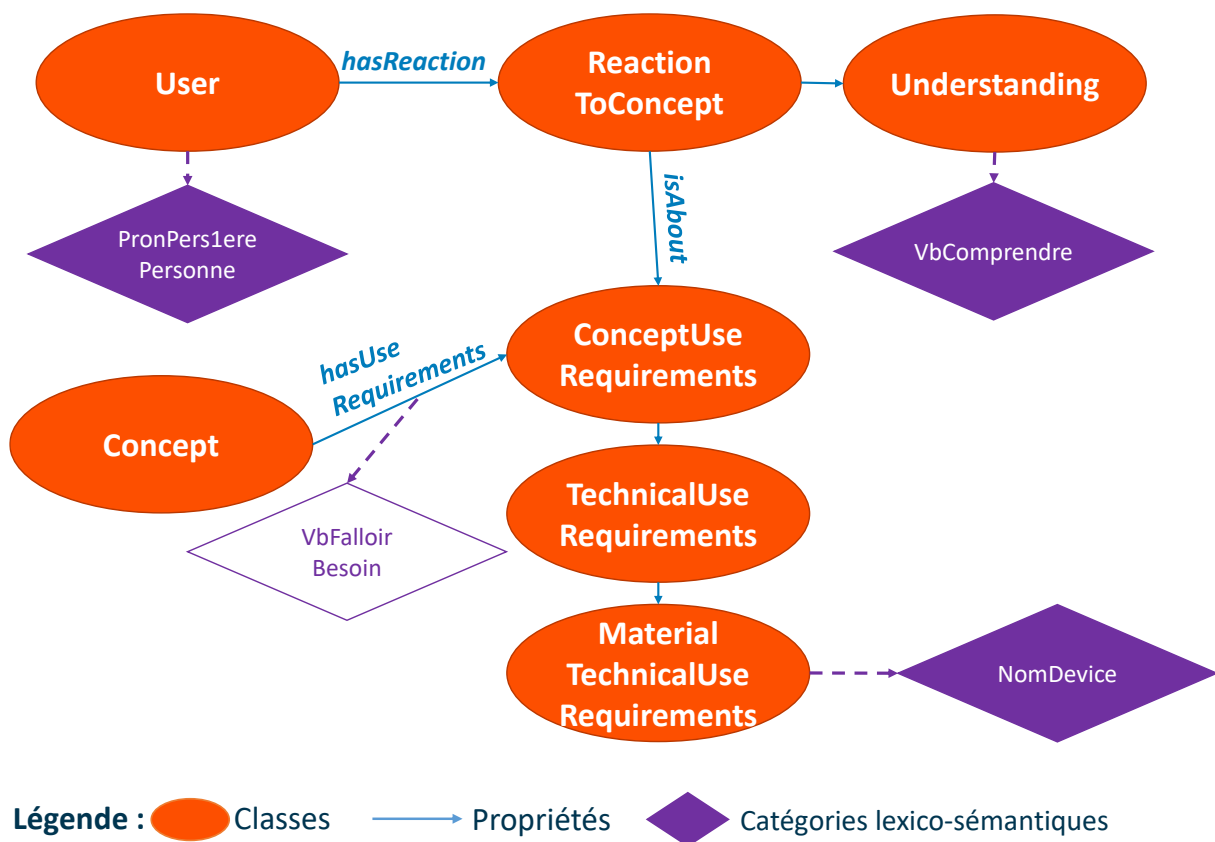
Nous refermons à présent la parenthèse pour présenter la méthode que nous avons utilisée pour augmenter l'ontologie. Au fil de l'analyse des verbatim types précédemment collectées, nous ajoutons les classes et propriétés nécessaires pour représenter l'ensemble des éléments pertinents de ces verbatim. Pour ce faire, nous transformons la phrase donnée en patron de catégories lexico-sémantiques *ad hoc* ; le verbatim « *j'ai compris qu'il y avait besoin d'un port USB* » donne ainsi lieu au patron *[PronPers1erePersonne]₍₁₎ [VbComprendre]₍₂₎ [que] [VbFalloirBesoin]₍₃₎ [NomDevice]₍₄₎*. Ensuite, à partir de ce patron, nous construisons une représentation sémantique constituée de triplets de l'ontologie, en faisant correspondre les catégories lexico-sémantiques à une classe ou une propriété, et en créant au fil de l'eau les classes et catégories lexico-sémantiques nécessaires à cette modélisation si elles n'existent pas dans l'ontologie. Nous obtenons ainsi, pour cet exemple :

```

Concept_hasUseRequirements(3)_MaterialTechnicalUseRequirements(4)
+ User(1)_hasReaction_Understanding(2)
+ Understanding(2)_isAbout_MaterialTechnicalUseRequirements(4)

```

Les chiffres en indice montrent la correspondance entre les catégories lexico-sémantiques et les classes et propriétés : *PronPers1erePersonne* est associée à *User*, *VbComprendre* à *Understanding*, *VbFalloirBesoin* à *hasUseRequirements* et *NomDevice* à *MaterialTechnicalUseRequirements*. La FIGURE 11 fournit une représentation de cet exemple et présente d'autres verbatim possiblement modélisés par cet ensemble de triplets. En outre, afin d'illustrer plus clairement le processus que nous venons de décrire ici, nous proposons dans l'ENCADRÉ EN ANNEXE 1 (pp. 317-319) une explication du processus – manuel – que nous utilisons pour modéliser chaque verbatim.



Exemples de verbatim correspondants :

- « J'ai compris qu'il fallait un câble USB. »
- « J'ai pas bien saisi si on avait quand même besoin d'une tablette ou si un téléphone suffit. »
- « Ah mais j'avais pas tilté que c'était obligatoire d'avoir des plaques à induction, je pensais que ça marchait même sur du gaz. »

FIGURE 11 : Exemple de modélisation

2.2.2. Vers un système de règles

Les grandes lignes du modèle ayant été décrites, nous expliquons ici comment nous entendons aller vers un système de règles sur la base de l'ontologie créée. Le principe

général appliqué est que nous partons de l'idée que la modélisation des verbatim stéréotypiques telle que décrite ci-dessus permet d'accéder à un niveau de généralisation suffisant pour que la modélisation de chaque verbatim constitue une règle d'annotation.

Nous présentons tout d'abord le fonctionnement envisagé d'un tel système puis la manière dont nous avons vérifié la validité des règles par consultation des experts de l'entreprise.

2.2.2.1. Fonctionnement du système de règles

Patron linguistique	[PronPers1erePersonne] ₍₁₎ [VbComprendre] ₍₂₎ [que] [VbFalloirBesoin] ₍₃₎ [NomDevice] ₍₄₎
Triplets	Concept_hasUseRequirements ₍₃₎ _MaterialTechnicalUseRequirements ₍₄₎ User ₍₁₎ _hasReaction_Understanding ₍₄₎ Understanding ₍₂₎ _isAbout_MaterialTechnicalUseRequirements ₍₄₎
Verbatim exemple	« J'ai compris qu'il y avait besoin d'un port USB »
Explication règle	L'utilisateur exprime sa compréhension des prérequis matériels nécessaires pour utiliser le concept.
Critère	1.2 – Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?

TABLEAU 26 : Exemple de règle de classification produite sur la base d'un verbatim stéréotypique

Nous obtenons ainsi pour chaque verbatim un ensemble de triplets qui constituent une règle de classification – nous lui attribuons le Critère qui était associé au verbatim type – et dont les différents constituants sont reliés à des catégories lexico-sémantiques. Autrement dit, il s'agit d'aller vers un système qui fonctionnera selon la logique suivante : le contenu textuel sera associé à des catégories lexico-sémantiques qui « activeront » des triplets de l'ontologie, et la présence d'un ensemble de triplets fonctionnera comme une règle qui provoque l'attribution d'un Critère. L'exemple fourni dans le point précédent donne ainsi lieu à la règle illustrée dans le TABLEAU 26. On peut y voir que la règle est expliquée par une phrase qui décrit l'événement que cette règle modélise (« L'utilisateur exprime sa compréhension des prérequis matériels nécessaires pour utiliser le concept. ») ; nous concevons cette phrase de description comme une possibilité de répondre au besoin de transparence de l'algorithme pour les analystes, en ce que le codage d'un verbatim pourrait être accompagné de cette explication de l'interprétation du texte par l'algorithme, justifiant le choix qu'il réalise. Afin donner un aperçu plus large des verbatim que nous traitons et règles que nous générons, nous proposons une liste de

règles et des verbatim à partir desquels elles ont été créées dans le [TABLEAU EN ANNEXE 18](#) (pp. 323-326). Les verbatim qu'il contient sont les verbatim stéréotypiques imaginés par les analystes. Le [TABLEAU EN ANNEXE 19](#) (pp. 327-330) est construit à l'identique, mais contient quant à lui des verbatim de notre corpus de données réelles d'études. Ces derniers ont été utilisés pour construire des règles dans une étape successive que nous décrirons dans le chapitre suivant. Nous les présentons cependant dès à présent au lecteur afin qu'il puisse se faire une idée du type de contenu que nous sommes amenée à traiter à terme, et pour souligner la différence de diversité de contenu entre les verbatim imaginés et les données réelles. Précisons que, par souci de lisibilité, nous ne transcrivons dans ce tableau que la partie d'intérêt du verbatim, et que les segments présentés étaient en réalité contenus dans des verbatim plus longs.

Cette proposition de système fait donc de l'ontologie non seulement un outil pour repérer des concepts d'intérêt dans le texte, mais lui fait également endosser le rôle de support de l'interprétation sémantique qui permet la classification. En d'autres termes, nous utilisons l'ontologie comme un artefact fournissant un « plan d'interprétation compositionnelle restreinte » du texte. Cette notion de « restriction » correspond au fait que seuls les concepts d'intérêt pour l'analyse seront pris en considération. Nous envisageons que cela puisse permettre de répondre à deux enjeux fondamentaux de notre tâche. Le premier est la distinction entre ce qui est pertinent pour l'analyse et ce qui ne l'est pas, dans la mesure où seuls les concepts d'intérêt seraient repérés et utilisés pour la classification. Le deuxième est la nécessité d'interpréter non seulement la présence de concepts, mais aussi leur articulation sémantique, comme des déclencheurs de l'annotation.

2.2.2.2. Validité des règles

Il nous paraît nécessaire de vérifier que les règles, en tant qu'association entre un ensemble de triplets basé sur un verbatim stéréotypique et un Critère CAUTIC® soient pertinentes. Autrement dit, est-ce que l'ensemble de triplets

Concept_hasUseRequirements_MaterialTechnicalUseRequirements
 + *User_hasReaction_Understanding*
 + *Understanding_isAbout_MaterialTechnicalUseRequirements*

doit bien donner lieu à une classification dans le Critère 1.2 ? Ou bien s'agit-il d'une surinterprétation de notre part à partir du contenu de la réunion de concertation ? Ou bien encore, ces triplets peuvent-ils correspondre à un autre Critère sans que cela n'eût

été évoqué lors de cette réunion ? Il s'agit donc de valider les règles, mais également d'identifier si le problème provient d'une modélisation maladroite à partir du verbatim d'exemple, ou bien si c'est le verbatim d'exemple lui-même qui s'avère, finalement, peu représentatif du Critère.

Afin de procéder à cette évaluation, nous réunissons trois analystes de l'équipe et leur présentons la méthode de modélisation que nous utilisons afin qu'ils se l'approprient et qu'ils soient en mesure de comprendre la signification des triplets¹⁸². Puis, nous leur demandons d'effectuer une évaluation des règles à l'aveugle via un questionnaire en ligne. Dans cette optique, nous leur proposons d'attribuer à chaque ensemble de triplets un Critère, puis, dans un deuxième temps, nous leur demandons d'attribuer également un Critère à chaque verbatim, sans qu'ils ne sachent quel triplet correspond à quel verbatim. Ces deux évaluations, réalisées séparément, permettent d'évaluer l'adéquation entre le verbatim et sa modélisation ainsi que la pertinence de la règle : si, évalués séparément, l'ensemble de triplets présenté ici en exemple et le verbatim qui a servi à le construire sont tous deux majoritairement attribués au Critère 1.2, alors la règle est validée. S'ils sont majoritairement tous deux attribués à un autre Critère, alors nous conservons la règle et la réattribuons à cet autre Critère. Si le verbatim et la règle sont attribués à deux Critères différents, il est alors nécessaire de remettre la modélisation en question, car on peut estimer qu'elle ne reflète pas le contenu du verbatim. Enfin, si la classification de la règle et/ou du verbatim donne lieu à un accord inter-juge peu satisfaisant, alors se pose la question de la pertinence du choix de conserver la règle ou bien de devoir prendre collectivement une décision quant à l'attribution d'un Critère à la règle.

Sur ce dernier point, les résultats de cette évaluation permettent d'identifier des configurations conceptuelles qui semblent être problématiques en termes de classification, sans que l'on ne puisse décider de ne pas les prendre en compte dans l'analyse, car l'information qu'elles contiennent reste importante. Par exemple, la réponse aux attentes de l'utilisateur par le Concept (« *ça répond exactement à ce que j'attendais !* ») est une information importante dans une perspective d'évaluation de l'acceptabilité mais sa classification ne donne pas lieu à un consensus ; les règles et verbatim qui comportent

¹⁸² Au moment où nous avons réalisé cette évaluation, les phrases d'explication de règles n'avaient pas été créées.

ce type d'informations sont classées dans les Critères 1.1, 1.5, 2.2, 2.3, 2.4, 3.1, 3.3 et 3.4. La perception du progrès, avec la règle

User_hasReaction_ProgressPerception
+ *ProgressPerception_isAbout_Concept*

est proposée deux fois aux évaluateurs (car elle a été créée deux fois à partir de deux verbatim différents : « *je trouve ça pas du tout innovant* » et « *c'est révolutionnaire* ») et donne lieu à deux classifications différentes ; cela se comprend de par le fait que l'idée de progrès et de nouveauté est rattachée aux connaissances de l'utilisateur (Niveau 1), mais aussi aux tendances du marché (Niveau 4). Il n'en reste pas moins que ce type de verbatim apporte une information intéressante et qu'ils devront être conservés dans l'analyse.

A la suite de cette évaluation nous décidons de présenter aux analystes les cas problématiques qui ont été identifiés et leur demandons de prendre des décisions pour chacun de ces cas. Il est décidé, pour quelques règles, de conserver deux Critères possibles.

2.2.3. Description de l'ontologie résultante

Au fur et à mesure de la création de règles et de leur évaluation, l'ontologie est augmentée. Si nous avons présenté jusqu'ici sa constitution selon une logique linéaire, dans la pratique, il s'est plutôt agi d'un processus itératif. Aussi, présenter sa volumétrie à l'instant T entre ce que nous sommes en train de décrire et la suite n'est pas pertinent. La description que nous présentons ici contient donc des classes que nous avons ajoutées dans les étapes que nous décrirons plus tard. Cette description vise à donner une idée au lecteur de ce que contient notre ontologie et la manière dont ses concepts sont articulés.

Tout d'abord, comme nous l'avons déjà souligné, nous considérons les réactions de l'interviewé comme des classes à part entière. Ces réactions se décomposent en plusieurs sous-classes, dont nous fournissons plusieurs exemples dans l'ENCADRÉ EN ANNEXE 2 (p. 320). Nous trouvons, par exemple, *ConceptTargetPerception*, qui a pour définition : « Réaction au Concept dans laquelle l'utilisateur s'exprime sur « à qui s'adresse le Concept ? » », ou bien *NewPossibilityPerception* qui dénote une « réaction qui implique la perception de possibilités nouvelles portées par le Concept ». Nous trouvons également *Opinion*, qui désigne des réactions positives ou négatives générales sans considération d'objet ou de réflexion spécifiques, comme dans des verbatim du type « *ce concept est*

super ! » ou « *je n'aime pas du tout ce que vous m'avez présenté* ». Les règles qui utilisent ces sous-classes fonctionnent souvent sur un principe de trois triplets de type

```
User_hasReaction_[sous-classe de ReactionToConcept]
+ [sous-classe de ReactionToConcept]_isAbout_Concept
+ [sous-classe de ReactionToConcept]_isAbout_[élément sur lequel porte la réaction]
```

par exemple

```
User_hasReaction_AdequacyPerception
+ AdequacyPerception_isAbout_Concept
+ AdequacyPerception_isAbout_UserStatus
```

qui signifient donc : « L'utilisateur exprime une réaction de perception d'adéquation. Cette perception d'adéquation porte sur le Concept. Cette perception d'adéquation porte sur son propre statut ». Ils peuvent être également interprétés comme « L'utilisateur exprime une réaction de perception entre le Concept et son propre statut » ou encore « L'utilisateur exprime une réaction de perception d'adéquation entre le Concept et son propre statut ». On comprend donc que, dans ce type fréquent de configurations de triplets formant une règle, la relation *isAbout*, lorsqu'elle est appliquée au Concept, a une signification qui se rapproche de 'est suscité par' plus que de 'porte sur'. Cependant dans d'autres cas de figure, la relation *isAbout* qui lie une réaction au Concept conserve bien le sens de 'porte sur', comme dans *Understanding_isAbout_Concept*. Il aurait été donc plus correct de créer deux propriétés différentes. Cependant, cela aurait nécessité des complexifications de l'ontologie qui n'apporteraient rien d'utile en termes d'efficacité opérationnelle du modèle, et nous décidons donc de ne pas corriger cette imprécision sémantique.

En outre, nous créons la classe *ReactionModulation* qui a pour sous-classes *APrioriReactionModulation*, *ChangesConditionsReactionModulation*, etc. Cette classe est reliée à *ReactionToConcept* par la propriété *modulatesReaction* et permet d'apporter une information qui vient modifier la nature de la réaction. Ainsi, nous faisons le choix de modéliser, par exemple « *j'ai du mal à comprendre* » par

```
User_hasReaction_Understanding
+ DifficultyToReactionModulation_modulatesReaction_Understanding
```

soit : « Le participant exprime une réaction qui porte sur la compréhension. La compréhension est impactée par une modulation qui implique une difficulté ». Autre exemple, la phrase « *Il faut que ce soit abordable* » est modélisée par les quatre triplets

User_hasReaction_ValuePerception

+ *GeneralConditionsReactionModulation_modulatesReaction_ValuePerception*

+ *ValuePerception_isAbout_Concept*

+ *ValuePerception_isAbout_LowCosts*

soit : « Le participant exprime une réaction qui consiste en la perception d'une valeur. La réaction de perception de valeur, qui porte sur le concept et un prix bas, comporte une condition » que l'on interprète aisément comme 'le participant dit que le prix du Concept doit être bas'.

Enfin, dans les grandes lignes, et afin de donner un aperçu global de l'ontologie, nous pouvons dire que l'ontologie contient :

- Des classes relatives aux caractéristiques du Concept, comme *ConceptCompany* qui correspond à l'entreprise qui propose ce Concept, ou bien *ConceptProofingParticularities* qui fait référence aux propriétés de résistance du Concept tel le fait d'être *waterproof* ou bien résistant à la foudre.
- Des classes qui peuvent être utilisées comme relatives au Concept ou non, comme *Functions* qui peut se référer si bien à une des fonctions du Concept qu'à une fonction d'un autre objet.
- Des classes relatives à l'utilisateur comme *UserKnowledge*, *UserSkills* et *UserAcquaintances* qui font respectivement référence aux connaissances, aux compétences, et à l'entourage de l'utilisateur, ou encore *UserTasksToAcheive* qui se rapporte aux tâches qu'il doit accomplir, qu'elles soient professionnelles (faire un compte rendu) ou bien quotidiennes (se laver les dents).
- Des classes plus générales que l'on va lier au Concept ou à l'utilisateur mais qui ne le sont pas par nature, comme *Frequency* qui pourra être utilisée pour décrire la fréquence imaginée d'utilisation du Concept, la fréquence avec laquelle l'utilisateur réalise une tâche, ou encore la fréquence à laquelle le Concept doit subir une maintenance.
- Des classes liées à la présentation du Concept. Ces dernières incluent la classe *PresentationParts* qui est subdivisée selon les sept questions de définition du Concept et dont chacune est reliée par la propriété *describes* à une sous-classe de *ConceptCharacterisitics* (par exemple, le triplet *HowDoesItWorkPart_describes_ConceptFunctionning* peut être traduit par « la partie « Comment ça

marche ? » de la présentation du Concept décrit le fonctionnement du Concept ».

Des exemples plus nombreux pour chacun de ces types de classes sont fournis dans les [ENCADRÉS EN ANNEXE 3 à 7](#) (pp. 320-322).

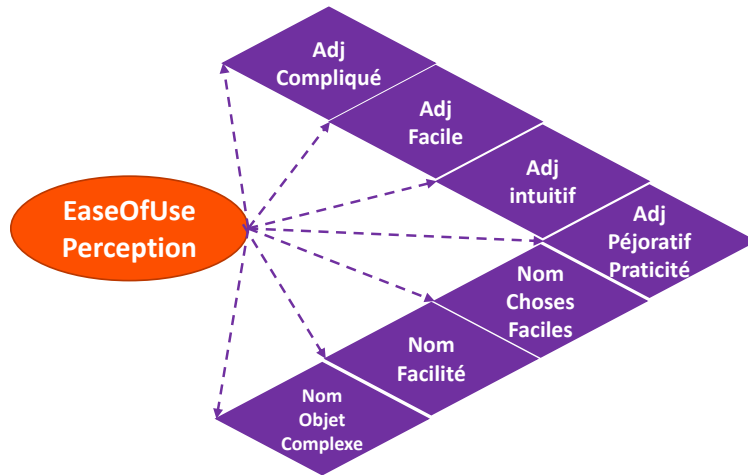
Les contours de l'ontologie réalisée ayant été tracés, nous allons à présent nous intéresser à la partie lexicale du système envisagé.

2.2.4. Constitution et peuplement du lexique

Comme nous l'avons dit, nous avons créé des catégories lexico-sémantiques *ad hoc*. Il convient alors de créer un lexique qui corresponde à ces catégories. Pour cela, nous structurons ce lexique selon les modalités que nous décrivons à présent.

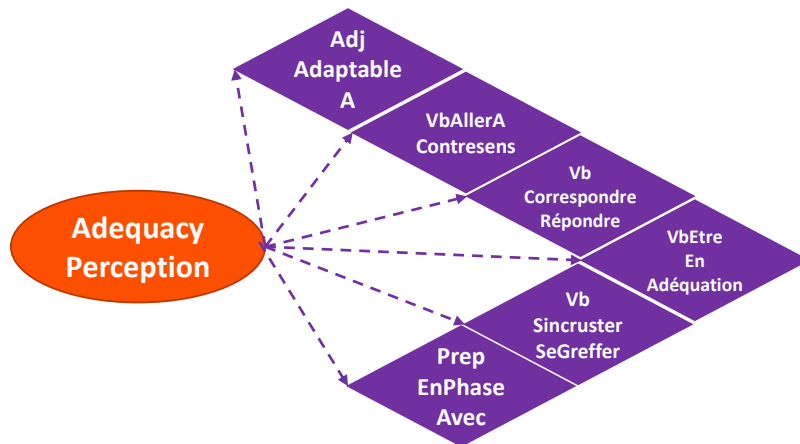
Cette description s'appuiera sur l'exemple des catégories associées à deux sous-classes de *ReactionToConcept*, à savoir *EaseOfUsePerception* et *AdequacyPerception*. La première vise à identifier l'expression d'une facilité ou difficulté d'utilisation du Concept imaginée par l'utilisateur. La seconde renvoie à l'expression de la perception d'une adéquation ou inadéquation du Concept avec n'importe quel élément que ce soit : l'organisation quotidienne de l'utilisateur, les tendances du marché, la taille de son appartement, etc. Les FIGURES 12 et 13 proposent une description des différentes catégories lexico-sémantiques associées à ces classes et des différentes unités lexicales qui sont, à leur tour, associées à ces catégories lexico-sémantiques. Comme pour les classes que nous avons présentées au point précédent, chacun des éléments présentés pour ces exemples n'avait pas forcément été ajouté à notre modèle à l'issue de cette phase de constitution préliminaire de l'ontologie. Les unités lexicales, notamment, proviennent en partie de travaux d'augmentation du lexique que nous présenterons au Chapitre IX.

Comme le lecteur pourra donc le constater sur les figures, plusieurs catégories lexico-sémantiques sont associées à la même classe. Ces catégories sont nommées par combinaison d'une ou plusieurs catégorie(s) morphosyntaxique(s) et d'une ou deux unité(s) lexicale(s) représentative(s) du sens que cette catégorie cherche à englober. Par exemple *VbSincrusterSeGreffer* désigne les constructions verbales dont le signifié renvoie à cette idée d'incrustation (dans)' et de 'greffe (sur)'. Ensuite, à chacune de ces catégories sont attribuées des unités lexicales qu'il s'agira de repérer dans le texte à analyser. Ces unités lexicales sont de quatre types possibles :



Classe lexicio-sémantique	Lemmes	Formes	Chaîne lemmes	Patron lemCat
AdjCompliqué	complexe, ardu	alambiqué, tarabiscoté		
AdjFacile	évident, enfantin			
AdjIntuitif	intuitif, accessible		facile de utilisation	
AdjPéjoratifPraticité	contre-intuitif, chiant		difficile à utiliser	
NomChoseFacile	jeu d'enfants, mince affaire			
NomFacilité	aisance, simplicité	facilité		
NomObjetComplexe	usine à gaz			

Figure 12 : Lexique pour les catégories lexicio-sémantiques associées à la classe *EaseOfUsePerception*



Classe lexicio-sémantique	Lemmes	Formes	Chaîne lemmes	Patron lemCat
AdjAdaptableA	flexible, modulable, aménageable		adaptable à	
VbAllerAContresens			aller à contresens	aller ADV à contre-courant de
VbCorrespondreRépondre			correspondre à	être ADV aligné avec
VbÊtreEnAdéquation	concorde, s'accorder		coller avec, être en harmonie avec	être ADV cohérent avec
VbSincrusterSeGreffer	s'insérer		s'incruster dans, se greffer sur	s'incruster ADV sur
PrepEnPhaseAvec	en phase avec, en accord avec			

FIGURE 13 : Lexique pour les catégories lexicio-sémantiques associées à la classe *AdequacyPerception*

- Des lemmes, qui peuvent être mono ou polylexicaux (*complexe, usine à gaz*). Les lemmes polylexicaux renvoient généralement à des unités ayant un certain degré de figement, mais il peut également s'agir de groupes non figés, du moment qu'ils forment un seul groupe nominal, verbal, adjectival, adverbial ou prépositionnel.
- Des formes, auxquelles nous avons recours pour préciser les éléments recherchés dans le texte particulièrement sujets à des ambiguïtés catégorielles. Par exemple, pour la catégorie *NomData*, nous mettons dans les formes *données* et *donnée*, pour qu'elles soient identifiées aussi bien quand elles seront étiquetées avec le lemme *donnée* que *donner* (le système d'étiquetage dans le corpus, nous y reviendrons, permettra de ne pas repérer dans ce cas les occurrences de *donnée(s)* dans des constructions verbales composées).
- Des chaînes de lemmes, qui peuvent constituer ou non des unités figées composées de groupes de catégories morphosyntaxiques différentes (ou susceptibles d'être étiquetées comme telles même si cela ne correspond pas strictement à la réalité linguistique), comme *difficile à utiliser* ou *aller à contresens*.
- Des patrons qui mêlent lemmes et catégories morphosyntaxiques, permettant notamment d'anticiper des insertions d'adverbes dans des groupes verbaux (*être ADV cohérent avec*).

D'autre part, précisons qu'une même catégorie lexico-sémantique peut être associée à plusieurs classes ou propriétés. Dans ce cas, les listes d'unités lexicales peuvent varier selon la classe ou propriété à laquelle la catégorie est associée. Par exemple, la catégorie *VbCorrespondreRépondre*, lorsqu'elle est associée à la propriété *answersProblems*, contiendra le lemme *résoudre* (« *ça permet de résoudre tous mes soucis* ») ce qui n'est pas le cas lorsqu'elle est associée à *meetsExpectations* (« *ça résout ce que j'attendais* »). En ce sens, une entrée du lexique est définie par une association « catégorie lexico-sémantique – classe/propriété » et non par la catégorie lexico-sémantique seule.

Deux éléments supplémentaires viennent compléter ce lexique. Tout d'abord, toute catégorie lexico-sémantique impossible à lexicaliser de manière à tendre vers l'exhaustivité, par exemple *GvAction*, reçoit une liste de « patrons de seconde couche ».

Cette dénomination renvoie au fait que ces patrons ne seront recherchés dans les textes qu'après que le reste a été étiqueté, sur les éléments qui n'ont pas déjà reçu une étiquette. Ainsi, un des patrons de seconde couche pour *GvAction* est *VERB NP* qui peut par exemple correspondre à *range la maison*. Ce patron permettra d'étiqueter ainsi ce syntagme si et seulement si *range* et *maison*, ou bien tout le syntagme, n'ont pas été étiquetés avec une autre catégorie. D'autre part, ce procédé étant particulièrement susceptible de créer du bruit, nous en limitons l'impact de deux façons. La première consiste simplement à limiter autant que possible le nombre de classes qui ont recours à ces patrons de seconde couche. La seconde consiste à créer une catégorie lexico-sémantique purement fonctionnelle qui ne servira qu'à attribuer une étiquette vide à des verbes dont on ne veut pas qu'ils soient repérés par des patrons de seconde couche. Ainsi, le verbe *constater* faisant partie de cette catégorie, *constate un défaut* ne pourra pas être étiqueté comme *GvAction* dans la mesure où le verbe *constater* aura déjà un étiquetage.

Le deuxième élément supplémentaire répond également à une logique d'étiquetage venant se superposer à une phase précédente. Il s'agit d'un lexique qui inclut des conditions de contexte. Il contient par exemple l'entrée constituée de la catégorie *PrepPour* associée à la classe *ConceptTargetPerception*. Cette entrée contient des lemmes tels que *pour*, à destination de ou à l'attention de et a pour condition de contexte droit la présence d'une classe telle que *Job* ou *TargetMarket*. Ainsi, elle permettra d'étiqueter *pour* dans *c'est pour les électriciens*, dans la mesure où *électricien* aura précédemment reçu l'étiquette *Job*, mais pas *c'est pour prendre la température*.

Enfin, soulignons les particularités de la catégorie associée à la classe *Concept*. Cette catégorie est particulièrement délicate à traiter car la détection de la mention du *Concept* est importante pour distinguer le pertinent du non pertinent ; *c'est relativement facile à utiliser* ne constitue un verbatim d'intérêt qu'à condition que *c'* se réfère au *Concept* et non à un autre objet. Or, le *Concept* étant le sujet central de l'entretien, il est rarement nommé et la résolution de l'anaphore n'est pas évidente. Cette catégorie lexico-sémantique, nommée simplement *C*, contient à ce stade des éléments lexicalisés tels que *concept*, *ce que vous me montrez*, *votre truc*, etc. Nous choisissons, dans un premier temps, de considérer que tous les pronoms *il*, *le*, *cela* peuvent s'y référer, ce qui aura bien évidemment comme conséquence de créer beaucoup de bruit. Nous reviendrons, par la suite, sur ce problème.

2.3. Conclusions



FIGURE 14 : Classes de l'ontologie

A partir de la collecte d'informations et exemples stéréotypiques experts, nous avons construit 1) une ontologie qui se veut être une modélisation du paradigme de l'entretien d'acceptabilité CAUTIC®, 2) un ensemble de catégories lexico-sémantiques *ad hoc*, 3) un lexique. Cet ensemble vise à nous permettre de construire un système de règles de classification qui fonctionne sur le principe suivant :

- Les unités lexicales du texte pertinentes pour l'analyse CAUTIC® seront associées à des classes lexico-sémantiques, elles-mêmes reliées à des classes et propriétés de l'ontologie.
- La présence de ces unités lexicales dans un segment de texte permet donc de lui associer une interprétation conceptuelle dénotée par les triplets possibles « activés » par ces unités lexicales.
- La présence concomitante de plusieurs triplets dans un segment de texte correspond à une règle de classification.

En cela, l'ontologie au cœur du système envisagé aura fonction de support à l'interprétation textuelle mais aussi de filtre entre le contenu pertinent et non pertinent. Cette ontologie comporte environ 160 classes dont la taxonomie est présentée dans la FIGURE 14.

Ce chapitre a donc permis d'expliquer le procédé de conceptualisation qui a été réalisé et de poser les bases d'une proposition de système d'analyse par règles sur une base d'ontologie. Les deux suivants présenteront les expérimentations menées sur la base de ce système ainsi que les améliorations proposées.

CHAPITRE VIII : PROJECTIONS DU MODÈLE EXPERT

Ce chapitre expose les modalités de projection du modèle expert décrit au chapitre précédent sur le corpus, ainsi que les premiers résultats en termes de classification que cette projection permet. Les expérimentations successives que nous y décrivons participent à améliorer le modèle de manière cumulative, c'est-à-dire que chaque amélioration permise par une de ces expérimentations est intégrée au modèle et en fait donc partie pour l'expérimentation suivante.

Nous commençons par décrire la manière dont cette projection du modèle est réalisée. Ensuite, nous présentons des expérimentations de classification menées sur la base de patrons, puis d'autres expérimentations basées, cette fois, sur le système de règles basées sur les triplets de l'ontologie telles qu'introduites au chapitre précédent.

3. Projection des classes sémantiques

Dans ce point, nous expliquons comment nous étiquetons notre corpus avec les catégories lexico-sémantiques – et classes et propriétés associées – de notre lexique. Nous apportons tout d'abord quelques précisions sur l'étiquetage morphosyntaxique réalisé, puis présentons les grands principes de la projection des catégories des modèles sur les données.

3.1. Étiquetage morphosyntaxique

Nous utilisons le corpus de verbatim filtré présenté dans le Chapitre IV. Dans les grandes lignes, nous conservons l'étiquetage qui avait été réalisé pour la projection de Verbænet – décrit au point 8.1.2 du Chapitre VI (pp. 149-151) – avec quelques modifications. La première est que nous remplaçons *TreeTagger* par *Stanza* (P. Qi et al., 2020). Cet étiqueteur est récent et obtient des résultats substantiellement meilleurs, principalement sur des textes susceptibles de contenir des fautes comme cela sera le cas pour notre application finale que devra traiter des *posts* issus d'internet¹⁸³.

¹⁸³ Selon une comparaison effectuée sur le français par Xiaou Wang et Xingyu Liu (consultable à https://nlpin french.fr/nlp/01_compara_anno_fr.html (consulté le 15/12/2021)), ainsi qu'une comparaison effectuée sur un échantillon de notre corpus réalisée par Alice Adnot dans le cadre de son stage de fin d'études en TAL au sein de l'entreprise.

Les autres modifications consistent de manière générale à déconstruire les groupes étiquetés pour la projection de Verbānet, pour laquelle nous avons besoin d'avoir des constituants de grande taille, en des groupes plus petits. En effet, si la précédente expérimentation nécessitait une analyse syntaxique centrée sur le verbe, cela n'est plus le cas à présent en ce que nous cherchons plutôt ici à obtenir des composants correspondant aux entrées de notre lexique, et que les relations syntaxiques n'ont plus d'importance. Ainsi, par exemple, l'étiquetage des propositions introduites par *que* n'ont plus lieu d'être. Nous n'avons pas besoin de l'information selon laquelle, dans « *je trouve que c'est très esthétique* », le groupe [*que c'est très esthétique*] est un actant du verbe *trouve*. Nous avons en revanche besoin d'identifier le groupe [*très esthétique*] de manière isolé, car il correspond à une entrée du lexique. Nous préférons donc dans ce cas un découpage en constituants qui ressemble davantage à : [*je*] [*trouve*] [*que*] [*c'*][*est*] [*très esthétique*]. Autre exemple, nous déconstruisons également certains GN ; le patron [*NP de NP*] permettant d'identifier des GN, nous avons dans le corpus des étiquetages tels que [*j'*][*ai*][*besoin d'informations*]. Or, *avoir besoin de* est une des unités lexicales du lexique et *information* en est une autre qui appartient à une entrée différente. Nous créons donc une liste de noms qui, s'ils sont suivis de *de*, viennent briser le patron [*NP de NP*]. Nous obtenons ainsi l'étiquetage [*j'*][*ai*][*besoin*][*d'*][*informations*] qui est plus approprié à la structure de notre lexique. Les noms appartenant à cette liste sont soit importants à identifier individuellement car leur sens a un intérêt pour nous, soit parce qu'ils constituent la partie nominale d'une expression polylexicale verbale de notre lexique. Le nom *besoin* de cet exemple remplit pour sa part ces deux critères. De telles règles spécifiques seraient très problématiques pour un corpus qui devrait être étiqueté selon un lexique très volumineux et généraliste. Nous utilisons pour notre part un lexique restreint et contrôlé qui permet de traiter les unités lexicales pertinentes pour notre analyse de manière spécifique tout en ignorant le reste. Nous pouvons donc nous permettre de créer un étiquetage qui peut contenir des incohérences linguistiques qui sont, certes, peu élégantes¹⁸⁴, mais sans impact.

¹⁸⁴ Si nous reprenons l'exemple que nous venons de citer avec *avoir besoin*, nous obtenons donc [*avoir*] [*besoin*] [*d'*] [*informations*]. Si notre lexique ne contient pas l'expression *avoir peur de* alors nous obtiendrions [*avoir*] [*peur des informations*]. Or, il n'existe aucune raison linguistique pour que ces deux syntagmes subissent un traitement différent, d'où une certaine incohérence.

D'autre part, nous choisissons de ne considérer comme verbes auxiliaires potentiels que *aller*, *faire* et *venir de*, afin de capturer plus facilement les nuances de sens qu'apportent les autres verbes. Par exemple, nous étiquetons [*pouvons*] [*nous connecter*], [*voulons*] [*nous connecter*] mais [*venons de nous connecter*]. L'identification des lemmes *pouvoir* et *vouloir* comme éléments autonomes permet en effet de relever des nuances de sens intéressantes pour l'analyse, tandis que le syntagme *venons de nous connecter* peut être considéré comme un seul verbe ayant pour lemme *connecter*.

Enfin, précisons – cela sera utile pour la compréhension du point suivant – que chaque groupe identifié reçoit une étiquette de forme, de lemme et de tête de groupe lemmatisée. Ainsi, le groupe nominal *une fonction intéressante* aura les étiquettes *form="fonction intéressante"*, *lemma="fonction intéressant"* et *headLemma="fonction"*.

3.2. Principes généraux de la projection

La projection du modèle va donc consister à étiqueter le corpus avec les classes et propriétés de l'ontologie par le repérage des unités lexicales de notre lexique qui les dénotent. Comme nous l'avons vu précédemment, les données lexicales dans le lexique appartiennent à différents types (lemmes, formes, chaînes de lemmes, patrons composés de lemmes et de catégories morphosyntaxiques, patrons de seconde couche). Ces types sont définis pour être en adéquation avec l'étiquetage du corpus et pour permettre de rechercher les syntagmes longs en priorité. En effet, pour chaque mot du corpus, nous cherchons s'il est le début d'un patron composé de lemmes et de catégories morphosyntaxiques, sinon, nous cherchons s'il est le début d'une chaîne de lemmes, puis s'il correspond à une forme et enfin à un des lemmes du lexique. L'application de cet ordre de priorité nous permet d'encoder des choix de désambiguïsation dans le lexique ; si nous prenons le mot *réseau*, nous pouvons décider qu'il doit par défaut recevoir une classe relative à son acception technique, mais que s'il est accompagné de l'adjectif *professionnel*, alors il devra être attribué à une autre classe. Il nous suffit pour cela d'ajouter au lexique *réseau professionnel* en tant que forme et *réseau* en tant que lemme. Ainsi, si l'on rencontre *réseau professionnel* dans le corpus, le repérage par la forme sera prioritaire par rapport à celui par le lemme, et ce groupe nominal recevra donc la classe souhaitée.

De plus, comme nous l'avons introduit au point 2.2.4 du précédent chapitre (§5, pp. 184-185) deux autres couches d'étiquetage sont appliquées : l'étiquetage des patrons

de deuxième couche qui vient « combler les trous » laissés par la première couche, et celui des étiquettes conditionnées par leur contexte droit et/ou gauche.

```
<SEM lingCat="C;PronGénérique" semClass="Concept;NULL"
lingLemmesAndPatterns="cela">Ça</SEM> <SEM lingCat="VbExister"
semClass="CompetitiveProductsServices "
lingLemmesAndPatterns="exister">existe</SEM> <SEM lingCat="AdvNomDomaine"
semClass="NonTargetMarketDomains;TargetMarketDomains" lingLemmesAndPatterns="dans
gestion">dans la gestion des interphonies</SEM>, <SEM
lingCat="NomMétier;NomMétierUtilisateur" semClass="Jobs;UserJob"
lingLemmesAndPatterns="technicien">les
techniciens</SEM> en agences, <SEM lingCat="C;PronPers3emePersonne"
semClass="Concept" lingLemmesAndPatterns="il">ils</SEM><SEM lingCat="VbOrganiser"
semClass="UserTimeOrganization"
lingLemmesAndPatterns="programmer">programment</SEM> les badges des locataires
<SEM lingCat="AdjFaçonDeMarcher" semClass="ConceptFunctionning"
lingLemmesAndPatterns="à distance">à distance</SEM>. Avant <SEM
lingCat="C;PronPers3emePersonne semClass="Concept"
lingLemmesAndPatterns="il">ils</SEM> <SEM lingCat="GvDéplacement"
semClass="MotionFunctions;UserMotionTasksToAcheive" lingLemmesAndPatterns="aller
sur NP">allaient sur site</SEM> pour <SEM lingCat="VbOrganiser"
semClass="UserOrganization"
lingLemmesAndPatterns="programmer">programmer</SEM>/<SEM lingCat="GvAction"
semClass="UserTasksToAcheive;ConceptPurpose;ConceptUseContext;UserExistingPractic
es;UserAuthorizedTasksToAcheive;UserAcquaintancesActions"
lingLemmesAndPatterns="VINP NP">déprogrammer les badges</SEM>.
```

ENCADRÉ 7 : Exemple de verbatim étiqueté sémantiquement

Les unités pertinentes ainsi repérées, l'étiquetage correspond à l'ajout d'une balise <SEM> qui précise dans ses attributs la/les catégorie(s) lexico-sémantique(s) possible(s) ainsi que la/les propriété(s) et classe(s) associée(s). Un exemple d'étiquetage est présenté dans l'ENCADRÉ 7. Par souci de lisibilité, nous ne mettons dans cet exemple que les balises <SEM> et non celles qui relèvent de l'étiquetage morphosyntaxique. Les attributs *lingCat* indiquent les catégories lexico-sémantiques, *semClass* renvoie aux classes et propriétés correspondantes, et *lingLemmesAndPatterns* précise quelle unité lexicale de notre lexique

est identifiée. Le verbatim, dénué de balises, est : « *Ça existe dans la gestion des interphonies, les techniciens en agences, ils programment les badges des locataires à distance. Avant ils allaient sur site pour programmer/déprogrammer les badges.* ». Il est classifié dans le Critère 1.3 – « *L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?* ».

Nous nous appuyons sur cet exemple pour formuler quelques explications et remarques. Tout d'abord, nous pouvons noter l'efficacité de certains choix de structuration de ressources et d'algorithme d'étiquetage : la préposition *pour* est bien laissée sans étiquette car elle n'est pas suivie d'un élément indiquant une cible potentielle du Concept (cf. Chapitre VII, point 2.2.4, §6 p. 185). Le patron de seconde couche *VINF NP* permet bien d'identifier une action (*déprogrammer les badges*). Le groupe prépositionnel *dans la gestion des interphonies* est repéré à juste titre comme un marché cible ou non cible du Concept ; cela est possible grâce à la chaîne de lemmes *dans gestion* contenue dans le lexique pour la catégorie lexico-sémantique *AdvNomDomaine*. Cette classe contient de nombreuses chaînes telles que *dans maintenance, en animation, en électronique*, etc. Nous expliquerons au chapitre suivant comment ces syntagmes sont générés.

D'autre part, certains choix de désambiguïsation entre les classes potentielles ne sont pas réalisés car il est difficile d'imposer des règles strictes. Ainsi, *les techniciens* se voit attribuer une étiquette de classe *Job* et une autre *UserJob*. Nous avons en effet pris le parti de considérer que toutes les mentions de noms de métier pouvaient désigner le métier du participant ou le métier d'une autre personne. Ce choix rend *de facto* la distinction entre ces deux classes impossible et remet donc en question le bien-fondé de leur coexistence. Cependant, nous choisissons de les conserver et de rendre la distinction conceptuelle possible en nous laissant la possibilité d'inclure des restrictions supplémentaires sur une des classes si cette ambiguïté s'avère génératrice de trop de problèmes de classification ; nous pourrions par exemple imposer que le nom de métier soit précédé de *en tant que* ou *étant moi-même* pour qu'il soit attribué à la classe *UserJob*. Ce genre de décisions constitue un des leviers d'actions de réajustement du modèle dont tout l'enjeu est de trouver le meilleur équilibre. Cette impossible désambiguïsation *a priori* prend bien entendu une dimension toute particulière pour les *GvAction* qui ne sont repérés qu'à l'aide d'un patron catégoriel. Six classes sont en effet attribuées à cette

catégorie. Il va donc s'agir d'aller vers une quantité importante de lexicalisation. Cependant, ici encore, nous devons trouver un équilibre ; le verbe *programmer* avait été attribué à la catégorie *VbOrganiser* et non à *GvAction* et reçoit donc une classe différente du verbe *déprogrammer* que n'appartenait pas à notre lexique, ce qui provoque donc une incohérence certaine.

D'autres erreurs et imprécisions ponctuent cette annotation. Par exemple, la locution adverbiale à *distance* est étiquetée comme appartenant à la classe *ConceptFunctionning*. Or, ce n'est pas le fonctionnement du Concept qui est concerné ici mais celui d'un autre système. Cette erreur entre dans la problématique plus globale de la distinction entre ce qui a trait au Concept et ce qui se réfère à d'autres objets. D'ailleurs, le choix d'étiqueter tous les pronoms (*ça, ils*) comme désignant potentiellement le Concept crée bien évidemment des erreurs notables.

Nous avons donné dans ce point une explication concernant la manière dont nous étiquetons sémantiquement le corpus avec les catégories de notre modèle. Dans la suite, nous présentons des expérimentations menées par des systèmes de patrons extraits à partir de cet étiquetage.

4. Expérimentations sur des systèmes par patrons

Dans ce point, nous laissons de côté temporairement le système de règles fondé sur les triplets de l'ontologie afin de nous concentrer sur ce qu'il est possible de réaliser avec notre seul système d'étiquetage sémantique. Nous menons plusieurs expérimentations sur la possibilité de classification par patrons seulement, pour voir ce que ces derniers, à eux seuls, permettent de déduire en termes d'attribution à un Critère. Nous commençons par vérifier la performance technique de notre système d'étiquetage puis évaluons dans quelle mesure et selon quelles méthodologies cet étiquetage permet d'apprendre des règles de classification à partir du corpus.

4.1. Vérification de la performance technique de l'étiquetage sémantique

Nous cherchons ici à vérifier que notre système de projection des étiquettes sémantiques est suffisamment précis par rapport à ce que le lexique est censé permettre d'identifier. Nous présentons la méthodologie, les résultats et conclusions de cette vérification.

4.1.1. Méthodologie

Nous utilisons un pseudo-corpus constitué des verbatim stéréotypiques d'amorce à la création du modèle. Nous appliquons l'étiquetage que nous venons de décrire et extrayons toutes les chaînes de catégories possibles pour chaque verbatim. Chacun de ces verbatim est donc traduit en une chaîne de catégories, et nous devrions donc, sur cette seule base, pouvoir repérer à quel Critère il est associé. Si nous reprenons l'exemple fourni dans le TABLEAU 26 du chapitre précédent (p. 176), le verbatim « *J'ai compris qu'il y avait besoin d'un port USB* » a été traduit par la chaîne *[PronPers1erePersonne] [VbComprendre] [que] [VbFalloirBesoin] [NomDevice]*, et cette chaîne est associée au Critère 1.2. L'étiquetage réalisé devrait donc permettre de retrouver cette chaîne et d'attribuer le bon Critère à ce verbatim. Précisons que cette évaluation n'a pas vocation à évaluer le potentiel de réussite global d'un système par patrons, dans la mesure où nous utilisons un corpus de données « artificielles » qui sont les mêmes que celles qui ont servi à construire le modèle. Aussi, de très bons scores sur cet exercice ne permettent en aucun cas, bien entendu, de tirer quelque conclusion que ce soit sur les potentielles performances du modèle sur des données réelles. Il s'agit simplement de vérifier que l'étiquetage permet une bonne identification des configurations sémantiques que nous avons modélisées. Il s'agit également d'identifier les problèmes liés aux ambiguïtés lexicales et à la manière dont notre système de catégories lexico-sémantique *ad hoc* permet de les traiter.

Pour chaque verbatim stéréotypique, nous extrayons toutes les chaînes et sous-chaînes possibles et les comparons à celles contenues dans notre base de règles. A chaque fois qu'une correspondance est trouvée, le Critère associé est attribué. Cela a pour effet de, potentiellement, attribuer plusieurs Critères à un même verbatim. Dans l'exemple que nous avons fourni, une sous-chaîne possible étant *[PronPers1erePersonne] [VbComprendre]*, si un verbatim a été modélisé par cette seule chaîne et est attribué, par exemple, au Critère 1.1, alors le verbatim sera attribué aux Critères 1.1 et 1.2.

4.1.2. Résultats

Les scores obtenus, en termes de rappel et de précision qui évaluent la bonne attribution d'un verbatim à un Critère sont de $r=0,66$ et $p=0,90$. Les erreurs sont principalement dues à plusieurs facteurs.

Le premier de ces facteurs est la présence de doublons dans les sorties qui créent du bruit. Cela n'est en aucune mesure étonnant car nous appliquons un système multilabel sur des données classifiées en monolabel. Certaines des classifications en double peuvent, dans l'absolu, être considérées comme correctes, mais cela n'est bien entendu pas le cas pour tous les doublons. D'autres phénomènes sont également à prendre en compte. Par exemple, nous trouvons le verbatim « *de ce que vous m'avez montré je trouve ça pas du tout innovant* », originellement classifié dans le Critère 1.1 et donné deux fois en sortie, une fois dans le Critère 1.1 et une fois dans le Critère 1.2. Cette erreur est due au fait que la perception du progrès n'a pas suscité une décision tranchée de la part des experts. Aussi, le patron *[PronPers1erePersonne] [VbJuger] [C] [AdjMélioratifsProgrès]* est présent deux fois, associé à deux Critères différents. Il s'agit donc à ce stade d'une erreur tout à fait justifiable, liée aux règles, et considérée comme acceptable. En revanche, nous trouvons des doublons qui sont dus à des problèmes de sous-chaînes et d'ambiguïté lexicale dans les patrons ; le verbatim « *ça m'étonnerait qu'on voie ça un jour* », dont la classification originale est le Critère 4.1, est donné deux fois en sortie. La première sortie, correcte, est détectée par le patron *[PronGénérique] [PronCOI1erePersonne] [VbEtonner] [que] [PronPers 1erePersonne] [VbAvoirSurMarché] [C] [AdvTempsIndéfiniFutur]*. La deuxième classification, dans le Critère 1.1, est due à la sous-chaîne *[PronPers1erePersonne] [VbComprendre] [PronGénérique]*. Ce patron correspond au syntagme *on voie ça*. Le verbe *voir* étant susceptible de dénoter la compréhension, comme dans, « *oui, je vois bien* », il est nécessaire qu'il fasse partie du lexique dans la catégorie *VbComprendre*. Cette erreur est donc provoquée, d'une part, par un problème d'ambiguïté lexicale. D'autre part elle est due au fait qu'une sous-chaîne a donné lieu à une classification alors qu'elle était incluse dans une chaîne de taille plus importante qui précisait le sens de l'énoncé. Ce problème est difficilement corrigeable, bien que potentiellement soluble, par exemple en utilisant un automate qui permette d'identifier les chaînes les plus longues possibles parmi toutes celles possibles contenues dans l'ensemble de patrons en une seule passe (Antoniadis, 1984).

Le deuxième facteur d'erreur est le repérage des *GvAction*. En effet, comme nous l'avons précédemment expliqué, cette catégorie est annotée par patrons catégoriels lors d'une deuxième couche d'annotation qui ne prend en compte que les éléments qui n'ont pas déjà été étiquetés sémantiquement. Or, la présence de certains verbes dans des catégories existantes empêche l'étiquetage des *GvAction*. Par exemple, dans « *je vais*

l'utiliser pour changer mes pneus », *changer mes pneus* n'est pas reconnu comme *GvAction* car le verbe *changer* est déjà étiqueté comme *VbChanger*. Ce problème a déjà été expliqué plus en détails au point 3.2 (pp. 192-193) de ce chapitre.

En outre, l'impossibilité de repérer les noms d'entreprises et de produits est également source de bruits et silences. Les catégories du type *NomCompétition* (qui renvoie à la classe *CompetitiveCompanies*) sont pour l'instant difficilement activables en l'absence d'un lexique de noms d'entreprises ou d'un système de reconnaissance des entités nommées.

Enfin, certaines catégories doivent être lexicalisées mais le sont difficilement. Par exemple, la catégorie *GvPorterVêtement* ne peut être modélisée par des patrons tels que *porter NP* car cela entraînerait une trop forte ambiguïté. Or, lexicaliser tous les noms de vêtements et créer des patrons pour toutes les configurations possibles est difficile à réaliser. Cette catégorie avait d'ailleurs été ajoutée sur la base du syntagme *avoir les épaules dénudées* (dans le verbatim « *mais ça m'obligerait à avoir les épaules dénudées !* »)¹⁸⁵, qui n'a pas été étiqueté car nous avons jugé qu'il était inutile de l'inscrire dans la base lexicale tant elle est particulière à ce verbatim. Elle représente par conséquent assez bien le fait qu'il est très peu probable que ce type de syntagmes puisse être repéré et attribué, à terme, à la bonne classe.

Au-delà de ces problématiques, d'autres peuvent être résolues mais de manière non systématique, au cas par cas, et demeurent, à ce titre, solubles par un travail de longue haleine. Parmi elles, nous trouvons celles provoquées par les ambiguïtés catégorielles, problème bien connu du TAL. Nous avons limité l'impact de ce problème en n'imposant pas de catégorie morphosyntaxique pour qu'une unité soit repérée. Cependant, il arrive qu'une mauvaise désambiguïsation dans l'étiquetage de base attribue un lemme erroné¹⁸⁶. C'est notamment le cas de l'ambiguïté entre verbe et participe passé adjectif. Pour pallier ce problème, nous essayons de prendre en compte plusieurs possibilités dans notre lexique ; pour *AdjImpressionnéBluffé*, nous indiquons les lemmes *impressionné*, *bluffé*

¹⁸⁵ *avoir les épaules dénudées* a été attribué à la classe *GvPorterVêtement* car, dans le cadre de l'analyse qui nous intéresse, il dénote une obligation vestimentaire qui n'est pas appréciée par l'utilisateur ; il devrait être interprété de la même façon que *ça m'obligerait à porter un uniforme* et doit donc être attribué à la même catégorie.

¹⁸⁶ Au moment où nous avons réalisé cette expérimentation, nous n'avions pas encore remplacé *TreeTagger* par *Stanza*. Ce remplacement a par la suite contribué à améliorer le traitement de ces ambiguïtés.

et *impressionner*, *bluffer*. Mais cette solution, peu élégante d'une part, ne suffit pas, d'autre part, à résoudre tous les problèmes.

4.1.3. Conclusions

Cette expérimentation a permis de démontrer que le système d'étiquetage, malgré des problèmes dont la résolution semble délicate, permet de repérer avec une précision correcte les éléments que nous cherchons dans le texte, malgré des problèmes difficilement solubles qui impactent notablement le rappel. Dans la suite, nous cherchons utiliser cet étiquetage pour apprendre des règles de classification.

4.2. Apprentissage de règles par patrons sur le corpus

Dans ce point, nous continuons à n'utiliser que les patrons mais avons cette fois recours à des données réelles en utilisant notre corpus de verbatim. Une première projection à partir des règles par patrons déjà construites sur ce corpus nous donnent les scores, faibles, de $r=0,04$ et $p=0,30$. Ces résultats ne sont pas étonnants dans la mesure où ces premières règles par patrons et ce début de lexique n'avaient pas prétention à l'exhaustivité mais plutôt à la preuve de concept, et que la rigidité d'une méthode par patron couplée à un modèle pauvre appris sur des données stéréotypiques ne sont pas prometteurs d'une forte capacité de généralisation à des données nouvelles et réelles. L'objectif des expérimentations que nous présentons dans la suite de ce point est par conséquent de tester différentes méthodes de création de règles et augmentation du lexique afin de choisir la plus efficace pour améliorer le modèle et ces scores. Au-delà du fait d'obtenir des résultats expérimentaux, donc, cette étape est également l'occasion d'un ajout incrémental de classes et propriétés, règles et entrées lexicales.

Nous divisons le corpus en quatre sous-corpus de tailles égales, chacun comportant des verbatim de tous les Niveaux. Dans le Sous-Corpus 1, nous nous concentrons tour à tour sur les verbatim d'un des Niveaux et nous créons des règles à partir de ces derniers et d'une des méthodes possibles de création de règles. Puis, nous projetons les nouvelles règles – ajoutée aux précédentes – sur les trois autres sous-corpus afin d'évaluer l'impact que la création de ces règles a sur les scores. Nous avons initialement 148 règles que nous avons établies sur la base des verbatim stéréotypiques.

4.2.1. Modélisation manuelle

Pour cette première expérimentation, nous nous concentrons sur le Niveau 1, à l'exception du Critère 1.1 qui, nous l'avons vu, est un peu particulier en tant que Critère de synthèse. Nous créons manuellement les règles comme nous l'avions fait précédemment sur les verbatim stéréotypiques, c'est-à-dire en augmentant progressivement l'ontologie et le lexique par modélisation des verbatim du corpus. L'objectif de ce test est de voir si la méthode que nous avons employée jusqu'ici, assez laborieuse mais offrant une possibilité de contrôle sur chaque élément ajouté au modèle, permet d'augmenter les scores à partir de données réelles d'apprentissage. Les 57 règles par patron ainsi créées sont ensuite projetées sur ce même Sous-Corpus 1 qui a servi à leur élaboration, puis sur les trois autres sous corpus. Nous comparons les scores – en ne prenant en compte que les Critères du Niveau 1 – avant et après l'ajout de ces règles.

Sur le Sous-Corpus 1, nous passons, de $r=0,2$ et $p=0,25$ à $r=0,53$ et $p=0,8$. Sur les autres sous-corpus, nous passons d'une moyenne de $r=0,01$ et $p=0,17$ à $r=0,12$ et $p=0,29$. L'augmentation des scores n'est donc pas très importante mais elle reste significative. Ce processus manuel a un certain potentiel – qui reste limité – mais il est particulièrement chronophage.

4.2.2. Annotation manuelle et extraction

Nous nous penchons ici sur les Critères du Niveau 2. Nous faisons tout d'abord une première annotation automatique des catégories lexico-sémantiques contenues dans notre modèle. Puis, nous corrigeons et augmentons cette annotation manuellement selon les principes suivants :

- Nous ne gardons que les annotations formant un patron qui nous semble utile pour la classification (i.e., les annotations correctes mais inutiles sont supprimées).
- Nous désambiguïsons chaque annotation considérée comme utile, c'est-à-dire que nous ne laissons que l'étiquette de catégorie lexico-sémantique pertinente.
- Nous ajoutons, lorsque nécessaire, de nouvelles catégories lexico-sémantiques et annotons la catégorie du lexique.

Cette annotation réalisée, nous extrayons premièrement les nouveaux éléments du lexique que nous ajoutons à notre lexique existant. Puis, nous extrayons tous les patrons

qui deviennent ainsi les nouvelles règles, qui sont au nombre de 67. Nous ré-étiquetons les sous-corpus 2, 3 et 4 avec le lexique augmenté et y projetons les nouvelles règles.

Nous obtenons ainsi pour les Critères du Niveau 2 un passage de $r=0,13$ et $p=0,8$ à $r=0,66$ et $p=0,83$ sur le Sous-Corpus 1 qui a servi à construire les règles, et d'une moyenne de $r=0,03$ et $p=0,27$ à $r=0,04$ et $p=0,17$ sur les autres sous-corpus.

Cette méthode n'augmente donc que très légèrement le rappel et fait baisser la précision. Elle est, d'autant plus, également très chronophage.

4.2.3. Extraction des patrons et filtrage

Cette dernière expérimentation porte sur les Critères du Niveau 3. Nous souhaitons, dans un premier temps, effectuer une simple extraction des patrons déjà présents dans le corpus afin de les transformer directement en règles. Il est cependant vite apparu, par observation de ces patrons, qu'ils comportaient trop d'ambiguïté. Nous avons donc procédé de la façon suivante : dans un premier temps, nous lisons les verbatim et ajoutons à notre lexique, au fur et à mesure, les unités lexicales qui nous semblent pertinentes. Puis, nous effectuons une nouvelle annotation du corpus avec le lexique augmenté. Nous extrayons ensuite tous les patrons. Nous lisons ces patrons et sélectionnons les sous-chaînes qui nous paraissent être autoporteuses, i.e. dont la lecture seule (sans le verbatim associé) permet d'inférer la classification. Nous obtenons ainsi 17 nouvelles règles par patron (le corpus contenant peu de verbatim du Niveau 3), que nous projetons sur les autres sous-corpus.

Nous obtenons pour les Critères du Niveau 3 un passage de $r=0$ et $p=0$ à $r=0,46$ et $p=0,83$ sur le Sous-Corpus 1 et une moyenne de $r=0,03$ et $p=0,20$ à $r=0,03$ et $p=0,09$ sur les autres sous-corpus. Ces scores moyens sont obtenus grâce au Sous-Corpus 4, les deux autres ayant des scores de $r=0$ et $p=0$ aussi bien avant et qu'après l'ajout de règles. On constate donc que, pour le seul sous-corpus sur lequel on peut observer une variation, cette dernière correspond à une baisse des scores. Cette méthode ne semble donc pas adaptée.

4.2.4. Conclusions

Parmi les trois méthodes testées, seule la méthode manuelle permet, à ce stade tout du moins, d'augmenter les scores de rappel et de précision sur une méthode basée sur des patrons. A l'issue de cette étape, nous avons 289 règles, dont toutes n'ont pas été

traduites en triplets car les expérimentations, étant basées sur des patrons, ne les nécessitaient pas ; cela signifie que toutes les règles sont présentes telles qu'illustrées dans les colonnes « Patron correspondant » des TABLEAUX EN ANNEXE 18 et 19 (pp. 323-330), mais pas sous forme de triplets comme dans la colonne « Règles » de ces mêmes tableaux. Tous Niveaux confondus, sur cette seule base des 289 règles par patrons, nous obtenons sur les sous-corpus 2 à 4, une précision de 0,22 et un rappel de 0,03 seulement. Ces scores sont très bas, mais ce système de classification par patrons ne correspond pas, quoi qu'il en soit, à celui envisagé. Nous avons cependant pu enrichir notre modélisation et notre lexique. D'autres méthodes pourraient être explorées mais, pour l'heure, nous nous concentrons sur la réalisation du système de règles prévu, basé sur les triplets de l'ontologie.

5. Expérimentations sur des systèmes de règles basées sur les triplets de l'ontologie

Nous mettons à présent en œuvre le système de classification basé sur les triplets de l'ontologie dont nous avons décrit le fonctionnement général au chapitre précédent. Nous expliquons les choix algorithmiques effectués, décrivons une étape d'harmonisation des règles puis présentons les résultats obtenus.

5.1. Algorithmes de classification par règles basées sur les triplets de l'ontologie

Afin de présenter l'algorithme choisi pour identifier dans le texte des ensembles de triplets correspondant à des règles, nous introduisons ses principes généraux, puis les choix de paramètres utilisés, les résultats préliminaires obtenus et la conclusion que nous pouvons en tirer.

5.1.1. Principes généraux

Le système que nous réalisons est celui qui a commencé à être décrit dans les pages précédentes. Dans les grandes lignes, il s'agit tout d'abord d'identifier les potentiels triplets de l'ontologie que peuvent former les classes et propriétés annotées. Ensuite, nous cherchons à savoir si un sous-ensemble des triplets possibles dans le segment analysé correspond à une combinaison qui définit une règle de classification. Afin d'identifier la meilleure façon de procéder, nous testons plusieurs paramètres.

5.1.2. Choix de paramètres

Le premier paramètre testé correspond aux différentes façons d'identifier les triplets potentiels à partir des classes et propriétés annotées dans le corpus. Nous avons en effet au moins trois possibilités :

- Soit nous estimons que pour qu'un triplet D_p_R ¹⁸⁷ soit considéré comme présent dans un segment, il faut que D , p et R soient annotés. Cette possibilité est rapidement écartée car peu d'occurrences d'une telle configuration sont présentes dans le corpus.
- Soit nous estimons que D et p ou p et R suffisent.
- Soit nous estimons que D et R sont suffisantes, ce qui revient à considérer que si deux classes sont présentes dans le segment analysé, alors toutes les propriétés de l'ontologie qui les unissent sont possibles (cela revient également à dire que l'identification des propriétés par le lexique est inutile).

D'autre part, le Concept et l'utilisateur sont centraux dans un nombre important de règles. Cependant, comme nous l'avons déjà expliqué, la résolution de l'anaphore sur tout le texte est compliquée pour ce qui concerne le Concept. Quant à l'utilisateur, il est identifié dans le texte par des pronoms et déterminants de première personne. Or, il est loin d'être évident que ces indices soient systématiquement présents. Nous décidons donc de tester la possibilité que les classes *Concept* et *User* soient considérées comme présentes par défaut. Si ce choix ne crée pas de bruit excessif dans les résultats, il pourrait s'avérer d'une grande aide en ce qu'il apporterait une solution – au moins temporaire – à ces besoins de repérage.

Enfin, nous nous posons la question de la longueur du segment à utiliser. En effet, si notre corpus est composé de verbatim de taille variable pré-extraits par des analystes, le système final devra faire des choix de segmentation. Nous décidons donc de segmenter chaque verbatim du corpus en phrases, et d'attribuer à chaque phrase le Critère du verbatim dont elle provient. Par exemple, le verbatim

« Premier abord, comme ça, c'est un produit top. Mais par contre, c'est pour l'industrie. C'est pour l'industrie, hein. Je ne

¹⁸⁷ Pour rappel, la classe dont part la propriété orienté est appelé *Domain* et celle vers laquelle elle pointe est appelée *Range*.

vois pas. Dans notre domaine, dans le bâtiment... où je suis passé peut-être à côté de quelque chose, mais... Parce qu'en fait, nous, tout est piloté par de la régulation. »

attribué au Critère 3.1 – « *L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?* », est segmenté en sept phrases toutes attribuées à ce même Critère. Dès lors, il est évident que ce choix implique une baisse notable de la fiabilité du corpus ; si ce verbatim considéré comme unité autonome est bien à attribuer au Critère 3.1, ce n'est pas le cas pour chaque phrase prise individuellement. « *Premier abord, comme ça, c'est un produit top.* » est lié à la réaction spontanée, qui correspond au Critère 1.1. « *Mais par contre, c'est pour l'industrie.* » et « *C'est pour l'industrie, hein.* » correspondent bien au Critère 3.1. « *Je ne vois pas.* » se réfère à la compréhension et l'on pourrait donc considérer soit que ce verbatim est à classer dans le Niveau 1, soit qu'il n'est pas assez pertinent pour être classifié. « *Dans notre domaine, dans le bâtiment...* » pointe vers le Critère 3.1 mais est incomplet. « *où je suis passé peut-être à côté de quelque chose, mais...* » se rapporte également à la compréhension. Enfin, « *Parce qu'en fait, nous, tout est piloté par de la régulation.* » seul ne permet pas une classification dans la mesure où il n'est pas évident qu'elle consiste en une remarque portant sur l'adéquation entre le Concept et les pratiques, même si l'on peut le supposer. Nous avons choisi ce verbatim car il illustre bien le problème que nous soulignons, mais ce phénomène n'est bien sûr pas aussi marqué sur l'ensemble du corpus. Partant du principe que nous ne souhaitons pas substituer notre jugement à celui des experts, nous choisissons de ne pas réétiqueter chaque phrase, au risque de rendre tout le mérite du filtrage expert du corpus caduque. Nous utilisons donc ce nouveau corpus segmenté en phrases, tout en gardant à l'esprit qu'il est porteur d'une imprécision certaine.

Précisons que, dans tous ces tests, le classifieur à base de règles créé est multilabel, afin, encore une fois, de nous rapprocher du fonctionnement final attendu. Ce choix a bien entendu l'inconvénient de compliquer la comparaison entre les classes cibles et les classes proposées par l'algorithme, mais nous ne voulons pas, à ce stade tout du moins, imposer de choix entre différentes classes qui seraient nécessairement relativement arbitraires.

5.1.3. Résultats

Algorithme	Segmentation	Précision	Rappel
Triplet activé si deux Classes	Verbatim	0,1	0,09
Triplet activé si deux Classes + User et Concept par défaut	Verbatim	0,1	0,1
Triplet activé si deux Classes + ou si Domain et Property	Verbatim	0,1	0,09
Triplet activé si deux Classes + ou si Property et Range	Verbatim	0,1	0,09
Triplet activé si deux Classes	Phrases	0,11	0,06
Triplet activé si deux Classes + User et Concept par défaut	Phrases	0,14	0,09

TABLEAU 27 : Résultats des tests de paramètres algorithmiques pour le système de règles basé sur les triplets de l'ontologie

Les scores obtenus pour différentes configurations algorithmiques qui implémentent ces possibilités de paramétrage sont présentés dans le TABLEAU 27. Ces scores sont bas, et les écarts sont légers, mais il apparaît que la meilleure alternative est d'utiliser une segmentation en phrases, un système d'identification des triplets qui n'utilise que les classes, et qui considère que les classes *Concept* et *User* sont par défaut présentes dans ces phrases.

5.1.4. Conclusions

La première conclusion que l'on peut tirer, à l'issue de la réalisation du système de classification par les triplets de l'ontologie, est que les scores obtenus sont assez faibles sur des données réelles. Cependant, cette faiblesse de score est à relativiser. La première raison est que nous disposons à ce stade de peu de règles et d'un lexique très restreint, qui ne permet donc pas de traiter une variété importante de données. D'autre part, nous comparons un corpus étiqueté en monolabel à des sorties multilabel. Or, certaines propositions de sortie d'algorithme peuvent constituer des suggestions de classification secondaires à celle annotée qui sont en réalité pertinentes. De plus, le fait que des scores légèrement meilleurs soient obtenus à partir de phrases plutôt que des verbatim entiers, malgré les problèmes de fiabilité évoqués au point 5.1.2, est plutôt encourageant. En effet, cela est relativement surprenant, dans la mesure où nous aurions pu imaginer que la quantité de phrases impossibles à classifier extraites des verbatim ne fasse baisser le score. Or, malgré la présence de phrases inclassifiables ou dont la classification n'est pas pertinente dont on attend de l'algorithme qu'il les classifie, les performances sont tout de même meilleures. Une observation des sorties confirme d'ailleurs que, sur la base seule des phrases, certaines classifications évaluées comme fausses seraient en fait potentiellement recevables. Cela nous encourage donc à choisir de travailler à l'échelle de la phrase, même si cela reste un choix arbitraire imposé aux analystes.

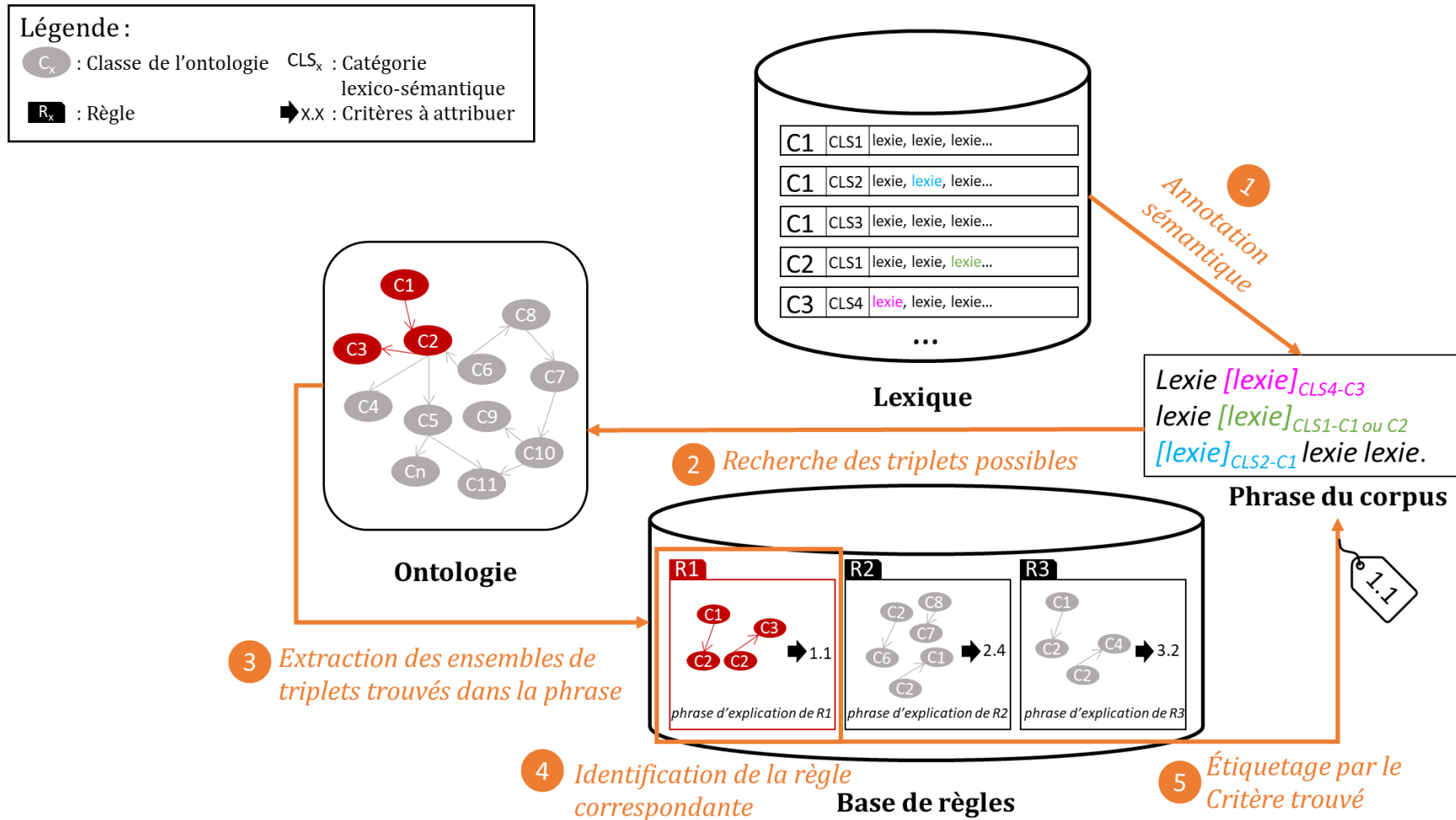


FIGURE 15 : Fonctionnement du système proposé

D'autre part, étant donnés les faibles écarts sur lesquels reposent les choix algorithmiques que nous effectuons, nous estimons que ces choix ne sont pas nécessairement définitifs. Pour cette raison, bien que le choix opéré rende caduque la nécessité de l'étiquetage des propriétés et des classes *Concept* et *User*, nous les conservons au cas où nous décidions finalement de les utiliser par la suite.

Nous obtenons donc un système opérationnel, dont l'architecture est présentée dans la [FIGURE EN ANNEXE 4](#) (p. 268) et dont la [FIGURE 15](#) propose une illustration du fonctionnement. Ce système ne permet pas en l'état d'obtenir des performances satisfaisantes, mais nous estimons que la taille du modèle ainsi que les différents problèmes que nous venons d'évoquer y concourent. Nous décidons alors d'augmenter le modèle et d'affiner les règles.

5.2. Apprentissage sur corpus et harmonisation des règles

Nous présentons dans ce point la manière dont nous avons créé de nouvelles règles à partir du corpus et les avons ensuite harmonisées. Nous commençons par expliquer le principe général de la démarche, puis celui plus précis de la méthode d'harmonisation à travers la réalisation de ce que nous appelons « tableaux conceptuels ».

5.2.1. Principe général

A la mise en place du système que nous venons de décrire succède un processus manuel d'augmentation et d'ajustements du modèle. Sur notre corpus segmenté en phrases, nous appliquons notre système, repérons quelques erreurs, corrigeons notre modèle en ajoutons des classes, propriétés, entrées lexicales et règles. Nous répétons ce processus de manière itérative. Étant donné l'aspect extrêmement chronophage de ce travail, ainsi que de la nécessité de conserver une quantité suffisante de données de test, nous ne cherchons pas à « corriger » chaque sortie obtenue, mais plutôt à identifier les erreurs récurrentes et à mesurer l'impact de chaque choix d'ajout ou de suppression que nous suggère un cas particulier sur l'ensemble du corpus.

Nous identifions ainsi des manques à combler dans le modèle, mais aussi des ambiguïtés lexicales récurrentes que nous tentons de résoudre. L'ajout itératif de règles produit d'autre part des doublons et des incohérences. Afin de résoudre ce problème, et de construire un système qui ne soit pas un ensemble décousu de règles en tout genre,

nous décidons, après avoir conséquemment augmenté le modèle, d'effectuer un travail d'harmonisation.

5.2.2. Création de tableaux conceptuels par Critères

Afin de procéder à cette harmonisation, nous prenons le parti de créer des « tableaux conceptuels » pour chaque Critère. Afin d'expliquer leur fonctionnement, nous utiliserons l'exemple de celui réalisé pour le Critère 1.2 – « *Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?* », représenté dans la FIGURE 16. Par souci de lisibilité, cette figure est segmentée pour s'étaler sur deux pages.

La volonté de construire ces tableaux naît du constat du fait que la majorité de nos règles mettent en jeu une des classes « primaires » (i.e. taxonomiquement positionnées immédiatement sous la racine de l'ontologie) ou « secondaires » (sous-classes d'une classe primaire) ainsi qu'une réaction du participant représentée par une sous-classe de *ReactionToConcept*. Nous pouvons ainsi généraliser – même si ce n'est pas exact dans certains cas particuliers – qu'une règle de classification consiste en l'association entre un type de réaction et un sujet qui suscite/sur lequel porte cette réaction. Il nous a donc paru opportun de réaliser un inventaire, qui se veut tendre vers l'exhaustivité, des combinaisons types de réaction-sujet associées à chaque Critère. Cet inventaire se matérialise donc sous la forme de ces tableaux conceptuels.

Dans l'exemple présenté, les colonnes *Sujet* et *Sous-sujet* incarnent respectivement les classes secondaires et primaires. La colonne *Réaction* indique la sous-classe de *ReactionToConcept* concernée. La colonne *Phrases* renseigne la phrase d'explication associée à la règle induite par cette combinaison, ainsi qu'un exemple de verbatim qui y correspond. Ainsi, la deuxième ligne du tableau est interprétée comme : « Dans le Critère 1.2, on classe les verbatim dans lesquels on trouve une réaction de compréhension et qui traite du sujet « composants *hardware* » qui est un sous sujet de « composants ». L'interprétation de ce verbatim est : *L'utilisateur s'exprime sur sa compréhension des composants hardware du Concept* et un exemple de verbatim qui y correspond est « *J'ai compris qu'il y avait un terminal dédié* ». ». Bien entendu, chacune des lignes est traduite en un ensemble de triplets qui constituent une règle. Ici, il s'agit de :

User_hasReaction_Understanding
+ *Understanding_isAbout_HardwareComponents*

1.2 - Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?

Sujet (hors concept)	Sous-sujet	Réaction	Phrases	
Fonctions		Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction du concept. <i>Je ne comprends pas cette fonction.</i>	
		Question	L'utilisateur pose une question sur une fonction du concept. <i>C'est une fonction qui est intégrée ?</i>	
	AccessToComponents Fonctions	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction d'accès à un composant du concept. <i>Je ne comprends pas bien si on peut accéder aux capteurs.</i>	
		Question	L'utilisateur pose une question sur une fonction d'accès à un composant du concept. <i>Du coup, je peux accéder à l'appareil photo depuis l'interface ?</i>	
	CentralizationFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction de centralisation du concept. <i>Je comprends que ça permet de regrouper les fonctions de deux de mes logiciels dans une seule application.</i>	
		Question	L'utilisateur pose une question sur une fonction de centralisation du concept. <i>Mais est-ce que ça permet de tout agréger au même endroit ?</i>	
	DataProcessingFunction	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction de traitement des données du concept. <i>Je comprends que ça analyse toutes les données par paquets.</i>	
		Question	L'utilisateur pose une question sur une fonction de traitement des données du concept. <i>Du coup, ça analyse les données en temps réel ?</i>	
	FeedbackFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction de feedback du concept. <i>Je comprends pas bien si ça me fait remonter toutes les infos ou certaines en particulier.</i>	
		Question	L'utilisateur pose une question sur une fonction de feedback du concept. <i>Du coup, ça vibre à chaque événement ?</i>	
	LinkingFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction de lien du concept. <i>Je comprends que ça fait le lien entre mon planning et celui de mes collègues ?</i>	
		Question	L'utilisateur pose une question sur une fonction de lien du concept. <i>Mais ça lie les données de tous les sites ?</i>	
	MotionFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension des fonctions de mouvement du concept. <i>J'ai compris que ça fermait les volets automatiquement.</i>	
		Question	L'utilisateur pose une question à propos d'une fonction de mouvement du concept. <i>Y a des scénarios où tout se met en route en même temps : mes volets se ferment ?</i>	
	RecordingFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction d'enregistrement du concept. <i>Je suis pas sûr d'avoir bien saisi si ça enregistrait tous mes parcours ou juste ceux pour lesquels j'active moi-même le tracking.</i>	
		Question	L'utilisateur pose une question sur une fonction de feedback du concept. <i>Ça enregistre toute la réunion ?</i>	
	UploadFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension d'une fonction d'upload du concept. <i>Je vois bien que ça permet de charger tous mes devis dans le logiciel en un seul clic.</i>	
		Question	L'utilisateur pose une question à propos d'une fonction d'upload du concept. <i>Je peux charger mes photo sur la plateforme ?</i>	
	Competition		Comparison	L'utilisateur fait une comparaison entre le concept et des produits/services concurrents . <i>Ça ressemble à ce que propose [Entreprise].</i>

Sujet (hors concept)	Sous-sujet	Réaction	Phrases
Components	HardwareComponents	Understanding	L'utilisateur s'exprime sur sa compréhension des composants hardware du concept. <i>J'ai compris qu'il y avait un terminal dédié.</i>
		Question	L'utilisateur pose une question sur les composants hardware du concept. <i>Est-ce qu'il y a un capteur ?</i>
	SoftwareComponents	Understanding	L'utilisateur s'exprime sur sa compréhension des composants software du concept. <i>Je comprends bien que c'est un logiciel spécifique.</i>
		Question	L'utilisateur pose une question sur les composants software du concept. <i>Du coup, c'est un logiciel spécifique ?</i>
	NetworkComponents	Understanding	L'utilisateur s'exprime sur sa compréhension des composants réseau du concept. <i>Du coup, si j'ai bien compris, ça communique soit par radio soit par Bluetooth.</i>
		Question	L'utilisateur pose une question sur les composants réseau du concept. <i>C'est de la radio ?</i>
ConceptFunctioning		Understanding	L'utilisateur s'exprime sur sa compréhension du fonctionnement du concept . <i>J'ai compris que ça marchait par itérations. / Ça fait boîte noire!</i>
		Question	L'utilisateur pose une question sur le fonctionnement du concept . <i>Mais là du coup, ça marche comment ?</i>
Use Requirements	MaterialTechnical UseRequirements	Understanding	L'utilisateur exprime sa compréhension des prérequis matériels nécessaires pour utiliser le concept. <i>J'ai compris qu'il y avait besoin d'un port USB.</i>
		Question	L'utilisateur pose une question sur les prérequis matériels nécessaires pour utiliser le concept. <i>Est-ce qu'il faudrait refaire toute l'installation électrique ?</i>
	NetworkTechnical UseRequirements	Understanding	L'utilisateur s'exprime sur sa compréhension des conditions préalables d'utilisation du concept liées aux réseaux . <i>J'ai compris qu'il fallait une connexion.</i>
		Question	L'utilisateur pose une question sur les prérequis de réseaux nécessaires pour utiliser le concept. <i>Ouais mais là, faut une connexion internet ?</i>
Components + Fonctions	HardwareComponents + FeedbackFonctions	Understanding	L'utilisateur s'exprime sur sa compréhension des composants hardware et des fonctions du concept. <i>Il y a un concentrateur qui vous permet d'informer chaque module que vous avez brancher de ce qu'il doit faire .</i>
Functions + Material Environment	DevicesMaterial Environment + BuildingMaterial Environment + LinkingFonctions	Opinion	L'utilisateur donne une opinion générale positive ou négative sur une fonction du concept qui consiste à lier des éléments de son environnement matériel . <i>C'est super intelligent de faire le lien entre ma maison et mon téléphone.</i>

FIGURE 16 : Exemple de tableau conceptuel

Réaliser ainsi cette cartographie permet tout d'abord de nous doter d'un moyen simple de partage du modèle avec l'équipe d'analystes, qui peuvent alors suggérer des modifications pour le modèle sans avoir à entrer dans ses méandres.

Bien entendu, cela nous permet surtout d'apporter de la cohérence à notre système de règles sur plusieurs points. Premièrement, nous pouvons limiter les doublons de règles à ceux qui sont réellement nécessaires et en éliminant ceux qui résultaient d'une erreur ou d'une surgénéralisation à partir d'un seul verbatim. D'autre part, cela nous permet également de combler des manques que les données n'avaient pas permis d'identifier mais dont il était évident qu'ils méritaient d'être traités. Par exemple, dans les lignes 2 à 6 du tableau donné en exemple dans la figure, les données du corpus nous avaient permis d'identifier la combinaison réaction-sujet « question-composants réseau ». Or, d'une part, il était évident que si les composants réseau étaient concernés, les composants *hardware* et *software* l'étaient aussi. D'autre part, nous avons identifié dans les données que ce Critère implique soit des questions soit l'expression d'une (in)compréhension. Nous pouvons donc créer toute la combinatoire présentée dans ces six lignes (compréhension-composants *hardware*, question-composants *hardware*, compréhension-composants *software*, etc.).

Ces tableaux conceptuels nous permettent donc d'apporter des réponses cohérentes et potentiellement finies à une des questions centrales de nos travaux : « Que doit contenir un verbatim pour qu'il soit classé dans tel ou tel Critère ? ». Cette question, par ailleurs, résume finalement au moins la moitié du problème à résoudre, la seconde moitié étant : « Comment faire pour identifier les contenus correspondants dans les textes ? ».

A l'issue de cette réorganisation conceptuelle, nous obtenons un total de 231 règles¹⁸⁸. Un déséquilibre certain dans le nombre de règles par Critère est observable ; certains n'en ont que très peu, soit parce que la variété de configurations trouvées dans les données était très faible, soit car peu de verbatim étaient attribués à ces Critères dans notre corpus. Quoi qu'il en soit, nous avons un nombre de règles que nous estimons – de manière tout à fait intuitive – suffisant pour prétendre à des résultats qui puissent être

¹⁸⁸ Nous en avons 288 à l'issue de l'expérimentation par patrons. Cette diminution est liée à deux facteurs. Premièrement, différentes règles par patrons peuvent correspondre à une seule combinaison de triplets. Deuxièmement, le processus d'harmonisation ici décrit a permis de supprimer des règles inutiles ou redondantes.

considérés comme significatifs. Nous présentons dans le prochain point ces résultats et quelques procédés d'optimisation de la classification que nous avons employés pour les améliorer.

5.3. Résultats et optimisations de la classification

Les règles ainsi réorganisées, nous réalisons une dernière évaluation de notre système sur notre corpus d'entraînement. Précisons qu'une étape d'augmentation du lexique, qui sera présentée au chapitre suivant, a été menée en parallèle de ces ajouts et optimisations des règles, et que les scores que nous présentons ici ont été obtenus non seulement par les modifications des règles mais aussi par ce lexique augmenté.

Nous obtenons donc, sur le corpus segmenté en phrases, une mesure de rappel $r=0,09$ et de précision $p=0,22$. Nous décidons d'ajouter deux règles à notre algorithme. La première consiste à considérer que, le Critère 1.1 étant un Critère général parfois attribué par défaut, un verbatim ne peut lui être attribué que si l'algorithme ne l'attribue à aucun autre Critère. D'autre part, si un verbatim n'est attribué à aucun Critère, nous l'attribuons au Critère 4.5 – « *Quel prix l'utilisateur consent-il à payer pour ce concept ?* » s'il contient €, euro, prix, coût, tarif ou somme. Ce Critère appelle en effet des verbatim courts de type « *100€ par mois* » qui sont des réponses directes à une question de l'interviewer. Ces deux modifications font augmenter le rappel de $+0,01$ et la précision de $+0,04$, ce qui nous laisse à penser que ce type de règles particulières peut en effet améliorer un peu les performances.

Nous avons ainsi un score final, sur notre corpus d'entraînement, de $r=0,10$ et $p=0,26$. Nous avons donc obtenu, par rapport au système mobilisant le jeu de règles initiales (TABLEAU 27, p. 203), une progression de rappel de $+0,01$ et de précision de $+0,12$. Les scores obtenus demeurent bien en-dessous des attentes. Cependant, comme précisé plus haut, le corpus utilisé présente des inconvénients certains pour l'évaluation de ces résultats. D'autre part, nous rappelons que le système est entraîné sur des données orales transcrites mais doit pouvoir permettre d'analyse des *posts* issus d'internet. L'impact de cette différence dans les données n'est, à ce stade, pas évaluée¹⁸⁹. Elle le sera lors d'un test

¹⁸⁹ Principalement pour des raisons de calendrier ; les études menées en ligne sur la plateforme propriétaire *yomane* ayant été amorcées tardivement par rapport à la chronologie de notre thèse, nous n'avons pu disposer de données analysées collectées ainsi qu'à la fin de ces travaux.

en conditions réelles d'études qui sera présenté dans le prochain et dernier chapitre de cette thèse.

Ce chapitre a présenté comment le système basé sur des règles elles-mêmes basées sur une ontologie a été construit. Nous avons également exposé son architecture finale ainsi que les premiers résultats obtenus. La réalisation effective d'un tel système a permis de mettre en évidence les limites de notre corpus d'entraînement en termes d'évaluation. Dans le prochain chapitre, nous expliquons comment nous avons achevé de rendre le système potentiellement utilisable par les experts en complétant sa base lexicale et en le rendant accessible aux analystes sur une plateforme web afin d'initier son utilisation effective et de l'évaluer.

CHAPITRE IX : VERS UN SYSTÈME EN APPRENTISSAGE CONTINU

Ce dernier chapitre vise à expliquer comment nous entendons nous diriger vers un système en apprentissage continu à partir du modèle que nous avons construit. Cette volonté est conditionnée, premièrement, par l'augmentation du potentiel de généralisation de ce modèle grâce à un lexique plus important. Deuxièmement, ce modèle augmenté doit être mis à l'épreuve de données nouvelles dans un contexte d'étude réelle, afin d'évaluer ses performances et la possibilité, ou non, de l'utiliser de manière effective. Ces deux axes structurent ce qui suit ; dans un premier point, nous présentons la méthode employée pour augmenter le lexique, puis exposons dans un second une évaluation du modèle à travers son utilisation pour l'autocodage d'une étude. A l'issue du développement de ces deux axes, nous proposons dans un troisième point un bilan de l'expérience ainsi que des pistes d'améliorations futures envisagées.

6. Augmentation du lexique

Notre lexique a la forme présentée au point 2.2.4 du Chapitre VII (pp. 182-185). Les premières unités lexicales ont été ajoutées sur la base de celles que nous rencontrons dans le pseudo-corpus de verbatim stéréotypiques ainsi que sur le corpus d'apprentissage de données réelles d'études anciennes. Nous ajoutons de manière artisanale des synonymes évidents de ces unités au fur et à mesure que nous les traitons. Cette méthode n'a été employée que dans le but d'amorcer la mise en œuvre de notre système et nous devons donc explorer d'autres méthodes aboutissant à une augmentation plus conséquente et plus systématique. Notre corpus étant de faible taille, nous avons préféré avoir recours à des ressources lexicales plutôt qu'à une méthode uniquement basée sur un apprentissage. Les travaux que nous exposons dans ce point ont dans l'immense majorité été réalisés par Alice Adnot dans le cadre de son stage de fin d'études effectué au sein de l'entreprise. Nous décrivons dans la suite le processus suivi et les résultats obtenus.

6.1. Méthodologie

Nous nommons « UL existantes » les unités lexicales que notre lexique contient avant l'augmentation et « UL cibles » celles que ce procédé doit être en mesure d'identifier pour les y ajouter. D'autre part, au moment où nous initions ce procédé, le lexique est composé de 324 entrées pour 3 816 UL existantes qui avaient été ajoutées à la main à partir du corpus et par consultation de dictionnaires de synonymes en ligne. Pour rappel, une entrée du lexique correspond à une association entre une catégorie lexico-sémantique et une classe ou propriété à laquelle est assignée une liste d'unités lexicales de plusieurs types (lemmes, formes, chaînes de lemmes, patrons de lemmes et catégories morphosyntaxiques). Par exemple, une des entrées du lexique consiste en l'association entre la catégorie lexico-sémantique *NomEntourage* et la classe *UserRelatives* et contient, entre autres, les lemmes *proche* et *cousin*, la forme *fil*, et la chaîne de lemmes *membre de mon famille*.

Afin d'exposer la méthodologie suivie, nous commençons par présenter les méthodes particulières d'augmentation du lexique que nous envisageons. Nous faisons ensuite un point sur le choix de ressources qui nous avons effectué, puis précisons les détails de la mise en œuvre du processus d'augmentation.

6.1.1. Identification des modalités d'extension

Pour chaque entrée du lexique, nous commençons par identifier quelle est *a priori* la modalité d'extension la plus pertinente. Nous entendons par « modalité d'extension » la méthode qui permet, à partir des UL existantes dans les entrées, d'identifier des UL cibles. Puis nous choisissons, parmi les UL existantes, celle que nous considérons être la plus représentative de l'entrée et qui sera le « déclencheur de l'extension ». Bien souvent, le déclencheur correspond à l'UL qui apparaît dans le nom de la catégorie – le déclencheur, pour la catégorie *AjdStupide* est *stupide* – mais cela n'est pas toujours le cas. Nous répertorions ainsi cinq modalités d'extension:

- **L'extension par synonymie** est la modalité la plus évidente et la plus fréquemment nécessaire que nous identifions (215/324 entrées). Elle consiste, tout simplement, à partir d'une UL déclencheur, à identifier ses synonymes. Par exemple, pour la catégorie *NomParcoursPro* associée à la classe *UserProfessionalPath*, il s'agira d'identifier les synonymes du déclencheur *carrière*.

- **L'extension par hyponymie** consiste à identifier tous les hyponymes du déclencheur. C'est le cas par exemple pour la catégorie *NomHabitation*, associée à la classe *BuildingMaterialEnvironment*, et pour laquelle il faut extraire tous les hyponymes de *habitation*. Nous avons 14 entrées de ce type.
- **L'extension par synonymie à déclencheurs multiples** est similaire à l'extension par synonymie, mais nous identifions plusieurs déclencheurs car nous estimons qu'un seul ne suffira pas à couvrir les besoins de l'extension. Par exemple, pour la classe *NomCatégorieAge* associée à la classe *TargetMarketPersonalProfile*, nous utilisons plusieurs déclencheurs comme *adolescent*, *adulte*, *vieux* ou *bébé*, afin d'identifier les synonymes de chacun d'entre eux. Nous avons 26 entrées de ce type.
- **L'extension par combinatoire** concerne les entrées qui appellent naturellement des UL polylexicales, comme c'est le cas pour *GnChosesEquivalentes* qui est associé à la classe *CompetitiveProductsServices*. Cette entrée doit comporter des UL de type *solution existante*, *produit équivalent* ou *chose similaire*. Nous devons donc lister plusieurs noms pouvant occuper la première place du GN (*chose*, *solution*, *truc*, *système*) et plusieurs adjectifs pouvant en occuper la seconde (*existant*, *similaire*, *équivalent*). Nous devons ensuite obtenir une liste de synonymes pour chacun de ces mots et reconstruire tous les GN possibles. Notons par ailleurs que, dans l'exemple présenté, les adjectifs peuvent aussi être remplacés par des relatives (*que je connais*, *qui existe déjà*) que nous devons également prendre en compte dans la combinatoire. Nous avons 45 entrées de ce type.
- **L'extension par ressource spécialisée** est réservée aux trois entrées pour lesquelles nous n'envisageons pas qu'une ressource lexicale générale puisse offrir une couverture suffisante. Il s'agit d'entrées destinées à lister des noms de composants ou actions techniques (*disjoncteur*, *réarmement*, *concentrateur*, etc.). Ces entrées n'étant qu'au nombre de trois, nous choisissons de ne pas les traiter pour le moment.

Nous avons donc identifié quatre modalités d'extension pour augmenter le lexique de 303 des 324 entrées. Les 21 restantes n'ont pas besoin d'être augmentées car il s'agit de catégories qui appellent des ensembles de mots grammaticaux très restreints (adjectifs

possessifs, pronoms COD ou COI, etc.). Nos besoins ainsi précisés, nous pouvons choisir une ressource lexicale adaptée.

6.1.2. Choix de ressources

Parmi les ressources lexicales qui proposent une structuration du lexique en relations de synonymie et hyperonymie, nous retenons comme possibles JeuxDeMots (Lafourcade & Joubert, 2008), WoNef (Pradet et al., 2014) et WOLF (Sagot & Fišer, 2008). La première a été construite à partir d'un jeu sérieux en ligne, les deux suivantes sont des adaptations françaises de WordNet (Miller, 1995).

Afin d'évaluer le potentiel de couverture et de précision de ces trois ressources, nous effectuons un test préliminaire sur un échantillon d'entrées. La modalité d'extension par synonymie étant celle qui a été identifiée comme la plus importante, nous nous concentrons sur celle-ci. Nous utilisons 14 entrées, qui ont pour catégories lexico-sémantiques : *AdjAgréable*, *AdjStupide*, *AdvActuellement*, *AdvRapidité*, *VbEvoquer*, *VbCraindre*, *NomCible*, *NomCompétences*, *AdjFacile*, *AdjCoûteux*, *AdvDavantage*, *AdvDifféremment*, *VbAbandonner* et *VbCimenterConsolider*. A partir des déclencheurs de ces différentes entrées, nous extrayons des UL cibles candidates à partir des trois ressources. Nous séparons dans les sorties les UL monolexicales et polylexicales afin de faciliter les classements dans nos différents types d'unités lexicales¹⁹⁰.

Une étape de filtrage manuel est ensuite réalisée. Pour aider ce filtrage, nous créons pour chaque entrée une phrase d'aide à la désambiguïsation qui contient le mot déclencheur. Cette dernière permet de préciser quelle acception de ce mot est recherchée. Par exemple, pour *VbAbandonner*, nous créons la phrase « Je vais pouvoir abandonner mes cahiers. ». En cas de doute sur validité d'une UL candidate, nous pouvons donc effectuer un test de substitution du déclencheur *abandonner* par cette UL candidate ; *laisser tomber* sera retenu car le sens de « Je vais pouvoir laisser tomber mes cahiers. » est identique à la phrase d'exemple, *abdiquer* sera éliminé car *« Je vais pouvoir abdiquer mes cahiers. » est incorrect, et *déposer* est également éliminé car « Je vais pouvoir déposer mes cahiers. » n'a pas un sens assez proche de celui de la phrase d'aide à la désambiguïsation. Bien entendu, cet exercice comporte son lot de difficultés et de choix délicats. Il nécessite également de devoir toujours prendre en considération le contexte plus global du type de

¹⁹⁰ Lemmes, formes, chaînes de lemmes et patrons de lemmes et catégories morphosyntaxiques.

données que nous traitons, de ce que l'analyse cherche à identifier, ainsi que de la classe à laquelle est reliée la catégorie lexico-sémantique. Par exemple, le candidat *balancer* est retenu car nous cherchons ici à identifier toutes les lexies qui peuvent dénoter le fait qu'un utilisateur puisse supprimer un objet ou une pratique dans ses usages grâce au Concept. Dans un autre contexte, il eut été possible de considérer que *balancer* n'était pas à retenir comme synonyme sur la base de cette seule phrase d'exemple. D'autre part, pour aider davantage à la désambiguïsation, nous ajoutons également, lorsque nous le jugeons nécessaire, des indications qui précisent les emplois possibles des mots recherchés. Par exemple, *AdjStupide* est relié à la classe *Opinion* – qui désigne un jugement général sur le Concept – et doit servir à identifier des verbatim tels que « *c'est stupide votre truc* ». Nous indiquons que les adjectifs retenus doivent pouvoir qualifier des objets, et que nous ne retenons pas ceux qui ne peuvent porter que sur des noms d'êtres animés. Nous pouvons ainsi éliminer *sans cervelle*. Précisons également que, pour cette étape de test préliminaire, nous ne filtrons pas les UL cibles candidates que nous avons déjà dans notre lexique en tant qu'UL existantes, afin d'obtenir une idée du potentiel de chaque ressource sans inclure de biais aléatoire. Les UL cibles retenues pour chaque ressource à l'issue de ce filtrage sont consultables dans les [TABLEAUX EN ANNEXE 20 à 22](#) (pp. 331-335).

Une fois les données filtrées, chaque ressource est évaluée en termes de couverture et de précision. Nous entendons par couverture le nombre d'UL cibles correctes que chaque ressource a permis d'identifier. Il s'agit donc d'évaluer quelle ressource est la plus susceptible d'enrichir notre lexique avec un grand nombre d'UL cibles pertinentes. Les résultats obtenus sont présentés dans le [TABLEAU 28](#). La précision désigne le nombre d'UL cibles correctes par rapport à toutes celles que la ressource a fournies. Il s'agit donc de prendre la mesure de la quantité de filtrage manuel que nécessiterait chacune des ressources. Ces résultats sont exposés dans le [TABLEAU 29](#).

Il en ressort que *JeuxDeMots* est la ressource la plus à même de satisfaire nos besoins. Elle offre en effet un potentiel de couverture environ deux fois supérieur à *WoNef* et quatre fois supérieur à *WOLF*. Cette couverture est meilleure concernant les quatre catégories morphosyntaxiques considérées et concerne – à deux exceptions près – aussi bien les UL cibles monolexicales que polylexicales. Sa précision est également globalement meilleure que celles des deux autres ressources, mais cela est principalement dû un bon score atteint sur les adjectifs. Pour les autres catégories, la précision est

globalement inférieure à celle de WoNef et WOLF. Au vu de ces résultats, nous choisissons d'utiliser JeuxDeMots pour augmenter notre lexique car, même s'il est susceptible d'entraîner la nécessité d'un filtrage important sur certaines catégories, le potentiel d'enrichissement qu'il propose est bien meilleur. Nous trouvons d'autre part que la richesse des expressions (semi-)figées extraites, en particulier pour les adverbes de rapidité (*illico presto, aussi sec, en deux coups de cuillère à pot*, etc.), est prometteuse. Nous observons également qu'il offre une variété de registre de langue plus importante avec davantage de vocabulaire familier (*viteuf, avoir la trouille, bateau* ('facile'), etc.) mais aussi soutenu (*dispendieux, diligemment*, etc.).

	WoNef	Wolf	JeuxDeMots
Adjectifs	39	15	65
- dont monolexicaux	33	15	62
- dont polylexicaux	6	0	3
Adverbes	16	14	67
- dont monolexicaux	9	10	22
- dont polylexicaux	7	4	45
Noms	6	4	7
- dont monolexicaux	5	4	7
- dont polylexicaux	1	0	0
Verbes	31	12	60
- dont monolexicaux	23	12	51
- dont polylexicaux	8	0	9
Total	92	45	199

TABLEAU 28 : Evaluation du potentiel de couverture des trois ressources lexicales envisagées pour augmenter le lexique

	WoNef	Wolf	JeuxDeMots
Adjectifs	0,73	0,46	0,75
- dont monolexicaux	0,74	0,46	0,74
- dont polylexicaux	0,60	-	0,85
Adverbes	0,43	0,39	0,27
- dont monolexicaux	0,36	0,47	0,42
- dont polylexicaux	0,50	0,00	0,17
Noms	0,86	0,60	0,81
- dont monolexicaux	0,88	0,60	0,81
- dont polylexicaux	0,00	-	-
Verbes	0,82	0,87	0,76
- dont monolexicaux	0,84	0,86	0,74
- dont polylexicaux	0,77	1,00	0,85
Global	0,24	0,29	0,31

TABLEAU 29 : Evaluation du potentiel de précision des trois ressources lexicales envisagées pour augmenter le lexique

6.1.3. Mise en œuvre

La ressource ainsi choisie, nous pouvons procéder à l'enrichissement du lexique à partir de cette dernière. Le procédé utilisé pour les modalités d'extension par synonymie et synonymie à déclencheurs multiples est celle que nous venons de décrire pour la phase d'évaluation du potentiel des ressources (extraction à partir du déclencheur et filtrage manuel à partir d'éléments aidant à la désambiguïsation). La démarche est également identique pour ce qui concerne la modalité par hyponymie, la seule différence étant que nous nous servons de cette relation pour identifier des UL cibles candidates dans JeuxDeMots. Certaines entrées donnant lieu à peu de résultats, elles sont complétées par des listes externes trouvées sur le Web (les noms d'animaux et de pays par exemple). La modalité par combinatoire est réalisée telle que décrite dans le point 6.1.1, en utilisant JeuxDeMots pour identifier les synonymes candidats pour chaque partie de l'unité polylexicale que nous souhaitons construire.

6.2. Résultats

En termes de volume, la différence entre le lexique construit manuellement (V1) et celui qui a été augmenté selon les modalités que nous venons de décrire (V2) est présentée dans le TABLEAU 30. La taille du lexique a donc été augmentée de 318% au total avec – sans grande surprise – des augmentations moins importantes sur les patrons de lemmes et de catégories, très spécifiques et augmentées par combinatoire.

	Lemmes	Formes	Chaînes de lemmes	Patrons de lemmes et de catégories	Total
V1	1 725	91	1 177	823	3 816
V2	7 777	516	6 349	1 327	15 969
Augmentation	+6 052	+425	+5 172	+504	+12 153

TABLEAU 30 : Volumétrie de l'augmentation du lexique

Pour évaluer l'impact de cette augmentation du lexique sur le potentiel de repérage d'éléments d'intérêt dans les textes, nous utilisons notre corpus de verbatim et comptons combien d'unités lexicales sont étiquetées avec le lexique V1 et le lexique V2. Nous en avons 13 704 avec V1 et 15 213 avec V2. Cela correspond à 11,01% d'unités repérées en plus avec le passage au lexique augmenté. Cette augmentation est moins importante que nous aurions pu l'espérer en triplant le nombre d'unités lexicales du

lexique. Cependant, la V1 ayant été créée à partir de ce même corpus, il n'est pas étonnant que les unités lexicales d'intérêt qu'il contient avaient déjà en bonne partie été identifiées.

En termes de différences de résultats de classification du système global, la différence apportée par cette augmentation du lexique est très peu concluante sur notre corpus de verbatim ; si nous comparons les scores obtenus sur le corpus segmenté en phrases, pour un même jeu de règles, avec les lexiques V1 et V2, la progression du rappel n'est que de 0,01 et la précision reste inchangée. Cela est directement lié à la même raison que celle évoquée pour la couverture, à savoir qu'une majorité d'unités lexicales d'intérêt contenues dans le corpus avaient déjà été ajoutées manuellement au lexique. Une évaluation comparative entre ces deux lexiques sur un nouveau corpus, plus représentative de l'impact de cette évolution, sera présentée dans le point suivant.

Nous avons expliqué comment nous avons procédé à l'augmentation du lexique, ce qui constituait la dernière étape nécessaire à la réalisation d'une expérimentation en contexte réel d'analyse.

7. Expérimentation en conditions réelles

Malgré les scores peu encourageants obtenus jusqu'à présent, nous souhaitons mettre notre système à l'épreuve d'un usage en conditions réelles d'utilisation, sur un corpus constitué des textes d'une étude réelle conduite en ligne. Il s'agit en effet d'estimer ce que ce système est capable de réaliser sur des données issues d'internet et en nous affranchissant des problèmes d'évaluation liés à notre corpus d'entraînement. Nous avons donc conduit une expérimentation pour aller dans ce sens. Dans ce point, les conditions logicielles et l'étude traitée sont présentées avant d'introduire la méthodologie suivie et, enfin, les résultats obtenus et les conclusions que nous pouvons en tirer.

7.1. Conditions logicielles et projet test

Nous décrivons ici les conditions de réalisation de l'expérimentation qui ont été pensées pour se rapprocher au plus près de celles dans lesquelles seront menées les analyses CAUTIC® à l'aide du module d'analyse que nous réalisons. Nous présentons pour cela la plateforme d'analyse dans laquelle ce module a été intégré ainsi que l'étude de laquelle proviennent les données traitées.

7.1.1. Plateforme d'analyse

L'entreprise a développé une plateforme destinée à être le support de divers versants des analyses qualitatives, dont l'analyse CAUTIC®. Nous avons donc intégré le système que nous avons construit à celle-ci. Le lecteur trouvera des captures d'écran qui illustrent notre description de son fonctionnement et de ses interfaces dans les [FIGURES EN ANNEXE 5 et 6](#) (pp. 269-270).

Sur cette plateforme, l'analyse du contenu qualitatif d'une étude est réalisée sur plusieurs textes. Chacun des collaborateurs impliqué est chargé d'analyser un ou plusieurs textes, et plusieurs collaborateurs peuvent travailler en simultané sur les mêmes données. Un « texte » correspond à : une transcription d'entretien, d'un *focus group*, ou bien l'export des *posts* d'utilisateurs collectés lors d'une étude menée sur *yoomaeno*. Pour rappel, *yoomaeno* est une plateforme propriétaire sur laquelle les études en ligne sont menées¹⁹¹. L'ambition de l'entreprise est d'aller vers une utilisation majoritaire de ce mode de collecte et c'est donc celui auquel notre système doit être le mieux adapté.

Chaque texte est analysé par les modules¹⁹² de TAL qui prennent pour longueur de segment par défaut la phrase. Les analystes accèdent donc au texte entier pré-annoté depuis l'interface illustrée sur la [FIGURE EN ANNEXE 5](#). Cette interface est typique des systèmes de codage qualitatif. Elle propose une partie supérieure qui contient le texte que l'on peut surligner pour créer une nouvelle annotation. La partie inférieure affiche une liste de tous les verbatim du texte. Lorsqu'un verbatim a été extrait par notre système, le survol d'une icône d'information donne accès, dans une infobulle, à la phrase d'explication de la classification effectuée par l'algorithme¹⁹³. L'analyste peut donc choisir de conserver la classification, de la corriger, ou bien de supprimer le verbatim s'il le juge peu pertinent. Un code couleur dans le texte et une indication au-dessus de chaque verbatim permettent de différencier les verbatim 1) qui ont été annotés par l'algorithme¹⁹⁴ 2) qui ont été créés par un analyste, 3) qui ont été annotés par l'algorithme puis corrigés par un analyste, 4) qui ont été annotés par un analyste et corrigés par un autre analyste.

¹⁹¹ Cf. Partie I, Chapitre I, point 2.3.4, §2 (p. 16).

¹⁹² Le nôtre, ainsi que celui réalisé par notre collègue pour la classification en freins, motivations en conditions.

¹⁹³ « L'utilisateur s'exprime sur la facilité/difficulté d'utilisation du Concept » dans la capture d'écran. Ces phrases ont déjà été évoquées précédemment et l'on peut en trouver des exemples dans les colonnes « Phrase d'explication » des [TABLEAUX EN ANNEXE 18 et 19](#) (pp. 323-330).

¹⁹⁴ En orange sur la capture d'écran.

Une deuxième modalité de visualisation et classification des verbatim – présentée dans la [FIGURE EN ANNEXE 6](#) (p. 270) – permet d'accéder à tous les verbatim des différents textes de l'étude. Ils peuvent être filtrés selon plusieurs dimensions (Critères CAUTIC®, freins/motivations/conditions, origine de la classification, participant, etc.) afin que l'analyste puisse accéder directement aux catégories qui l'intéressent. Cette interface permet également de corriger les classifications.

7.1.2. Etude test

L'étude que nous utilisons pour notre test est réalisée sur *yoomaneo*. Il s'agit d'un projet qui évalue un Concept fictif que l'équipe a conçu pour faire une démonstration d'études réalisées sur cette plateforme de collecte. Elle a mobilisé une trentaine de participants. Hormis le fait que le Concept testé n'existe pas réellement, les modalités de présentation de ce Concept, les méthodes de collecte, d'analyse, et de production du livrable final sont identiques à celles déployées lors d'une étude réalisée pour un client. Les *posts* produits par les participants sont également de même nature et collectés dans la même quantité que lors d'études « réelles ». Il nous a alors paru opportun d'utiliser cette étude pour notre expérimentation pour deux raisons. La première est que, le Concept étant fictif – et donc porté par aucun client –, nous ne sommes pas contrainte par des dimensions de confidentialité qui entraveraient notre liberté de présenter, ici, les détails de ce Concept et des verbatim collectés à son propos. La deuxième raison est qu'aucune question d'analyse spécifique ne vient se superposer à l'analyse CAUTIC® ; dans le cadre de projets clients, il est courant qu'une ou plusieurs dimension(s) particulière(s) ai(en)t besoin d'être étudiée(s) en profondeur, ou encore qu'une question très spécifique au projet ne soit posée aux participants¹⁹⁵. N'étant pas attendu – et difficilement imaginable – que notre système soit en mesure d'aider à l'analyse d'éléments spécifiques à chaque projet n'entrant pas dans le cadre de CAUTIC®, cette étude « puriste » permettra d'évaluer au mieux les performances du système.

Le Concept fictif testé et appelé DOLOSPORT. Il s'agit d'une solution de vêtements connectés pour soulager les douleurs musculaires. Il a été présenté aux participants, lors de l'étude menée sur *yoomaneo*, sous forme d'une vidéo dont des captures d'écran sont

¹⁹⁵ Il est possible que le client demande, par exemple, à ce que soit étudiée la préférence des participants entre une commercialisation de son Concept sous forme d'achat ou d'abonnement, ou bien encore d'évaluer la légitimité qu'attribuent les participants à son entreprise sur le marché qu'il entend adresser avec ce Concept.

présentées dans la [FIGURE EN ANNEXE 21](#) (p. 336). Ce Concept et sa présentation ont été imaginés pour être représentatifs de ceux que l'entreprise est amenée à tester pour ses clients. On notera par exemple qu'un certain flou est laissé quant aux détails techniques des dispositifs utilisés par la solution ; aucune précision n'est fournie quant à la nature des capteurs sensoriels et, bien qu'une image de téléphone portable soit représentée, il n'est mentionné à aucun moment que la solution est ou non envisagée comme étant associée à une application mobile spécifique. Le même flou est laissé quant à la manière dont la solution va formuler ses recommandations pour corriger les postures. Ce manque de précision est souvent de mise dans les présentations des Concepts testés ; s'agissant de solutions innovantes en cours de développement, de nombreux éléments ne sont pas encore déterminés et l'étude est précisément réalisée pour permettre de prendre des décisions à leur propos. Nous pouvons également noter que l'avant dernière vignette de la [FIGURE EN ANNEXE 21](#) mentionne le fait que l'utilisation régulière de la solution permettrait d'obtenir des remises sur la mutuelle ; l'ajout de cette information au Concept a vocation à introduire un élément qui peut être perçu ou non, selon les sensibilités des participants, comme éthiquement dérangeant (aspect de surveillance, questions autour du partage des données, etc.). Cela est pensé pour que ce Concept donne matière, comme beaucoup d'innovations, à de potentielles réactions de questionnement ou même de rejet par les utilisateurs suscités par l'inadéquation de ce que propose ce Concept avec leurs valeurs personnelles.

La collecte des *posts* a donc eu lieu sur *yoomaneo* à partir de cette vidéo selon les modalités normales suivies pendant les études. Les participants doivent répondre à – en moyenne – une question par jour. Afin de combiner les mérites de l'expression individuelle des entretiens semi-directifs et ceux de l'échange d'opinions des *focus groups*, chaque participant doit dans un premier temps répondre aux questions de manière individuelle, sans voir les réponses des autres. Il les découvre après avoir posté la sienne et peut y réagir en les commentant ou en leur attribuant des points de pertinences. Les chargés d'études jouent le rôle d'animateurs et postent également des commentaires qui consistent souvent en des questions d'approfondissement (par exemple : « Pourriez-vous s'il vous plaît détailler les raisons pour lesquels vous trouvez que le Concept est inadapté à une pratique extérieure ? »). Les interventions des chargés d'études/animateurs ne sont pas formellement différenciées des commentaires des autres participants.

Les huit questions journalières qui ont été posées aux participants dans le cadre de cette étude sont consultables dans le [TABLEAU EN ANNEXE 23](#) (p. 337). On peut y voir que certaines questions sont en lien direct avec les Critères CAUTIC®, mais cela n'est pas le cas pour toutes. Un extrait des *posts* produits lors de la collecte est présenté dans l'[ENCADRÉ EN ANNEXE 8](#) (pp. 338-339). Nous y avons remplacé les pseudonymes des participants par des codes génériques. En outre, les participants peuvent associer des émoticônes à leur *posts*, d'où la présence de ces dernières dans l'extrait présenté. Le « 2★ » qui suit le pseudonyme d'un des participants signifie que son *post* a été évalué comme particulièrement pertinent par deux personnes. 193 *posts* (réponses aux questions journalières et commentaires confondus) ont été collectés au cours de cette étude.

7.2. Méthodologie

Le Concept étant principalement destiné à résoudre des problèmes de douleurs dorsales, les participants ont rempli un questionnaire leur demandant s'ils rencontraient ce genre de soucis. Cette question est initialement posée pour segmenter l'échantillon afin de réaliser une analyse comparative entre les deux populations. Nous utilisons donc cette variable pour diviser le texte en deux. L'un contient les réponses des participants ayant régulièrement mal au dos, et l'autre celles des participants dont ce n'est pas le cas. Nous les désignons respectivement par *MalDos* et *PasMalDos*. Nous soumettons donc ces textes à une analyse automatique sur la plateforme dédiée.

Nous souhaitons que notre évaluation permette d'évaluer la performance du système sur la base d'annotations réalisées par des analystes différents. Nous voulons également évaluer dans quelle mesure le niveau d'expertise des analystes éloigne ou rapproche leur codage de celui qui est réalisé par l'algorithme. Enfin, nous cherchons à identifier si le fait que l'annotation soit révisée par une seule ou plusieurs personnes a un impact sur les résultats. Nous demandons pour cela à quatre personnes de l'équipe de corriger les sorties de l'algorithme. Parmi elles, nous comptons deux analystes seniors (*Sen.1* et *Sen.2*) qui pratiquent la méthode CAUTIC® depuis plusieurs années, une analyste junior (*Jun.*) qui exerce le métier de chargée d'études depuis environ six mois, et une analyste novice (*Nov.*) qui occupe dans l'entreprise un poste qui n'implique pas de réaliser des analyses qualitatives et qui n'a donc jamais procédé à une analyse CAUTIC®. *Jun.* a été chargée d'analyser les deux textes avant notre expérimentation – pour des besoins indépendants de celle-ci – et est donc déjà familière avec les données. Nous demandons à

Sen.1 et *Jun.* d'analyser séparément le texte *MalDos*, qui est donc dupliqué. Nous demandons à *Nov.* d'analyser *PasMalDos* et à *Sen.2* de corriger ses annotations (i.e. d'ajouter une couche de corrections à celles que *Nov.* réalise déjà sur nos sorties). Cette répartition est résumée dans le TABLEAU 31.

même texte dupliqué

MalDos1	MalDos2	PasMalDos
Sen.1	Jun	Nov. puis Sen.2

TABLEAU 31 : Répartition textes/analystes pour l'expérimentation en conditions réelles

Une fois les analyses réalisées, nous comparons les sorties de l'algorithme avec les corrections réalisées. Les scores de rappel et de précisions obtenus et présentés au prochain point sont donc calculés en considérant chaque texte corrigé comme le référentiel de vérité auquel confronter les résultats de notre algorithme. En outre, précisons que nous sommes bien dans le cadre d'une tâche multilabel, car l'algorithme comme les analystes peuvent attribuer plusieurs Critères à un même verbatim.

7.3. Résultats

Nous présentons dans ce point les résultats en termes de performances traduites en scores, puis en entrant dans les textes par repérage des cas de réussite et d'échecs récurrents, et enfin en termes d'appréciations des analystes à travers leur retour sur expérience. A chaque étape, des pistes d'améliorations possibles du système sont proposées sur la base des éléments que cette analyse des résultats permet de mettre en exergue.

7.3.1. Scores obtenus

Le TABLEAU 32 présente les scores obtenus dans le cadre de la présente expérimentation. La colonne *PasMalDos_Nov* montre les scores obtenus sur la base de la correction par *Nov.* avant que cette correction ne soit à son tour revue par *Sen.2* ; sa présence sert seulement à illustrer la différence entre ces deux étapes et le *Total* n'en tient pas compte. D'autre part, les contenus de *MalDos1* et *MalDos2* étant identiques, nous n'utilisons pas *MalsDos2* non plus pour le calcul du total. Ce dernier est donc calculé par agrégation des annotations effectuées sur *MalDos1_Sen.1* et *PasMalDos_Nov.Sen.2*.

	MalDos1_Sen.1	MalDos2_Jun	PasMalDos_Nov.	PasMalDos_Nov.+Sen.2	Total
Précision	0,42	0,49	0,54	0,38	0,40
Rappel	0,38	0,42	0,49	0,28	0,33
F-score	0,40	0,45	0,51	0,32	0,36

TABLEAU 32 : Résultats pour l'expérience en conditions réelles

Nous obtenons ainsi au total une précision de $p=0,40$ et un rappel de $r=0,33$. Ces résultats ne sont pas très élevés, mais la différence avec ceux qui avaient été obtenus sur le corpus de verbatim d'études anciennes ($p=0,22$ et $r=0,09$) est significative. Ce corpus comportait, comme nous l'avons souligné, des inconvénients à être utilisé pour l'évaluation (segmentation en phrases non prévue initialement et étiquetage monolabel pour une évaluation multilabel). Cependant, nous pensons que la nature-même du contenu a joué un rôle. En effet, si la variation diamésique n'entraîne pas d'autres variations sociolinguistiques importantes¹⁹⁶, les réponses aux questions en ligne sont plus courtes et comportent moins de digressions que celles collectées à l'oral, ce qui a tendance à beaucoup restreindre la quantité de contenu non pertinent que notre système doit éviter.

Ces scores ne sont donc pas à la hauteur de nos attentes, mais ils doivent cependant être mis en perspective avec la complexité de la tâche : le système doit écarter le pertinent du non pertinent et traiter une problématique de classification multilabel sur 20 classes qui sont sémantiquement complexes. Les scores obtenus sur les quatre travaux d'autocodage que nous avons présentés dans le Chapitre III de la Partie II (point 8.2, pp. 60-72) étaient plus élevés. Cependant, les tâches et techniques étaient différentes. Crowston et al. (2012) obtenaient des scores de $p=0,68$ et $r=0,76$, sur des classes qui pouvaient être facilement identifiées par des indices linguistiques et typographiques plus simples que ceux que nous manipulons (*Apologies* et *Capitalization* par exemple). Les travaux de Kaufmann et al. (2020) résultaient en des scores de $p=0,3$ et $r=0,69$. Leur système était cependant basé sur des mots clés qui permettaient d'identifier des classes dont certaines peuvent être considérées comme purement thématiques (*People*, *IpRights*, *Project/Product*, *Education*) même si d'autres étaient plus complexes (*Day-to-Day issues*, *Trust*). Marathe & Toyama (2018) parvenaient à obtenir des scores de $p=0,88$ et $r=0,83$. Néanmoins, premièrement, certains des codes cibles étaient de type *in vivo*, c'est-à-dire qu'ils consistaient à étiqueter un segment avec des mots qu'il contenait. Deuxièmement,

¹⁹⁶ Cf. Chapitre I, point 4.2.3 (pp. 26-29).

parmi les approches qu'ils testaient, celle qui obtenait ces scores élevés était la plus laborieuse ; elle nécessitait de traduire chaque code en une requête complexe pour identifier des syntagmes d'intérêt qui permettaient, finalement, un simple repérage par mots clefs¹⁹⁷. Les deux autres méthodes testées, plus automatiques, ne permettaient d'obtenir que des scores bien plus bas ($p=0,15$ et $r=0,33$ pour l'une, $p=0,18$ et $r=0,43$ pour l'autre). Enfin, Rietz (2021) obtenait des performances de $p=0,92$ et $r=0,89$. Ces performances étaient cependant atteintes par ML seulement par le biais, en amont, d'une création manuelle de règles par les utilisateurs, et les codes, quoique plus complexes que ceux des trois précédents travaux, restaient plus simples que les nôtres¹⁹⁸.

D'autre part, étant donné que, sur une tâche similaire – classification multilabel avec élimination des inclassifiables – nous avons réussi à obtenir des scores meilleurs ($p=0,43$, $r=0,37$) avec un des modèles de ML sur notre corpus d'entraînement, nous nous demandons si le choix d'éliminer cette piste a finalement été judicieux. Nous appliquons donc le modèle de ML qui avait été identifié comme le plus prometteur sur ces nouvelles données. Nous obtenons des scores plus bas de $p=0,28$ et $r=0,24$. Il est fort probable que la différence de nature entre notre corpus d'entraînement et nos données cibles ait joué en la faveur de notre système par règles et aurait pénalisé un système de ML. En effet, en identifiant manuellement les caractéristiques importantes des textes et en les ramenant à un niveau conceptuel, nous avons fait en sorte que le modèle soit plus généralisable qu'il ne l'aurait été par apprentissage ML sans couche sémantique. Il n'est cependant pas exclu qu'à l'avenir ces deux approches soient utilisées de manière complémentaire. Il serait tout à fait possible d'imaginer, par exemple, que les phrases qui n'ont pas été classifiées par le système de règles soient soumises à un classifieur et que nous retenions les classifications sortantes qui dépassent un certain seuil de probabilité. Il est également possible de tester un système hybride qui utiliserait les informations sémantiques issues de la modélisation pour de la classification statistique. Les modalités exactes d'un tel système restent bien

¹⁹⁷ Pour rappel, il s'agissait par exemple, pour identifier des segments correspondant au code *Art research examples*, de produire la requête : *(video OR film OR animation OR (oral AND history) OR folktale OR poem OR poetic OR book OR anthology OR ((idea OR musical OR sketch) AND design) OR graphics) AND (project OR collection OR studio OR workshop OR performance OR (communication AND public))*. Cette requête complexe permettait par exemple d'attribuer le code *Art research examples* à un segment contenant *video project*.

¹⁹⁸ Il s'agissait par exemple du code *Motivation : Not Financial* qui était repéré sur la base des mots clefs *pay* et *money* accompagnés d'une négation par le système de règles.

entendu à définir, mais des modèles hybrides de ce type sont actuellement considérés comme prometteurs (Gomez-Perez et al., 2020).

En outre, nous observons une variabilité notable entre les scores obtenus sur les différents textes. Il est intéressant de constater que les analystes les moins expérimentés corrigent moins de sorties que les analystes seniors ; les différences de scores que nous obtenons entre la correction de *Nov.* et celle complétée ensuite par *Sen.2* est flagrante. Il serait prématuré à ce stade de tirer des conclusions définitives à partir de cette observation. Il est possible que davantage d'expertise confère une plus grande liberté interprétative aux analystes qui se permettent alors des écarts par rapport à la méthode, ou au contraire qu'ils l'appliquent avec une rigueur plus importante. Il est également possible que les analystes moins expérimentés se laissent plus facilement influencer par les choix de l'algorithme, ce qui constituerait alors un phénomène qui mériterait une certaine vigilance ; le système construit se veut, du point de vue pédagogique, être neutre ou bénéfique, mais en aucun cas il ne doit pousser vers une pratique de l'analyse qui incite à moins de recul. Nous retenons donc cette différence de taux de corrections par rapport au niveau d'expertise comme un point d'attention qu'il s'agira de continuer à mesurer et dont les raisons sont à éclaircir.

Nous pouvons par ailleurs observer le graphique présenté dans la FIGURE 17 pour nous intéresser à la répartition des scores selon les Critères. Les éléments en bleu représentent les attributions de classes qui ont été conservées par les analystes, en violet celles qui n'avaient pas été identifiées par l'algorithme, et en rouge celles qui avaient été effectuées par le système mais ont été supprimées par les analystes. Sans surprise, les Critères qui ont suscité le plus de classifications correctes (2.1 et 4.5) sont parmi ceux que nous avons identifiés comme ayant un vocabulaire stable lors de notre exploration textométrique¹⁹⁹. D'autre part, les verbatim du Niveau 4.5 sont des réponses à une question formulée directement dans l'étude (« Quel est le prix minimum en-dessous duquel la solution ne vous semblerait pas crédible et quel est le prix maximum au-dessus duquel la solution vous semblerait trop chère ? Merci d'indiquer une fourchette de prix. »). Mais, cela n'est pas le cas pour le Critère 2.1 – « *Le concept est-il simple à utiliser ?* »,

¹⁹⁹ cf. Chapitre V de la Partie II, point 3.2.2 (pp. 111-119).

ce qui signifie qu'il s'agit bien là d'expressions spontanées des participants à ce propos que le système a su identifier.

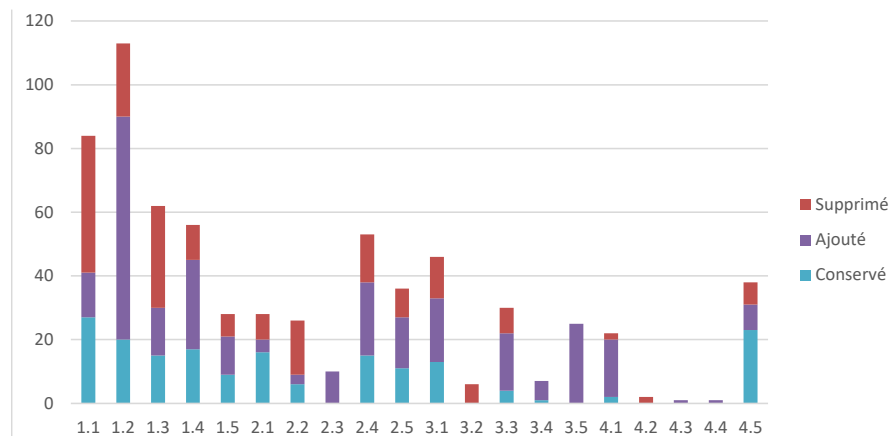


FIGURE 17 : Corrections effectuées par Critères

En outre, cette répartition des corrections selon les Critères permet d'identifier plusieurs phénomènes particuliers sur lesquels nous pourrions essayer d'agir par la suite et dont nous donnons ici quelques exemples. Premièrement, la majorité des classifications dans le Critère 1.1 ont été supprimées. Il semble donc que le choix que nous avons fait de limiter la classification dans ce Critère à des verbatim qui n'étaient attribués à aucun autre était judicieuse mais insuffisante. Deuxièmement le Critère 1.2 – « *Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?* » donne lieu à une couverture très limitée qu'il s'agira d'améliorer. Troisièmement, le Critère 3.5 – « *Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?* » n'est jamais identifié par l'algorithme, ce qui signifie que les règles qui lui sont associées ne sont pas assez précises et/ou nombreuses. Enfin, notons que les trois Critères 4.1, 4.2 et 4.3 sont extrêmement peu représentés, aussi bien dans les sorties de notre algorithme que dans les corrections des analystes. Ces derniers donnaient en effet lieu à peu de règles, dû notamment à cette même sous-représentation de ces mêmes Critères dans notre corpus d'apprentissage. Étant donné que ces Critères semblent peu utilisés, la question se pose de les conserver, à terme, dans notre système ; s'ils ne sont pas identifiés, la quantité d'identification manuelle est moindre pour les analystes, et la suppression de ces classes pourrait participer à augmenter le potentiel de précision du système. Cette décision pourrait être prise si ce phénomène venait à se répéter systématiquement sur toutes les études.

Nous jugeons qu'il est intéressant d'observer les différences de correction par Critères selon les textes et les analystes. Des graphiques similaires à celui que nous venons de commenter sont présentés, par texte, dans la [FIGURE EN ANNEXE 22](#) (p. 340). On peut y voir que répartition des verbatim par Critères est sensiblement identique d'un texte à l'autre et entre les analystes, tout comme la proportion de corrections des sorties de l'algorithme par Critère. Cela tend donc à indiquer que les résultats que nous obtenons ici, en termes de problématique par Critère à résoudre, ne sont pas rendus moins généralisables par une pratique de l'analyse potentiellement idiosyncratique de la part d'un analyste ou bien des particularités d'un texte. Il est en revanche possible que d'autres facteurs liés à l'étude entrent en jeu, et d'autres expérimentations sur des études différentes sont à mener. En outre, nous observons que les corrections réalisées par *Sen.1* et *Jun.* sur le même texte et à partir des mêmes sorties de l'algorithme sont extrêmement similaires. Cela tend à confirmer ce que nous avançons dans le point 2.3 (pp. 98-107) du Chapitre IV de la Partie II ; le manque d'accord entre les différents analystes qui avait été observé lors de l'annotation des verbatim hors contexte était bien majoritairement lié au fait, précisément, de ce manque de contexte, étant donné que lorsque deux analystes sont confrontés à des textes entiers, leur codage est bien plus similaire. Il n'en reste pas moins que ce paradigme d'analyse hors contexte est bien celui dans lequel s'inscrit notre algorithme, qui traite les corpus phrase par phrase sans pouvoir traiter tout ce qui relève d'un niveau d'informations pragmatique. D'autre part, et toujours à propos de cette problématique d'accord inter-juge, les nombreuses corrections apportées par l'analyse sénior *Sen.2* sur le codage réalisé par l'analyste novice *Nov.* – qui transparaissent à travers les chiffres présentés dans le [TABLEAU 32](#) – démontrent bien que la pratique de l'analyse CAUTIC® revêt un niveau de difficulté particulier même pour un être humain.

Enfin, la dernière évaluation quantitative que nous effectuons consiste à mesurer l'apport que la phase d'augmentation du lexique a constitué pour le système global, ce qui n'était pas possible à réaliser dans de bonnes conditions précédemment. Pour ce faire, nous utilisons le texte *MalDos_Sen.1* que nous faisons réanalyser à l'algorithme mais en nous servant du lexique non augmenté (V1). Nous comparons les sorties obtenues à celles corrigées pour par *Sen.1*. Avec le lexique V2, nous avons obtenu pour ce texte des scores de $p=0,42$ et $r=0,38$. Avec V1, nous aurions obtenu un rappel identique, mais pour une précision de $p=0,27$. Nous pouvons donc en conclure que cette étape d'augmentation du lexique a bien participé à améliorer le système, en faisant passer le f-score de 0,32 à 0,40.

En revanche, nous nous attendions à ce que cette augmentation ait un plus fort impact sur le rappel que sur la précision, ce qui est l'inverse de ce qui a été finalement observé. Il est donc fort probable que l'augmentation du lexique joue un rôle davantage centré sur la désambiguïsation, en ce qu'il aide à mieux étiqueter les classes de l'ontologie dans le texte, et que les leviers d'action dont nous disposons pour augmenter le rappel se situent davantage au niveau de la création, en amont, des catégories lexico-sémantiques. Cette supposition reste bien entendu à confirmer.

Ayant présenté une analyse quantitative des résultats, nous entrons à présent un peu plus dans le détail des annotations obtenues en nous intéressant aux régularités observés dans les réussites et erreurs de l'algorithme.

7.3.2. Réussites et erreurs récurrentes observées

Afin d'illustrer notre propos sur les réussites et erreurs que nous avons observées, nous présentons dans les [TABLEAUX EN ANNEXE 24 et 25](#) (pp. 341-344) des exemples sur lesquels nous nous appuyerons. Nous commençons par faire quelques commentaires sur les réussites. Ces dernières sont présentées dans le premier de ces deux tableaux. La première colonne contient le verbatim en question et la seconde le ou les Critère(s) correctement identifié(s). La quatrième colonne contient quant à elle la règle qui a permis la classification. Pour chaque règle sont fournis les deux éléments qui la caractérisent, à savoir une phrase d'explication et un ensemble de triplets de l'ontologie. Enfin, la dernière colonne – « Réussite description » – indique si, oui ou non, la phrase d'explication de la règle décrit bien le contenu du verbatim.

On peut voir dans ce tableau des analyses qui nous semblent particulièrement pertinentes au vu du contenu du verbatim. Par exemple le verbatim « *Il est vrai qu'elle peut apporter des informations intéressantes sur notre posture et les origines de nos douleurs.* » est correctement associé au Critère 1.4 – « *L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?* ». Cette classification a été opérée grâce à la règle qui identifie que « L'utilisateur exprime son intérêt pour une fonction de feedback du concept. ». D'autres classifications sont plus « hasardeuses » en ce que le bon Critère est attribué malgré le fait que l'interprétation proposée dans la phrase d'explication n'est pas correcte. Par exemple « *Il me paraît difficile de répondre affirmativement à cette question sans avoir essayé d'abord la solution.* » est correctement attribué au Critère 2.1 – « *Le concept est-il simple à utiliser ?* ». Cette classification est expliquée par la phrase

« L'utilisateur s'exprime sur la facilité/difficulté d'utilisation du concept. ». Or, cette phrase ne décrit pas ce qui est en train d'être dit dans le verbatim. En réalité, le système identifie *me paraît difficile de répondre* comme l'expression d'une difficulté d'utilisation, ce qui est faux. Cette bonne classification, si le système était plus rigoureux, n'aurait pas eu lieu. Enfin, quelques phrases de description sont plus ou moins correctes car peu précises. C'est le cas du verbatim « *J'utiliserais cette solution afin de me motiver à faire des activités sportives au quotidien, ou à des moments où la fatigue accentue mes douleurs musculaires.* ». Il est interprété comme « L'utilisateur se projette dans l'utilisation du concept pour accomplir une tâche. ». Cette interprétation réussit à saisir une projection dans l'usage associée à une tâche mais omet la précision d'un contexte d'utilisation.

Nous observons également que des classifications multilabel correctes sont possibles. C'est le cas du verbatim

« Ainsi, il serait intéressant que l'application puisse quand même donner des conseils et astuces personnalisés sur nos postures même en dehors des activités sportives avec par exemple des conseils pour les personnes qui travaillent devant un écran toute la journée, ou encore pour ceux qui sont souvent en déplacement en voiture... »

qui est à juste titre attribué aux Critères 3.1 – « *L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?* » et 1.4 – « *L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?* ».

Enfin, notre règle spécifique au Critère 4.5, qui consiste à repérer des mots clés appartenant au champ lexical des coûts, permet des classifications correctes telles que celle présentée dans la sixième ligne du tableau d'exemples.

Concernant les erreurs récurrentes, nous pouvons observer le deuxième tableau. Ce dernier comporte en première colonne le verbatim et en deuxième colonne la classification de l'algorithme. La troisième colonne indique la règle utilisée. La quatrième indique la correction de l'analyste. Enfin, dans la dernière colonne, nous ajoutons un commentaire d'explication de l'erreur.

La première ligne donne un exemple de bruit créé par le fait que l'algorithme a classifié une question de l'animatrice, ce qui n'intéresse pas l'analyste. Cette erreur est

assez fréquente. Comme nous l'avons expliqué, les *posts* des animateurs ne sont pas formellement différenciés de ceux des participants sur *yoomaneo*, et notre système les a donc analysées. Ce problème serait très facilement corrigeable ; nous pourrions tout simplement envisager de demander aux analystes de préciser dans une formulaire sur la plateforme d'analyse quels étaient les pseudonymes des animateurs afin d'exclure leurs *posts* des sorties.

La deuxième ligne démontre un problème de couverture du lexique. Le verbatim « *Elle est aussi en phase avec la tendance qui est de faire plus attention à soi et de faire plus d'activités sportives (notamment avec le covid et les confinements).* » n'est pas reconnu comme appartenant au Critère 4.1 – « *L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?* ». La règle qui stipule que si un participant perçoit une adéquation entre le Concept et les tendances aurait dû l'identifier. Cependant la locution *en phase avec* n'était pas dans notre lexique²⁰⁰, et la perception d'adéquation n'a donc pas pu être identifiée. Nous noterons par ailleurs que cette locution faisait partie de la question posée aux participants : « *En quoi la solution DOLOSPORT est-elle en phase ou non avec les tendances actuelles ?* ». Or, nombre d'entre eux ont répondu à cette question en reprenant cette expression, ce qui correspond à autant de silences dans nos sorties. Cette lacune peut être facilement comblée et nous permet d'identifier une opportunité d'amélioration du système. En effet, étant donné que les participants ont tendance à réutiliser mot pour mot des éléments des questions qui leur sont posées dans leurs réponses, si nous demandons aux analystes de stabiliser la formulation d'une même question sur toutes leurs études, nous aurions des éléments clairement identifiés à ajouter à notre lexique.

Nous trouvons en troisième ligne un cas de manque de contexte lié au fait qu'il s'agisse d'une réponse directe à une question : le verbatim « *Je suis intéressée par les deux premières.* » répond à « *Parmi ces trois fonctions de la solution [...] laquelle ou lesquelles vous intéressent le plus et pourquoi ?* ». Sur la base de ce seul verbatim, le système ne peut pas résoudre l'anaphore et déduire que *les deux premières* renvoient à des fonctions. Il serait donc possible, pour pallier ce phénomène, d'utiliser un système de résolution des anaphores pour ce cas spécifique. Nous pourrions alternativement avoir recours, à l'instar

²⁰⁰ Elle aurait dû faire partie de la catégorie lexico-sémantique existante *PrepEnAdequationAvec* qui contenait des unités lexicales telles que *en adéquation avec*, *en harmonie avec* ou *en cohérence avec*.

de Marathe & Toyama (2018), à une segmentation en « paragraphes contextuels » (cf. p. 65) qui associe chaque réponse à la question posée. La quatrième ligne expose également un manque de contexte et d'informations d'ordre pragmatique : « *Un signal sonore quand la position n'est pas bonne ?* » ne permet pas d'identifier que le *signal sonore* indiqué correspond à une suggestion pour améliorer le Concept. La formulation ne permet pas en effet à elle seule d'inférer avec certitude qu'il s'agit d'une suggestion, et cela est également impossible à définir si l'on n'est pas familier avec le Concept testé. De la même manière, il est impossible que le système identifie le fait que

« *Pour la troisième solution, hors de question de me faire fliquer par ma mutuelle. Et finalement, est-ce que ça ne sera pas ceux qui sont le plus cassé, qui finiront par payer le plus ?* »

mentionne une des fonctions du Concept. Il est d'ailleurs possible que des noms de fonctions spécifiques, en plus de n'être pas repérés comme tels, créent un bruit important. Cela n'a pas été le cas ici, mais *mutuelle* aurait pu être associée dans notre lexique à un type d'organismes ; de nombreuses classifications auraient alors été réalisées par compréhension du fait que le participant s'exprime à propos de l'adéquation entre le Concept et des organisations/entreprises. Pour pallier ce problème, nous envisageons d'ajouter un formulaire sur la plateforme qui demanderait aux participants d'indiquer les noms des différentes fonctions du Concept. Nous ferions alors une modification *ad hoc* du lexique, en associant ces lexies à la classe *Functions* tout en les retirant des autres entrées. Cela permettrait de mieux repérer les mentions des fonctions tout en réduisant les bruits. Cela est également envisagé pour la nature du Concept. En effet le Concept testé pourrait par exemple être un programme de coaching. Or, le mot *programme* étant associé à des classes relatives à l'organisation des participants ou bien à des composants *software*, chaque mention du Concept pour les participants pourraient créer un bruit important. Nous pourrions donc, sur la base des noms saisis par les analystes, éliminer du lexique, pour les besoins d'une étude en particulier, ce genre de lexies problématiques.

La cinquième ligne présente une occurrence de contresens, dû à la non prise en compte de *encore faudrait-il* dans le verbatim « *Encore faudrait-il pouvoir le porter pendant l'exercice physique.* ». Ce dernier est en effet interprété comme une perception de possibilité nouvelle offerte par le Concept. Ce type de problématiques est lié à un manque dans le lexique ; *encore faudrait-il* n'est fait pas partie, et l'expression d'une condition n'est

donc pas repérée. De la même manière, la notion de souhait n'est pas détectée, à la ligne suivante du tableau, dans « *une fonction permettant de mesurer son rythme cardiaque pendant la pratique sportive serait bien je pense* ». L'algorithme interprète ce verbatim comme un commentaire sur une fonction effectivement présente dans le Concept. Nous pourrions ajouter la forme *serait bien* comme entrée dans le lexique, mais il est très incertain que cela ne crée pas un bruit excessif tout en ne permettant d'identifier de tels cas. Ce type de problèmes est donc à traiter au cas par cas et ils constituent la partie la plus importante, en termes de temps à investir, des pistes d'amélioration que nous présentons ici en même temps que nous commentons les erreurs observées. Enfin, dans la dernière ligne du tableau, nous pouvons observer que, *se faire fliquer* ne faisant pas partie du lexique, la notion de conséquence néfaste et d'inadéquation avec les valeurs n'a pas été identifiée. Ce manque est facilement résoluble par un simple ajout de cette expression verbale, qui n'est que très peu susceptible de créer du bruit. Cette modification s'inscrirait dans un processus simple d'augmentation du lexique au fur et à mesure de la confrontation entre le système et les textes.

Enfin, plus globalement, nous n'identifions pas de manques importants au niveau des règles ; de manière générale, ce qui n'est pas identifié n'est pas dû à une configuration sémantique (i.e., association entre une réaction au Concept et un sujet) non anticipée par notre système de règles. Cela ne signifie pas que la couverture maximale de cet ensemble est atteinte, mais simplement que l'augmentation du nombre de ces règles n'est pas la démarche prioritaire à suivre pour améliorer les performances. Enfin, nous notons que la segmentation *a priori* en phrases ne pose pas de problème particulier ; très peu de corrections consistent à modifier la longueur des segments identifiés.

Des explications sur les raisons des échecs et réussites de notre système ayant été apportées, nous présentons à présent les opinions des analystes quant au système d'analyse qui leur a été proposé.

7.3.3. Retours d'expérience des analystes

Afin d'estimer quelle est l'adéquation de notre système avec les besoins des analystes au-delà de la mesure des performances, nous avons souhaité récolter des informations plus qualitatives à ce propos. Nous avons donc soumis aux personnes qui ont participé à cette expérimentation un petit questionnaire destiné à collecter le retour sur expérience. Nous présentons ici une synthèse de leurs réponses.

Un de nos questionnements portait sur la « stratégie d'analyse » qu'ont adoptée les analystes à partir des deux écrans possibles présentés (cf. FIGURES EN ANNEXE 5 et 6, pp. 269-270). Nous voulions savoir s'ils préféraient, par exemple, corriger les sorties de l'algorithme puis ajouter de nouveaux verbatim par la suite où s'ils réalisaient l'analyse linéairement. Nous voulions également savoir si les deux écrans avaient été utilisés. Leurs réponses sont unanimes : tous sont passés par l'analyse sur le texte entier, par besoin d'accéder à un contexte large des verbatim (a.). Cependant, à partir de ce texte entier, certains ont préféré tout corriger linéairement (b.), tandis que d'autres ont d'abord corrigé les verbatim annotés et ont ensuite ajouté ceux qui n'avaient pas été identifiés (c.). Notons toutefois que cette dernière stratégie a été appliquée par l'analyste novice, qui n'a donc pas l'habitude de procéder de façon linéaire comme ses collègues qui pratiquent l'analyse non pré-annotée depuis un certain temps.

- a. « J'ai essayé de commencer par les verbatim mais je me suis perdu, je devais supprimer, revenir au texte pour comprendre des verbatim incomplets, je retrouvais des questions d'animateurs, etc. J'ai changé de stratégie : d'abord j'ai relu et "nettoyé" le texte (suppression erreurs sélection des questions par les animateurs, sélection de verbatim non sélectionnés). » (*Sen.2*)
- b. « Je suis passée directement par le texte intégral, en le relisant dans l'ordre chronologique, en catégorisant ce qui n'était pas catégorisé et en corrigeant la classification automatique au fur-et-à-mesure. » (*Sen.1*)
- c. « En arrivant dans le projet, je suis restée dans l'onglet texte avec en dessous la liste des verbatim. J'ai en amont corrigé / modifié cette liste en corrigeant en premier les Critères CAUTIC® [...] Une fois ce premier travail effectué, j'ai consulté dans le détail le texte pour ajouter ou compléter des verbatim. Et enfin je suis allée vérifier et relire mon analyse dans la partie dédiée au verbatim. » (*Nov.*)

D'autre part, en ce qui concerne l'efficacité-même du système, malgré les scores que nous avons obtenus – qui se sont avérés inférieurs à nos espérances – la performance du système est ressentie comme plutôt positive par les analystes (d., e. et f.).

- d. « [la classification est] plutôt pertinente. » (*Sen.1*)
- e. « Peu d'erreurs » (*Jun.*)
- f. « Je trouvais les analyses plutôt bien faites » (*Nov.*)

On note qu'une certaine indulgence à l'égard des erreurs du système est exprimée par les analystes qui estiment que certaines limites ne peuvent être franchies par un traitement automatique (g.), même si certains perçoivent le fait que les règles pourraient être affinées pour obtenir de meilleurs résultats (h.).

- g. « Lorsqu'il y a des erreurs, ce n'est pas une question de mauvaise logique de la règle mais plutôt des questions de contexte (que l'IA ne peut pas cerner). » (*Sen.1*)
- h. « Les erreurs sont liées à mon sens à des règles un peu ambiguës et/ou trop larges. » (*Nov.*)

Les analystes identifient plusieurs problèmes et réussites récurrentes. Le Critère 1.4 est considéré comme bien identifié (i.) ; de nombreux ajouts ont dû être faits en correction pour ce dernier, mais il est vrai que relativement peu de classifications ont été supprimées²⁰¹. La surreprésentation du Critère 1.1 est bien identifiée, mais elle est associée au fait que le système ne prend pas en compte la chronologie du texte (j.) ; c'est un élément que nous avons déjà identifié²⁰² comme potentiellement problématique (i.e. le *spontané* concerne principalement ce qui est exprimé en début de collecte, tout de suite après la présentation du Concept). Il est donc envisagé que, dans le futur, nous ajoutions une nouvelle limitation à l'annotation par le Critère 1.1 basée sur une telle considération. Du bruit est identifié comme particulièrement récurrent pour les Critères 2.1 et 1.3 (k. et l.) ; cela est confirmé par les scores dans le cas du Critère 1.3, mais les attributions au Critère 2.1 n'ont en majorité pas été supprimés. Il faut alors peut-être entendre dans le commentaire k. que le système ne doit pas donner l'impression de fonctionner par simples mots clefs, car ce qui est deviné comme tel est considéré comme peu fiable. A l'inverse, le silence est souligné pour ce qui concerne les Critères du Niveau 4 (m. et n.), malgré le fait que les Critères de ce Niveau n'aient pas non plus beaucoup été utilisés par les analystes.

²⁰¹ Cf. FIGURE 17 (p. 228).

²⁰² Cf. Partie II, Chapitre VII, point 2.1.2.1 (p. 166).

- i. « J'ai eu l'impression que ce qui marche bien est "l'intérêt pour les fonctions". » (*Sen.2*)
- j. « J'ai eu l'impression que beaucoup de choses allaient en "intérêt spontané" or (c'est un problème de fond de CAUTIC) peut-on appeler "spontané" quelque chose qui arrive lors des derniers moments d'expression des participants ? Moi je ne crois pas. » (*Sen.2*)
- k. « Souvent lorsqu'il y a le mot "simple" ou "compliqué" le verbatim est classé dans le Critère 2.1 alors que ce n'est pas toujours juste (*Jun.*)
- l. « J'ai aussi l'impression qu'il y a eu pas mal de Critères en "comparaison existant" qui pour moi ne reflétaient pas ce Critère... » (*Sen.1*)
- m. « Le Critère 4.1 n'est pas toujours bien identifié » (*Jun.*)
- n. « J'ai noté une difficulté de la machine à classer dans le niveau 4. » (*Sen.1*)

Enfin, les cas de verbatim recevant trop de classes sont soulignés comme problématiques (o. et p.). Il s'agirait donc d'imaginer des solutions à cela, qui pourraient consister en une segmentation des phrases longues ou bien par élimination de certains étiquetages par priorisation par règles ou bien, si un modèle de ML s'avérait satisfaisant, sur la base de probabilités.

- o. « Pour certains verbatim longs, on peut avoir jusqu'à 5 catégorisations mais je ne sais pas si c'est pertinent... » (*Sen.1*)
- p. « Parfois beaucoup de catégorisations pour un seul verbatim » (*Jun.*)

Enfin, en ce qui concerne le choix que nous avons fait d'ajouter des phrases d'explication des classifications pour satisfaire un besoin de transparence des choix de l'algorithme, nous constatons qu'il est plutôt bien accueilli par les participants qui trouvent ces explications claires et utiles (q. et r.). Cependant, il est à noter que l'utilité de ces phrases semble être perçue de moins en moins importante lorsque le niveau d'expertise des analystes augmente : l'analyste novice dit les avoir lues systématiquement (r.) tandis qu'une sénior dit les avoir consultées seulement lorsque la classification

proposée suscitait un certain doute (s.). Il est d'ailleurs à souligner que l'analyste novice dit que les règles l'ont aidée à réaliser l'analyse. Ce commentaire est positif mais peut révéler un potentiel problème suscité par le système ; il ne faudrait pas que ce dernier influence trop l'apprentissage de la classification en imposant une vision trop rigide et restreinte du contenu attendu dans chaque Critère.

- q. « La règle aide bien à comprendre pourquoi tel verbatim est classé dans tel Critère » (*Jun.*)
- r. « J'ai quasiment systématiquement lu les commentaires [i.e., phrases d'explication], car je les trouvais vraiment utiles pour comprendre certaines classification et choix. De plus, en prenant connaissance de certaines règles, cela a pu me guider un peu pour la suite de l'analyse. » (*Nov.*)
- s. « Je les ai lus uniquement lorsque je ne comprenais pas ou n'étais pas d'accord avec la catégorisation. Ils permettent de vérifier la logique appliquée. » (*Sen.1*)

En somme, les analystes sont plutôt satisfaits du système que nous leur proposons malgré ses limites de performances, qui sont par ailleurs relativement bien identifiées et en partie tolérées. Cette analyse des résultats était la dernière que nous proposons, et nous présentons désormais nos conclusions sur cette ultime expérimentation.

7.4. Résumé et conclusions

Nous avons mis en œuvre une expérimentation qui visait à anticiper les conditions réelles dans lesquelles le système sera utilisé. Pour cela, ce module d'analyse que nous avons construit a été intégré à une plateforme dédiée dont l'interface est conçue pour en faciliter l'utilisation. Les scores que nous avons atteints ne sont pas à la hauteur de nos attentes, mais le fait qu'ils soient meilleurs que ceux obtenus sur notre corpus d'apprentissage est encourageant. Ces scores sont par ailleurs à mettre en perspective avec la complexité de la tâche à résoudre et nous avons montré que, pour le moins à partir des seules données d'entraînement dont nous disposions, un système de ML ne permettait pas d'atteindre des performances aussi élevées. La proposition de système que nous avons formulée semble en effet avoir une plus forte capacité à appliquer son modèle sur des données nouvelles et de nature sensiblement différentes à celles du corpus

d'apprentissage. En l'occurrence, le changement de nature des données a plutôt joué à notre avantage.

Cette expérimentation a été l'occasion pour nous d'identifier plusieurs zones de lacunes plus ou moins faciles à combler, ainsi que plusieurs pistes d'amélioration passant par de petites modifications du modèle, des ajouts dans ce dernier, mais aussi par des changements plus importants comme par exemple le passage à un système hybride (i.e. complété par des classifieurs statistiques). Nous estimons donc que nous avons un système qui ne demande qu'à être perfectionné mais qui peut permettre d'atteindre des résultats satisfaisants dans la limite de ce que l'on est en droit d'attendre raisonnablement d'une solution d'automatisation pour une tâche si particulière.

8. Bilan et pistes d'amélioration

Ce chapitre avait pour ambition de montrer comment nous avons affronté la question du passage d'un proto-système de type *preuve de concept* à un outil utilisable par les analystes, quoique n'ayant pas encore atteint un stade de maturité suffisant pour permettre de réduire le temps d'analyse de façon assez significative. Nous avons pour cela augmenté sa base lexicale, puis l'avons intégré à une plateforme opérationnelle.

Grâce à cela, malgré le fait que notre système n'obtienne pas des scores très élevés, il peut d'ores et déjà être utilisé dans le cadre de projets réels. Il est par ailleurs bien accepté par les analystes qui sont conscients de ses limites, même si un point d'attention est à souligner en ce qui concerne l'impact que ce système pourrait avoir dans le cadre d'une utilisation par un analyste débutant. Cette opérationnalisation du système peut être perçue comme prématurée, mais elle a en réalité pour ambition de permettre le passage à une phase d'amélioration par apprentissage en continu. Il s'agira en effet de systématiser l'utilisation de ce système au sein de l'entreprise, afin de collecter rapidement, par le biais de la plateforme, tous les textes d'études contenant nos sorties corrigées. Cela nous permettra d'agréger de nouvelles données annotées plus proches de celles qui nous devons traiter par rapport au corpus qui nous a servi à entraîner notre modèle. Cela nous permettra aussi de repérer très facilement les erreurs et lacunes récurrentes de notre modèle, que nous pourrons ainsi affiner et modifier dans une logique d'amélioration continue.

Cette première expérimentation nous a d'ailleurs démontré que, sur la base des résultats d'une première étude seulement, de nombreuses pistes d'amélioration peuvent déjà être identifiées. Certaines de ces améliorations seraient relativement simples à mettre en place par le biais d'alignement des pratiques et d'ajouts de formulaires sur la plateforme d'analyse. Il a notamment été question :

- De permettre à l'analyste de saisir les pseudonymes des animateurs pour que leurs questions ne soient plus données en sortie par l'algorithme.
- De demander aux analystes de saisir dans des formulaires les différents noms par lesquels le Concept et ses fonctions sont susceptible d'être nommés par les participants, ce qui constituerait une façon d'améliorer la désambiguïsation sémantique grâce à un ajout simple d'informations pragmatiques.
- De demander aux analystes de stabiliser la formulation de leurs questions afin de mieux pouvoir anticiper le vocabulaire qui sera utilisé par les participants pour y répondre.

Un autre type d'amélioration concernerait l'ajout de restrictions spécifiques dans l'attribution des classes par l'algorithme. Il s'agirait :

- De limiter le nombre de classes données en sortie pour un seul verbatim, en utilisant une priorisation par règles ou par probabilité.
- D'ajouter une règle pour le Critère 1.1 qui intégrerait la notion de chronologie du texte.

Enfin, un troisième type d'améliorations possibles nécessiterait l'intégration de nouvelles techniques du TAL dans notre système. Il s'agirait dans ce cas, par exemple :

- D'intégrer la possibilité de résoudre les reprises anaphoriques des questions dans les réponses des participants, ce qui nécessiterait l'utilisation d'un module de résolution des anaphores.
- De tester des méthodes hybrides qui mêlent notre modèle à des approches statistiques (plus nous collecterons de données annotées rapidement, plus vite cette solution deviendra envisageable).

En outre, et au-delà du fait que la phase de travail présentée dans ce chapitre ait permis d'opérationnaliser le système conçu tout en identifiant des pistes d'amélioration, elle a également été l'occasion de nombreux retours sur des points abordés plus tôt. En

cela, elle a permis d'apporter des éléments de réponse à des interrogations qui ont ponctué cette thèse, ou bien de réaffirmer des faits qui avaient été supposés avec un degré plus ou moins élevé de certitude. Par exemple, nous avons confirmé le fait qu'une approche ascendante sur la base de classifieurs statistiques n'était certainement pas la plus adaptée à notre problématique. Nous avons également eu confirmation du fait que la répartition inégale des Critères est relativement constante dans la pratique de CAUTIC®. Enfin, nous avons aussi confirmé notre hypothèse qui stipulait que le faible accord inter-juge qui avait été mesuré lors de la phase d'évaluation des verbatim hors contexte était davantage lié à la nature de l'exercice qu'il n'était représentatif de ce qui est atteint dans le cadre d'une analyse normale. Cela est également l'occasion de réaffirmer que cette nature de tâche « hors contexte », qui soustrait au texte son interprétation linéaire et les informations pragmatiques qui l'alimentent, est la caractéristique initiale majeure qui rend insolubles de nombreux problèmes pour notre algorithme.

CONCLUSIONS ET PERSPECTIVES

Nous concluons cette thèse en proposant un bref résumé de ce qui y a été exposé puis faisons un retour au questionnement initial afin de faire le point sur les éléments de réponse que nous avons pu y apporter. Suite à cela, nous terminons par présenter les perspectives de travaux futurs.

1. Résumé et réponses aux questionnements

Les travaux que nous avons présentés visaient à explorer les possibilités du TAL pour créer un système d'autocodage pensé spécifiquement pour une méthode d'analyse qualitative précise. Nous avons tout d'abord cherché à préciser la demande formulée par l'entreprise puis avons investigué le domaine de l'analyse qualitative outillée. Cette exploration du domaine nous a démontré que l'autocodage était un domaine relativement peu exploré. Les quelques travaux qui en traitaient ont néanmoins constitué une source précieuse de recommandations pour la réalisation d'un tel système et nous ont permis de souligner plusieurs particularités propres à nos travaux, telles que la stabilité de la grille d'analyse et la complexité particulière des codes qu'elle utilise. La première a constitué une opportunité d'approfondir l'exploration des possibilités du TAL pour l'autocodage sans avoir à anticiper une variabilité importante des codes mobilisés, tandis que la seconde a représenté le principal défi qui nous était adressé. Cette exploration des possibilités de TAL nous a amenée à explorer deux grandes approches. La plus courante, ascendante, ne s'est pas montrée fructueuse, tandis qu'une approche descendante a donné de meilleurs résultats et a permis d'adresser des points d'exigence en termes d'outil final.

Le questionnement transverse que nous avons proposé dans cette thèse (p. 84) était articulé sur deux axes subdivisés chacun en trois points. Le premier axe portait sur la manière dont nous pouvions mobiliser le TAL pour créer un système d'autocodage selon les spécificités propres à nos travaux. La première de ces spécificités concernait le fait que l'analyse qui doit être réalisée s'inscrit dans un domaine spécifique qui est celui de l'évaluation de l'acceptabilité et est à visée de diagnostic. Sur ce point, nous avons apporté une réponse qui consistait à construire un système basé sur la modélisation, sous forme d'ontologie, du paradigme de l'entretien d'étude d'acceptabilité. Cela fait du modèle

proposé *in fine* un outil hyperspécialisé qui se distingue clairement de solutions généralistes pour l'analyse qualitative, ou d'analyse des sentiments souvent employées en satisfaction clients pour des produits déjà commercialisés. Cette hyperspécialisation permet donc d'analyser le texte selon les dimensions de l'acceptabilité des innovations dans toute leur complexité mais aussi uniquement selon ce prisme d'interprétation du texte.

Le second point portait sur le fait que la grille d'analyse était prédéfinie et comportait des codes caractérisés par une complexité sémantique certaine et qui devaient pouvoir être appliqués sur des corpus multi-domaines. Nous avons tout d'abord démontré que la complexité sémantique des codes nécessitait que le texte soit analysé avec un certain degré d'interprétation et en passant par la détection de concepts d'intérêt articulés selon des configurations spécifiques. Nous avons également mis en évidence le fait que les seules fréquences lexicales ne permettaient pas de détecter la présence de ces concepts d'intérêt. Enfin, nous avons également éliminé la possibilité d'utiliser une annotation des verbes reliés à des classes sémantiques identifiées par leur structure argumentative, car la seule ressource permettant de réaliser une telle annotation pour le français n'était pas adaptée au repérage, dans les textes, des éléments d'intérêt potentiellement discriminants. La solution qui a été finalement choisie propose d'identifier les différentes configurations sémantiques attendues par chaque classe afin de pouvoir les identifier dans les textes. En mettant l'accent sur la prise en compte d'un vocabulaire spécifique à l'acceptabilité des innovations, elle permet une moindre sensibilité du système aux variations lexicales liées à des domaines particuliers. La non-variabilité des résultats obtenus en fonction des domaines reste cependant à confirmer par le biais d'expérimentations multiples sur des corpus variés.

Enfin, le dernier point de ce premier axe portait sur le fait que nous ne disposions, pour réaliser notre tâche, que d'un petit corpus étiqueté. Cette taille de corpus a certainement eu un impact sur la certitude avec laquelle nous pouvons généraliser nos résultats obtenus par l'utilisation de classifieurs statistiques, que nous avons donc dû éliminer sans pouvoir évaluer tout leur potentiel. Il n'en reste pas moins que la disponibilité de corpus contenant des étiquetages similaires aux nôtres par leur degré de spécialisation est assez réduite, et que les performances potentielles générales du ML pour ce type de tâches s'en trouvent donc *de facto* amoindries. D'autre part, notre corpus

s'est avéré non seulement de taille réduite mais aussi particulièrement déséquilibré, ce qui constitue également un problème bien connu pour le ML. La méthode que nous avons proposée dans le cadre de l'approche descendante permet de tirer de ce petit corpus des règles qui ne sont pas statistiques mais basées sur la sémantique et l'identification humaine des caractéristiques utiles à la classification. Enfin, au-delà de sa taille et de son déséquilibre, le corpus d'entraînement avait pour caractéristique de contenir des transcriptions d'oral, tandis que nos données cibles consistaient en des *posts* issus d'internet. Cette différence a profité à notre système par règles sémantiques, alors qu'il a impacté négativement les performances des modèles de ML. Ne disposant pas d'assez de données issues d'internet pour entraîner des modèles statistiques, nous ne pouvons pas indiquer quelles auraient été leurs performances si ces données avaient pu servir pour l'entraînement. Il est très probable que ces performances auraient été meilleures, et même possible qu'elles auraient surpassé celles de notre système par règles. Cependant, cette indisponibilité de données nous a permis de proposer une alternative aux approches ML dans de tels cas.

Dans le deuxième axe de questionnement, nous nous demandions comment faire en sorte que le système réponde à plusieurs exigences. La première était une exigence d'efficacité en termes de réduction de temps de codage. Sur ce point, nous ne pouvons guère apporter de réponse plus complète que celle qui consiste à prédire que, plus le système sera performant, plus le temps d'analyse sera réduit. En effet, si cette exigence constituait le cœur de la demande de l'entreprise, le système final ayant été opérationnalisé à la fin du temps imparti pour la thèse, nous n'avons pas eu l'occasion de mesurer la réduction du temps effective qu'il permettait ; les besoins particuliers de chaque étude étant susceptibles d'impacter le temps d'analyse, cette mesure ne pourrait se faire de manière fiable qu'en comparant le temps d'analyse de plusieurs études pour en tirer des moyennes et analyser si l'amélioration progressive du système coïncide en effet avec une réduction du temps d'analyse.

Le deuxième point de cet axe concernait les besoins de transparence et de possibilité de contrôle permis par l'outil. Nous y avons répondu en proposant dans le système final des phrases qui explicitent l'interprétation que la machine fait du texte et qui justifie sa classification. Cette possibilité n'aurait pas été permise par des systèmes de ML et joue donc en faveur du système par règles. Le contrôle sur les sorties de l'algorithme

a davantage été fourni par le biais d'une conception d'interface d'analyse permettant une correction facile des sorties de l'algorithme que par les modalités de fonctionnement de ce dernier.

Enfin, nous exigeons que le TAL n'impose pas un nouveau paradigme d'analyse. Cette exigence a été traitée de par le fait que nous avons tous simplement veillé à ne pas proposer de méthodes de TAL efficaces mais qui demandent d'adapter la méthode CAUTIC® en fonction de cette efficacité ; nous avons plutôt forcé le TAL à coller au plus près à la méthode concernée, quitte à obtenir des performances moins élevées.

Après ce retour sur la problématique qui a été traitée dans cette thèse, nous tournons à présent notre regard vers le futur en présentant les perspectives envisagées.

2. Perspectives

Nous présentons ici les pistes dont nous disposons pour améliorer le système puis terminons par décrire les perspectives pour celui-ci à plus long terme.

2.1. Pistes d'amélioration

Nous décrivons dans ce point les grandes lignes des modalités d'amélioration envisagées puis précisons les pistes dont nous disposons aujourd'hui pour faire des ajouts sur le modèle existant, des tests de modification de l'algorithme, et des expérimentations pour aller vers des approches hybrides complémentaires.

2.1.1. Direction générale

Le système que nous avons proposé dans la troisième partie de ce document est aujourd'hui intégré à une plateforme qui permet aux analystes de l'utiliser. Si ces derniers sont plutôt positifs à son propos, ils en connaissent également les limites actuelles et celles qui ne seront *a priori* jamais dépassées par le TAL. Il est donc entendu et accepté que les interprétations textuelles d'ordre pragmatique resteront l'affaire des codeurs humains. Néanmoins, quelques informations relatives particulières au projet – le nom du Concept et de ses fonctions, ainsi que les pseudonymes des animateurs – pourraient être saisies sur la plateforme afin d'apporter un peu de connaissances particulières à un projet au système qui pourrait les exploiter. Les autres dimensions pragmatiques à prendre en compte ne pourront pas, ou difficilement, être traitées : la connaissance de certaines particularités du marché dans lequel s'inscrit le Concept, les grandes tendances d'opinion

dans la population sur certains sujets ou encore les dimensions techniques problématiques par essence dans un domaine précis.

Nous entendons, à l'avenir, profiter du fait que toutes les nouvelles études qui seront menées soient traitées via la plateforme d'analyse qui intègre notre système. Cela nous permettra de constituer rapidement un corpus qui sera non seulement annoté mais qui contiendra également les marques de corrections des sorties de notre algorithme. Nous serons donc en mesure d'identifier facilement les parties faibles du modèle afin de les corriger, tout en augmentant notre lexique à partir de celui contenu dans les nouvelles données. Il s'agira pour cela de penser et construire un procédé efficace de collecte des études corrigées, d'analyse des erreurs de l'algorithme, de priorisation des problèmes à résoudre et enfin de modifications à apporter. Cela ne peut se passer, dans un premier temps, d'une intervention humaine, mais notre volonté est d'automatiser autant que faire se peut ces différentes étapes afin que cette intervention humaine se résume de plus en plus à une simple supervision. En outre, les diverses modifications que nous apporterons à notre système pouvant être facilement déployées opérationnellement, il sera aisé de tester rapidement leur impact en observant si, pour chaque modification apportée, les corrections apportées sur quelques études réalisées après cette modification sont plus ou moins nombreuses.

D'autre part, la systématisation de l'utilisation du système par les analystes sera porteuse d'une certaine standardisation dans leurs pratiques de l'analyse CAUTIC® outillée, ce qui amènera à des possibilités d'affinités plus importantes entre ces pratiques et notre système. Il reste tout à fait exclu, cela faisait partie des exigences initiales, que le système impose un changement de paradigme d'analyse. Cependant, des changements de pratiques très légères comme la stabilisation lexicale des questions posées aux participants – qui ont tendance à reprendre mot pour mot des éléments de ces questions – ne représente pas une contrainte particulière pour les chargés d'études et peut participer à améliorer les performances du système.

2.1.2. Ajouts au modèle existant

Le processus d'amélioration par apprentissage continu à partir des corrections des erreurs de l'algorithme consistera certainement en l'ajout de nombreux éléments au modèle existant. Une bonne partie de ces ajouts nécessaires reste à découvrir, mais certaines opportunités peuvent d'ores et déjà être identifiées.

En termes d'exploitation de techniques du TAL qui restent à ce jour non exploitées dans notre système, nous pensons que des outils de détection des entités nommées et de résolution des anaphores pourraient lui être bénéfiques. Les premiers serviraient à mieux reconnaître des éléments tels que les noms d'entreprises. Les seconds pourraient être mobilisés pour mieux faire le lien entre les questions posées aux participants et leurs réponses, ou tout simplement pour détecter les anaphores pronominales entre deux phrases qui sont, rappelons-le, analysées séparément.

D'autre part, le lexique sera augmenté de plusieurs façons. Premièrement, les corpus nous serviront à identifier de nouvelles catégories lexico-sémantiques d'intérêt qui peuvent être liées aux classes de notre ontologie, voire nécessiter l'ajout d'une nouvelle classe. Concernant l'augmentation du nombre d'unités lexicales pour chacune des catégories lexico-sémantiques, au moins deux pistes sont envisageables. La première consiste à exploiter de nouvelles ressources lexicales ; WoNef et WOLF n'ont pas été utilisées car identifiées comme moins riches que JeuxDeMots, mais il n'en reste pas moins qu'elles sont fortement susceptibles de contenir des unités lexicales d'intérêt que nous n'avons pas identifiées. D'autres ressources méritent également notre attention, comme par exemple GLAWI (Hathout & Sajous, 2016) qui a été construite à partir du *Wiktionnaire* français²⁰³. Il est également possible d'envisager le recours à des ressources qui ne sont pas uniquement basées sur une structuration du lexique hiérarchique et/ou segmentée en classes (Polguère, 2009) ; en particulier, le modèle proposé par la Lexicologie Explicative et Combinatoire (Mel'čuk & Polguère, 1995) pourrait s'avérer particulièrement intéressant pour la modalité d'extension par combinatoire que nous avons proposée et qui reste tout à fait artisanale. Une deuxième piste serait bien sûr de repérer et ajouter de nouvelles unités lexicales à partir des corpus récupérés. Cette approche pourrait être semi-automatisée ; nous pourrions par exemple envisager, pour chaque nouvelle unité lexicale non-étiquetée dans le texte, de calculer sa similarité sémantique avec les ensembles lexicaux que constituent les unités lexicales de chaque entrée du lexique.

²⁰³ https://fr.wiktionary.org/wiki/Wiktionnaire:Page_d%E2%80%99accueil (consulté le 22/12/21).

2.1.3. Tests de modifications de l'algorithme

En plus de ces ajouts au modèle, quelques modifications de l'algorithme mériteraient d'être testées. Nous pourrions notamment étudier le bénéfice qui pourrait être apporté par une segmentation différente que celle que nous appliquons ; il est par exemple possible d'imaginer que le système considère des segments de plusieurs phrases si une seule n'a pas permis de conclure à une classification. Il est également possible, sur la base des nouveaux corpus que nous collectons, de vérifier si la modalité d'identification des triplets dans le texte que nous avons identifiée comme la plus efficace – par la présence de deux classes sans prendre en compte la propriété – est bien la meilleure option et, le cas échéant, apporter des modifications. Enfin, des tests de priorisation de certains Critères sur d'autres par l'algorithme lorsque plusieurs sont identifiés pourraient être facilement réalisés.

2.1.4. Expérimentations sur des approches complémentaires

Tout en conservant l'idée que le travail mené ici ne doit pas être recommencé, il est possible de continuer à tester d'autres approches complémentaires. Les classifieurs statistiques pourraient s'avérer bien plus efficaces que ce qui a été constaté dès lors que nous aurons un corpus assez important de *posts* sur lequel les entraîner. D'ailleurs, nos investigations sur les modèles de type BERT ont été très limitées et, étant données les performances remarquables constatées ces dernières années grâce à ces derniers, il serait tout à fait intéressant de les approfondir malgré les résultats préliminaires insatisfaisants que nous avons obtenus. D'autre part, sur cette partie statistique, il est possible qu'une certaine hybridation du système soit permise. Cette hybridation pourrait passer par la classification statistique des données non étiquetées par notre modèle de règles, ou par une pondération des sorties d'un modèle par celles de l'autre, ou bien encore par la prise en compte des informations sémantiques fournies par notre modèle en tant que caractéristiques à donner en entrée à des modèles de ML.

Les modalités d'amélioration du système que nous envisageons à court terme ayant été explicitées, nous concluons cette thèse par les perspectives possibles à plus long terme.

2.2. Perspectives à long terme

Le module que nous avons créé se veut être un premier pas vers un système de TAL ultraspécialisé dans les analyses qualitatives UX pour l'évaluation des innovations.

Nous avons commencé à définir un modèle conceptuel et lexical de l'acceptabilité qui doit pouvoir adresser n'importe quel domaine d'activité (médecine, électronique, sport, etc.) mais qui consiste, en réalité, en la modélisation du domaine *per se* qu'est l'acceptabilité. Ainsi, si le système construit est très spécifique à CAUTIC® et ne peut à lui seul servir à d'autres méthodes d'analyse, nous avons initié la construction d'un véritable lexique de l'acceptabilité qui pourrait servir d'autres objectifs.

D'autre part, l'UX englobant d'autres dimensions que l'acceptabilité psychosociale, des travaux visant à l'implémentation d'une nouvelle méthode d'analyse sont, au moment où nous écrivons ces lignes, en train d'être initiés ; il s'agira de créer un système permettant l'analyse de l'acceptabilité et de la désirabilité des innovations selon des critères psycho-cognitifs et émotionnels selon la méthode EMINOSA® (Loeser, 2019) qui a été élaborée dans le cadre de travaux de recherche également à l'initiative de l'entreprise Ixiade. Nous entendons donc, à terme, constituer un véritable ensemble cohérent d'outils d'analyse qui mettent le TAL au service de l'UX. Enfin, nous pensons que cette suite d'outils hyperspécialisés, une fois optimisés, pourrait bénéficier au plus grand nombre par le biais d'une adaptation multilingue.

BIBLIOGRAPHIE

- Abbe, A., Grouin, C., Zweigenbaum, P., & Falissard, B. (2016). Text mining applications in psychiatry : A systematic literature review. *International journal of methods in psychiatric research*, 25(2), 86-100. [↑p52](#)
- Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications : A survey. *Heliyon*, 4(11), Article e00938. [↑p127](#)
- Adeva, J. G., Atxa, J. P., Carrillo, M. U., & Zengotitabengoa, E. A. (2014). Automatic text classification to support systematic reviews in medicine. *Expert Systems with Applications*, 41(4), 1498-1508. [↑p126;p127](#)
- Agrawal, A., Fu, W., & Menzies, T. (2018). What is wrong with topic modeling? And how to fix it using search-based software engineering. *Information and Software Technology*, 98, 74-88. [↑p56](#)
- Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Mohammed, K., Malik, R., Almahdi, E., Chyad, M., Tareq, Z., & Albahri, A. (2020). Sentiment analysis and its applications in fighting COVID-19 and infectious diseases : A systematic review. *Expert systems with applications*, 167, Article 114155. [↑p55](#)
- Alkhatib, W., Sabrin, S., Neitzel, S., & Rensing, C. (2018). Towards Ontology-Based Training-Less Multi-label Text Classification. In M. Silberztein, F. Atigui, E. Kornysheva, E. Métais, & F. Meziane (Éds.), *Natural Language Processing and Information Systems* (p. 389-396). Springer International Publishing. [↑p164](#)
- Allen, M. (Éd.). (2017). *The SAGE Encyclopedia of Communication Research Methods* (Vol. 1). Sage Publications. [↑p99](#)
- Alom, Z., Carminati, B., & Ferrari, E. (2020). A deep learning model for Twitter spam detection. *Online Social Networks and Media*, 18, Article 100079. [↑p121](#)
- Andréani, J.-C. (2004). Marketing du produit nouveau : 95% des produits nouveaux échouent. Les managers sont en cause, les études de marché aussi. *Icade. Revista de las Facultades de Derecho.*, 61, 413-416. [↑p9](#)
- Anjembe, E. (2014). *La projection d'usage des TICs : La composition de fictions axiomatiques au service de la recherche technologique* [Thèse de Doctorat]. Université Grenoble Alpes. [↑p11;p12;p58](#)
- Antoniadis, G. (1984). *Élaboration d'un système d'analyse morphosyntaxique d'une langue naturelle : Application en informatique documentaire*. [Thèse de Doctorat]. Université de Grenoble II. [↑p195](#)
- Antons, D., Grünwald, E., Cichy, P., & Salge, T. O. (2020). The application of text mining methods in innovation research : Current state, evolution patterns, and development priorities. *Special Issue : Innovation Management Research Methods*, 50(3), 329-351. [↑p51;p52](#)
- Aragón, M. E., Jarquín-Vásquez, H. J., Montes-Y-Gómez, M., Escalante, H. J., Pineda, L. V., Gómez-Adorno, H., Posadas-Durán, J. P., & Bel-Enguix, G. (2020). Overview of MEX-A3T at IberLEF 2020 : Fake News and Aggressiveness Analysis in Mexican Spanish. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)*, 222-235. [↑p128](#)
- Archer, D. (Éd.). (2008). *What's in a word-list ? Investigating word frequency and keyword extraction*. Ashgate. [↑p53](#)

- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., & Abbasi, H. M. (2018). A survey of ontology learning techniques and applications. *Database, 2018*. ↑[p163](#)
- Asmussen, C. B., & Møller, C. (2019). Smart literature review : A practical topic modelling approach to exploratory literature review. *Journal of Big Data, 6*(1), 1-18. ↑[p55](#)
- Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The berkeley framenet project. *Proceedings of the COLING-ACL'98 conference, 86-90*. ↑[p142](#)
- Bardin, L. (1977). *L'analyse de contenu*. Presses Universitaires de France. ↑[p38;p39](#)
- Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational research, 45*(2), 143-154. ↑[p78](#)
- Bauer, M. W. (2012). Classical content analysis : A review. In M. W. Bauer & G. Gaskell (Éds.), *Qualitative researching with text, image and sound (Electronic version)* (p. 131-151). Sage. ↑[p38](#)
- Béchet, F., Damnati, G., Heinecke, J., Marzinotto, G., & Nasr, A. (2017). CALOR-Frame : Un corpus de textes encyclopédiques annoté en cadres sémantiques. *ACor4French-Les corpus annotés du français-Atelier TALN*. ↑[p143](#)
- Beliga, S., Kitanović, O., Stanković, R., & Martinčić-Ipšić, S. (2017). Keyword extraction from parallel abstracts of scientific publications. *Semanitic keyword-based search on structured data sources, 44-55*. ↑[p53](#)
- Bencini, G. M. L., & Goldberg, A. E. (2000). The Contribution of Argument Structure Constructions to Sentence Meaning. *Journal of Memory and Language, 43*, 640-651. ↑[p141](#)
- Bendaoud, R., Toussaint, Y., & Napoli, A. (2010). L'analyse Formelle de Concepts au service de la construction et l'enrichissement d'une ontologie. *Revue des Nouvelles Technologies de l'Information, 133-164*. ↑[p164](#)
- Beniwal, R., Jain, M., & Gupta, Y. (2021). Opinion Mining to Aid User Acceptance Testing for Open Beta Versions. In P. Bansal, M. Tushir, V. E. Balas, & R. Srivastava (Éds.), *Proceedings of International Conference on Artificial Intelligence and Applications* (p. 291-301). Springer. ↑[p58](#)
- Bienvenu, M., Leclère, M., Mugnier, M.-L., & Rousset, M.-C. (2020). Reasoning with Ontologies. In P. Marquis, O. Papini, & H. Prade (Éds.), *A Guided Tour of Artificial Intelligence Research: Vol. Volume I: Knowledge Representation, Reasoning and Learning* (p. 185-215). Springer. ↑[p160](#)
- Blanche-Benveniste, C., & Jeanjean, C. (1987). *Le français parlé. Transcription et édition*. Didier Erudition. ↑[p25](#)
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research, 3*(1), 993-1022. ↑[p55](#)
- Bonial, C., Corvey, W., Palmer, M., Petukhova, V. V., & Bunt, H. (2011). A hierarchical unification of LIRICS and VerbNet semantic roles. *IEEE Fifth International Conference on Semantic Computing, 483-489*. ↑[p145](#)
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. In D. Haussler (Éd.), *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory* (p. 144-152). ACM Press. ↑[p127](#)
- Boughzala, Y., Moscarola, J., & Hervé, M. (2014). *Sphinx Quali : Un nouvel outil d'analyses textuelles et sémantiques*. JADT 2014 : 12es Journées internationales d'Analyse statistique des Données Textuelles, Paris, France. ↑[p35](#)
- Buhler, T., Bendinelli, M., Lethier, V., & He, Y. (2018). La textométrie pour révéler évolutions et oppositions dans les discours des politiques urbaines. Le cas des PDU français (2000-2015). *Flux, 113*(3), 93-110. ↑[p107](#)

- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews : Problems of unitization and intercoder reliability and agreement. *Sociological methods & research*, 42(3), 294-320. [↑p41;p44](#)
- Caputo, A., Giacchetta, A., & Langher, V. (2016). AIDS as social construction : Text mining of AIDS-related information in the Italian press. *AIDS care*, 28(9), 1171-1176. [↑p52](#)
- Charalampakis, B., Spathis, D., Kouslis, E., & Kermanidis, K. (2016). A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets. *Engineering Applications of Artificial Intelligence*, 51, 50-57. [↑p121](#)
- Chen, N.-C., Drouhard, M., Kocielnik, R., Suh, J., & Aragon, C. R. (2018). Using machine learning to support qualitative coding in social science : Shifting the focus to ambiguity. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(2), Article 9. [↑p3;p59;p60;p73;p77;p78;p93](#)
- Chen, N.-C., Kocielnik, R., Drouhard, M., Peña-Araya, V., Suh, J., Cen, K., Zheng, X., Aragon, C. R., & Peña-Araya, V. (2016). Challenges of applying machine learning to qualitative coding. *CHI 2016 Workshop on Human-Centered Machine Learning*. [↑p78](#)
- Chen, Y.-W., & Nakazawa, M. (2017). Emotions and Pan-Asian organizing in the US southwest : Analyzing interview discourses via sentiment analysis. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 28(6), 2785-2806. [↑p55](#)
- Cheng, H.-F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through UI : Strategies to help non-expert stakeholders. *Proceedings of the 2019 chi conference on human factors in computing systems (CHI '19)*, 1-12. [↑p78](#)
- Choi, D., Ko, B., Kim, H., & Kim, P. (2014). Text analysis for detecting terrorism-related articles on the web. *Journal of Network and Computer Applications*, 38, 16-21. [↑p121](#)
- Chung, S. K. G., Rodriguez, M., Lanier, P., & Gibbs, D. (2021). Text-mining open-ended survey responses using structural topic modeling : A practical demonstration to understand parents' coping methods during COVID-19 pandemic in Singapore. *OSF Preprint*. [↑p56](#)
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46. [↑p99](#)
- Comby, E. (2016). L'analyse de données textuelles et l'acceptation sociale. In S. Depraz, U. Cornec, & U. Grabski-Jieron (Éds.), *Acceptation sociale et développement des territoires*. ENS éditions. [↑p107](#)
- Cori, M. (2008). Des méthodes de traitement automatique aux linguistiques fondées sur les corpus. *Langages*, 3, 95-110. [↑p106](#)
- Creswell, J. W. (2015). *30 essential skills for the qualitative researcher*. Sage. [↑p40](#)
- Croce, F., Cima, G., Lenzerini, M., & Catarci, T. (2020). Ontology-based explanation of classifiers. *Workshop Proceedings of the EDBT/ICDT 2020 Joint Conference*. [↑p165](#)
- Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology*, 15(6), 523-543. [↑p3;p45;p51;p61;p62;p72;p78;p225](#)
- Crowston, K., Liu, X., & Allen, E. E. (2010). Machine learning and rule-based automated coding of qualitative data. *Proceedings of the American Society for Information Science and Technology*, 47, 1-2. [↑p63](#)
- Dahm, J. (2021). Couverture médiatique de la pandémie de COVID-19 en France et en Allemagne : Quelques pistes pour une analyse discursive et comparative au moyen des cadres sémantiques. *Synergies Pays germanophones*, 14, 61-74. [↑p143](#)

- Danlos, L., Pradet, Q., Barque, L., Nakamura, T., & Constant, M. (2016). Un Verbenet du français. *Traitement Automatique des Langues, ATALA*, 57(1), 33-58. [↑_{p5;p141;p143;p144;p145;p150;p151}](#)
- Dany, L. (2016). Analyse qualitative du contenu des représentations sociales. In G. Lo Monaco, S. Delouée, & P. Rateau (Éds.), *Les représentations sociales* (p. 85-102). de Boeck. [↑_{p38;p39}](#)
- Das, S., Muhuri, S., Chakraborty, S., & Biswas, S. (2020). Graph Based Keyword Extraction for Similarity Identification among Born-Digital News Contents. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1-7. [↑_{p53}](#)
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS quarterly*, 13(3), 319-340. [↑_{p10}](#)
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology : A comparison of two theoretical models. *Management science*, 35(8), 982-1003. [↑_{p10}](#)
- Daza, A., & Frank, A. (2020). X-SRL : A parallel cross-lingual Semantic Role Labeling dataset. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 3904-3914. [↑_{p144}](#)
- De Ruyter, K., & Scholl, N. (1998). Positioning qualitative market research : Reflections from theory and practice. *Qualitative market research: An international journal*, 1(1), 7-14. [↑_{p43}](#)
- Dejond, A., & Mercier, J. (2002). *La cyberl@ngue française*. Renaissance du livre. [↑_{p26}](#)
- Denzin, N. K. (1978). *The research act*. Aldine. [↑_{p39}](#)
- Desmarais, C., & Moscarola, J. (2004). Analyse de contenu et analyse lexicale, le cas d'une étude en management public. *Lexicométrica, Actes du colloque « L'analyse de données textuelles : De l'enquête aux corpus littéraires »*. [↑_{p40}](#)
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert : Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.- [↑_{p126}](#)
- Djemaa, M. (2017). *Stratégie domaine par domaine pour la création d'un FrameNet du français : Annotations en corpus de cadres et rôles sémantiques* [PhD Thesis]. Université Sorbonne Paris Cité. [↑_{p142}](#)
- Drouhard, M., Chen, N.-C., Suh, J., Kocielnik, R., Pena-Araya, V., Cen, K., Zheng, X., & Aragon, C. R. (2017). Aeonium : Visual analytics to support collaborative qualitative coding. *2017 IEEE Pacific Visualization Symposium (PacificVis)*, 220-229. [↑_{p77}](#)
- Dubois, J., & Dubois-Charlier, F. (1997). *Les verbes français*. Larousse. [↑_{pxi;p142}](#)
- Eensoo, E., & Valette, M. (2012). Sur l'application de méthodes textométriques à la construction de critères de classification en analyse des sentiments. *TALN 2012*, 2, 367-374. [↑_{p107}](#)
- Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, 23(11), 2850-2861. [↑_{p40;p42}](#)
- Ezike, T., & Sivanesan, M. (2020). Chrestotes at HASOC 2020 : Bert Fine-tuning for the Identification of Hate Speech and Offensive Language in Tweets. *Fire (Working Notes)*. [↑_{p128}](#)
- Fallery, B., & Rodhain, F. (2007). *Quatre approches pour l'analyse de données textuelles : Lexicale, linguistique, cognitive, thématique*. 1-16. [↑_{p36}](#)
- Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology : From ontological art towards ontological engineering. *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*, 33-40. [↑_{p253}](#)

- Fielding, N. G., & Lee, R. M. (2002). New patterns in the adoption and use of qualitative software. *Field methods*, 14(2), 197-216. ↑_{p48}
- Fillmore, C. J. (1982). Frame semantics. In *Linguistics in the Morning Calm* (p. 111-137). Hanshin Publishing. ↑_{p142}
- Fischbach, J., Vogelsang, A., Spies, D., Wehrle, A., Junker, M., & Freudenstein, D. (2020). *Specmate : Automated creation of test cases from acceptance criteria*. 321-331. ↑_{p58}
- Flick, U. (2017). Triangulation. In N. K. Denzin & Y. E. Lincoln (Éds.), *The SAGE Handbook of Qualitative Research (5th Edition)* (p. 777-804). SAGE Publications, Inc. ↑_{p39}
- Flisar, J., & Podgorelec, V. (2020). Improving short text classification using information from DBpedia ontology. *Fundamenta Informaticae*, 172(3), 261-297. ↑_{p129}
- Fort, K., & Guillaume, B. (2007). PrepLex : Un lexique des prépositions du français pour l'analyse syntaxique. *Traitement Automatique des Langues Naturelles - TALN 2007*. ↑_{p151}
- François, J., Le Pesant, D., & Leeman, D. (2007). Présentation de la classification des Verbes français de Jean Dubois et Françoise Dubois-Charlier. *Langue française*, 153, 3-19. ↑_{p142}
- Ganji, A., Orand, M., & McDonald, D. W. (2018). Ease on Down the Code : Complex Collaborative Qualitative Coding Simplified with « Code Wizard ». *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), Article 132. ↑_{p74}
- Ganter, B., & Wille, R. (1999). *Formal Concept Analysis. Mathematical Foundations*. Springer. ↑_{p164}
- Garnelo, M., & Shanahan, M. (2019). Reconciling deep learning with symbolic artificial intelligence : Representing objects and relations. *Current Opinion in Behavioral Sciences*, 29, 17-23. ↑_{p83;p83}
- Ghadfi, S., Béchet, N., & Berio, G. (2014). Building Ontologies from Textual Resources : A Pattern Based Improvement Using Deep Linguistic Information. *Proceedings of 5th Workshop on Ontology and Semantic Web Patterns (WOP2014)*, 14-25. ↑_{p164}
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine Publishing. ↑_{p39}
- Godbert, E. (2014). Typing verbs with LVF, for anaphora resolution (Typage sémantique de verbes avec LVF, pour la résolution d'anaphores)[in French]. *21ème Conférence annuelle sur le Traitement Automatique des Langues Naturelles*, 97-102. ↑_{p143}
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological Engineering : With examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. Springer-Verlag. ↑_{p163}
- Gomez-Perez, J. M., Denaux, R., & Garcia-Silva, A. (2020). Hybrid Natural Language Processing : An Introduction. In *A Practical Guide to Hybrid Natural Language Processing* (p. 3-6). Springer. ↑_{p82;p83;p227}
- Goncharova, O. (2017). *Méthodes et modèles de construction automatisée d'ontologies pour des domaines spécialisés* [PhD Thesis]. Université de Lyon. ↑_{p163}
- Goyal, A., Gupta, V., & Kumar, M. (2018). Recent Named Entity Recognition and Classification techniques : A systematic review. *Computer Science Review*, 29, 21-43. <https://doi.org/10.1016/j.cosrev.2018.06.001> ↑_{p53;p54}
- Grimmer, J., & Stewart, B. M. (2013). Text as data : The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297. ↑_{p254}
- Gross, M. (1975). *Méthodes en syntaxe. Régime des constructions complétives*. Hermann. ↑_{p141}
- Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5-6), 907-928. ↑_{p161}

- Gruber, T. (2009). Ontology. In L. Liu & M. T. Özsu (Éds.), *Encyclopedia of database systems* (Vol. 1, p. 1963-1965). Springer. [↑p162](#)
- Gruninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*. IJCAI-95, Montréal, Canada. [↑p163](#)
- Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In *Handbook on ontologies* (p. 1-17). Springer. [↑p161](#)
- Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. V. (2018). Augmenting qualitative text analysis with natural language processing : Methodological study. *Journal of medical Internet research*, 20(6), Article e9702. [↑p45](#)
- Gülich, E. (1999). Les activités de structuration dans l'interaction verbale. In M.-J. Barbéris (Éd.), *Le français parlé. Variétés et discours*. (p. 21-47). Praxiling. [↑p25](#)
- Guzman-Silverio, M., Balderas-Paredes, Á., & López-Monroy, A. P. (2020). Transformers and Data Augmentation for Aggressiveness Detection in Mexican Spanish. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)*, 293-302. [↑p128](#)
- Hadi-Denoueix, M. (2014). *Étude des caractéristiques discursives et sociales de l'écrit numérique : De la communication écrite médiée à la co-construction interactive de l'écrit socio-numérique* [Thèse de doctorat]. Université Aix-Marseille. [↑p26](#)
- Hadouche, F., & Lapalme, G. (2010). Une version électronique du LVF comparée avec d'autres ressources lexicales. *Langages*, 3, 193-220. [↑p141;p142](#)
- Halawi, B., Mourad, A., Otrok, H., & Damiani, E. (2018). Few are as good as many : An ontology-based tweet spam detection approach. *IEEE Access*, 6, 63890-63904. [↑p164](#)
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data : An overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23-34. [↑p99](#)
- Halté, P. (2020). Emojis, émoticônes, smileys? Proposition de classement terminologique selon des critères sémiotiques et énonciatifs. *Interfaces numériques*, 8(2), 386-386. [↑p30](#)
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(23), 146-162. [↑p125](#)
- Hasan, K. S., & Ng, V. (2014). Automatic keyphrase extraction : A survey of the state of the art. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1262-1273. [↑p53](#)
- Hathout, N., & Sajous, F. (2016). Wiktionnaire's wikicode GLAWified : A workable French machine-readable dictionary. *Proceedings of the tenth International Conference on Language Resources and Evaluation (LREC 2016)*, 1369-1376. [↑p247](#)
- He, L., Han, D., Zhou, X., & Qu, Z. (2020). The Voice of Drug Consumers : Online Textual Review Analysis Using Structural Topic Model. *International Journal of Environmental Research and Public Health*, 17(10), Article 3648. [↑p55](#)
- Healy, A. F., & Miller, G. A. (1970). The verb as the main determinant of sentence meaning. *Psychonomic Science*, 20(6), 372. [↑p141](#)
- Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. *COLING '92: Proceedings of the 14th conference on Computational linguistics*, 2, 539-545. [↑p163](#)
- Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, 72, 417-426. [↑p55](#)
- Hussenot, A. (2006). Vers une reconsidération de la notion d'usage des outils TIC dans les organisations : Une approche en termes d'« enaction ». *Pratiques et usages*

- organisationnels des sciences et technologies de l'information et de la communication*, 158-160. ↑[p13](#)
- Jenny, J. (1997). Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine. Etat des lieux et essai de classification. *Bulletin de Méthodologie Sociologique*, 54(1), 64-122. ↑[p45](#)
- Jo, T. (2019). *Text Mining : Concepts, Implementation, And Big Data Challenge* (Vol. 45). Springer. ↑[p51](#)
- Jung, H., & Lee, B. G. (2020). Research trends in text mining : Semantic network and main path analysis of selected journals. *Expert Systems with Applications*, 162(30), Article 113851. ↑[p52](#)
- Karami, A., Shah, V., Vaezi, R., & Bansal, A. (2020). Twitter speaks : A case of national disaster situational awareness. *Journal of Information Science*, 46(3), 313-324. ↑[p55](#)
- Kaufmann, A. (2014). *Using Natural Language Processing to Support Interview Analysis* [Master Thesis]. Freiderich-Alexander Univsersität, Erlangen-Nürnberg. ↑[p43](#); ↑[p49](#); ↑[p63](#)
- Kaufmann, A., Barcomb, A., & Riehle, D. (2020). Supporting Interview Analysis with Autocoding. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 752-761. ↑[p3](#); ↑[p49](#); ↑[p50](#); ↑[p51](#); ↑[p60](#); ↑[p61](#); ↑[p63](#); ↑[p225](#)
- Kleinknecht, A., & Van Der Panne, G. (2012). Predicting new product sales : The post-launch performance of 215 innovators. *International Journal of Innovation Management*, 16(2), 1-14. ↑[p9](#)
- Koltcov, S., Koltsova, O., & Nikolenko, S. (2014). Latent dirichlet allocation : Stability and applications to studies of user-generated content. *Proceedings of the 2014 ACM conference on Web science*, 161-165. ↑[p56](#)
- Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert systems with applications*, 40(10), 4065-4074. ↑[p164](#)
- Korhonen, A., & Briscoe, T. (2004). Extended lexical-semantic classification of English verbs. *Proceedings of the Computational Lexical Semantics Workshop at HLT-NAACL 2004*, 38-45. ↑[p145](#)
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms : A survey. *Information*, 10(4), Article 150. ↑[p124](#); ↑[p125](#); ↑[p126](#); ↑[p127](#); ↑[p129](#)
- Krippendorff, K. (1995). On the reliability of unitizing continuous data. *Sociological Methodology*, 25, 47-76. ↑[p41](#)
- Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability*. Article 43. ↑[p99](#)
- Kripper Schuler, K. (2005). *VerbNet : A broad-coverage, comprehensive verb lexicon* [PhD Thesis]. University of Pennsylvania. ↑[p144](#)
- Kuckartz, U. (2010). *Einführung in die computergestützte Analyse qualitativer Daten* (3rd ed.). VS Verlag für Sozialwissenschaften. ↑[p47](#)
- Kulviwat, S., Bruner II, G. C., Kumar, A., Nasco, S. A., & Clark, T. (2007). Toward a unified theory of consumer acceptance technology. *Psychology & Marketing*, 24(12), 1059-1084. ↑[p10](#)
- Kumar, R., Lahiri, B., Ojha, A. K., & Bansal, A. (2020). ComMA@ FIRE 2020 : Exploring Multilingual Joint Training across different Classification Tasks. *FIRE (Working Notes)*, 823-828. ↑[p128](#)
- Kumari, S. (2021). Nohate at hasoc2020 : Multilingual hate speech detection. *arXiv preprint*. ↑[p128](#)

- Lacity, M. C., & Janson, M. A. (1994). Understanding qualitative data : A framework of text analysis methods. *Journal of Management Information Systems*, 11(2), 137-155. [↑p35;p36](#)
- Lafourcade, M., & Joubert, A. (2008). JeuxDeMots : Un prototype ludique pour l'émergence de relations entre termes. *JADT'08: Journées internationales d'Analyse statistiques des Données Textuelles*, 657-666. [↑p157;p215](#)
- Langridge, M. E., & Ahern, K. (2003). A case report on using mixed methods in qualitative research. *Collegian*, 10(4), 32-36. [↑p40](#)
- Lapalme, G., & Le Pesant, D. (2020). Deux dictionnaires informatisés de Jean Dubois et Françoise Dubois-Charlier, leurs ultimes travaux. *Linx. Revue des linguistes de l'université Paris X Nanterre*, 80, Article 80. <https://doi.org/10.4000/linx.6671> [↑p141;p142](#)
- Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., Crabbé, B., Besacier, L., & Schwab, D. (2019). Flaubert : Unsupervised language model pre-training for french. *arXiv preprint arXiv:1912.05372*. [↑p139](#)
- Le Nagard-Assayag, E. L., Manceau, D., & Morin-Delerm, S. (2015). *Le marketing de l'innovation. Concevoir et lancer de nouveaux produits et services*. (3ème édition). Dunod. [↑p9](#)
- Lejeune, C. (2008). Au fil de l'interprétation. L'apport des registres aux logiciels d'analyse qualitative. *Swiss Journal of Sociology*, 34(3), 593-603. [↑p46](#)
- Lejeune, C. (2017). Analyser les contenus, les discours ou les vécus? À chaque méthode ses logiciels! In M. Santiago-Delefosse & M. Del Rio Carral (Éds.), *Les méthodes qualitatives en psychologie et sciences humaines de la santé* (Dunod, p. 203-224). Dunod. [↑p46;p47;p49](#)
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press. [↑p142;p144;p145](#)
- Lewis, D. D. (1998). Naive (Bayes) at forty : The independence assumption in information retrieval. In C. Nédellec & C. Rouveirol (Éds.), *Proceedings of ECML98, 10th European Conference on Machine Learning*. (p. 4-15). SpringerVerlag. [↑p126](#)
- L'Homme, M.-C., Kalichenko, T., Le Serrech, A., Laneville, M.-E., & Charlebois, J.-C. (2019). *Guide d'annotation des contextes du DiCoInfo et du DiCoEnviro*. [↑p149](#)
- Liu, Z., Huang, W., Zheng, Y., & Sun, M. (2010). Automatic keyphrase extraction via topic decomposition. *Proceedings of the 2010 conference on empirical methods in natural language processing*, 366-376. [↑p53](#)
- Loeser, F. (2019). *Modélisation probabiliste de l'influence des émotions sur l'acceptabilité des innovations* [Thèse de Doctorat]. Université Grenoble Alpes. [↑p249](#)
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication : Assessment and reporting of intercoder reliability. *Human communication research*, 28(4), 587-604. [↑p99](#)
- Lopez, P., & Romary, L. (2010). HUMB : Automatic Key Term Extraction from Scientific Articles in GROBID. *Proceedings of the 5th International Workshop on Semantic Evaluation*, 248-251. [↑p53](#)
- Luo, J. M., Vu, H. Q., Li, G., & Law, R. (2020). Topic modelling for theme park online reviews : Analysis of Disneyland. *Journal of Travel & Tourism Marketing*, 37(2), 272-285. [↑p55](#)
- Ma, Y., Audibert, L., & Nazarenko, A. (2009). Ontologies étendues pour l'annotation sémantique. *20es Journées francophones d'ingénierie des connaissances*, 205-216. [↑p164](#)

- MacPhail, C., Khoza, N., Abler, L., & Ranganathan, M. (2015). Process guidelines for establishing intercoder reliability in qualitative studies. *Qualitative research*, 16(2), 198-212. ↑[p44](#)
- Madhavan, A. (2014). *Semi Automated User Acceptance Testing using Natural Language Techniques* [Master Thesis]. Iowa State University. ↑[p58](#)
- Mandl, T., Modha, S., Shahi, G. K., Jaiswal, A. K., Nandini, D., Patel, D., Majumder, P., & Schäfer, J. (2020). *Overview of the HASOC track at FIRE 2020 : Hate speech and offensive content identification in Indo-European languages*. 14-17. ↑[p128](#)
- Mangabeira, W. C. (1996). Qualitative Sociology and Computer Programs : Advent and Diffusion of Computer-Assisted Qualitative Data Analysis Software (CAQDAS). *Current Sociology*, 44(3), 191-205. ↑[p45](#)
- Manguri, K. H., Ramadhan, R. N., & Amin, P. R. M. (2020). Twitter sentiment analysis on worldwide COVID-19 outbreaks. *Kurdistan Journal of Applied Research, Special Issue on Coronavirus*, 54-65. ↑[p55](#)
- Marathe, M., & Toyama, K. (2018). Semi-Automated Coding for Qualitative Research : A User-Centered Inquiry and Initial Prototypes. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 348. ↑[p3;p43;p48;p60;p61;p64;p70;p73;p74;p78;p225;p233](#)
- Marshall, H. (2002). What do we do when we code data. *Qualitative research journal*, 2(1), 56-70. ↑[p43;p45](#)
- Martin, A., & Stenner, P. (2004). Talking about drug use: What are we (and our participants) doing in qualitative research? *International Journal of Drug Policy*, 15(5), 395-405. ↑[p43](#)
- Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de La Clergerie, É. V., Seddah, D., & Sagot, B. (2019). Camembert : A tasty french language model. *arXiv preprint arXiv:1911.03894*. ↑[p139](#)
- Mayaffre, D., Pincemin, B., & Poudat, C. (2019). Explorer, mesurer, contextualiser. Quelques apports de la textométrie à l'analyse de discours. *Langue française*, 203(3), 101-115. ↑[p107](#)
- Mays, N., & Pope, C. (1995). Qualitative research : Rigour and qualitative research. *British Medical Journal*, 311(6997), 109-112. ↑[p43](#)
- McCartney, A., Hensman, S., & Longo, L. (2017). "How short is a piece of string?" The Impact of Text Length and Text Augmentation on Short-text Classification Accuracy. In J. McAuley & S. McKeever (Éds.), *Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science* (Vol. 2086, p. 102-114). CEUR-WS.org. ↑[p129](#)
- Meaney, J. A., Wilson, S., Chiruzzo, L., Lopez, A., & Magdy, W. (2021). Semeval 2021 task 7 : Hahackathon, detecting and rating humor and offense. *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, 105-119. ↑[p128](#)
- Mekaoui, M., Tisserant, G., Dodard, M., & Lopez, C. (2020). Extraction de tâches dans les e-mails : Une approche fondée sur les rôles sémantiques. *EGC*, 193-204. ↑[p143;p.145;p.148;p.157](#)
- Mel'čuk, I., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duclot. ↑[p247](#)
- Mertens, P. (2010). *Restrictions de sélection et réalisations syntagmatiques dans Dicovalence : Conversion vers un format utilisable en TAL*. <https://lirias.kuleuven.be/handle/123456789/269856> ↑[p142](#)
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. ↑[p125](#)

- Miles, M. B., & Weitzman, E. A. (1996). The state of qualitative data analysis software : What do we need? *Current Sociology*, 44(3), 206-224. ↑[p45](#)
- Miller, G. A. (1995). WordNet : A lexical database for English. *Communications of the ACM*, 38(11), 39-41. ↑[p164:p215](#)
- Mironćzuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, 106, 36-54. ↑[p121;p122](#)
- Mishraa, A. K., Saumyab, S., & Kumara, A. (2020). IIIT_DWD@ HASOC 2020 : Identifying offensive content in Indo-European languages. *Fire (Working Notes)*. ↑[p128](#)
- Mondary, T. (2011). *Construction d'ontologies à partir de textes. L'apport de l'analyse de concepts formels*. [Thèse de Doctorat]. Université Paris-Nord-Paris XIII. ↑[p163;p164](#)
- Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, 58(302), 415-434. ↑[p127](#)
- Moscarola, J., & Boughzala, Y. (2016). *Analyser les corpus d'avis en ligne : Analyse lexicale exploratoire et/ou modélisation sémantique?* JADT 2016 : 13ème Journées internationales d'Analyse statistiques des Données Textuelles, Nice, France. ↑[p39](#)
- Mountford, J. B. (2018). Topic modeling the red pill. *Social Sciences*, 7(3), 42. ↑[p55](#)
- Mouriño-García, M., Pérez-Rodríguez, R., Anido-Rifón, L., & Gómez-Carballa, M. (2016). Bag-of-concepts document representation for Bayesian text classification. *2016 IEEE International Conference on Computer and Information Technology (CIT)*, 281-288. ↑[p124](#)
- Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization : State-of-the-art. *Information Processing & Management*, 56(6), Article 102088. ↑[p53](#)
- Neale, J., Allen, D., & Coombes, L. (2005). Qualitative research methods within the addictions. *Addiction*, 100(11), 1584-1593. ↑[p43](#)
- Negura, L. (2006). L'analyse de contenu dans l'étude des représentations sociales. *SociologieS [En ligne], Théories et recherches (mis en ligne le 22 octobre 2006)*. <http://journals.openedition.org/sociologies/993> ↑[p39](#)
- Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2018). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods & Research*, 50(1), 202-237. ↑[p73](#)
- Nemes, L., & Kiss, A. (2021). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, 5(1), 1-15. ↑[p55](#)
- Neuhaus, F. (2018). What is an Ontology? *arXiv preprint arXiv:1810.09171*. ↑[p259](#)
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™. *L'année psychologique*, 101(3), 447-462. ↑[p108;p149](#)
- Nguyen, T. D., & Kan, M.-Y. (2007). Keyphrase extraction in scientific publications. *International conference on Asian digital libraries*, 317-326. ↑[p53](#)
- Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, 43(1), 88-102. ↑[p55](#)
- Nouwen, R., Brasoveanu, A., van Eijck, J., & Visser, A. (2010). Dynamic semantics. In E. N. Zalta (Éd.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). <https://seop.illc.uva.nl/entries/dynamic-semantics/> ↑[p79](#)
- Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101 : A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05. ↑[p162](#)

- OCDE/Eurostat. (2005). *Oslo Manual : Guidelines for Collecting and Interpreting Innovation Data* (3rd Edition). Éditions OCDE. ↑[p.8](#)
- O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research : Debates and practical guidelines. *International Journal of Qualitative Methods*, 19, 1-13. ↑[p44;p99](#)
- Paillé, P., & Mucchielli, A. (2016). *L'analyse qualitative en sciences humaines et sociales* (4ème édition). Armand Colin. ↑[p36;p37;p42](#)
- Pantel, P., & Pennacchiotti, M. (2006). Espresso : Leveraging generic patterns for automatically harvesting semantic relations. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, 113-120. ↑[p164](#)
- Paredes, P., Rufino Ferreira, A., Schillaci, C., Yoo, G., Karashchuk, P., Xing, D., Cheshire, C., & Canny, J. (2017). Inquire : Large-scale early insight discovery for qualitative research. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1562-1575. ↑[p78](#)
- Patel, D., & Srivastava, T. (2014). Ant colony optimization model for discrete tomography problems. *Proceedings of the Third International Conference on Soft Computing for Problem Solving*, 785-792. ↑[p127](#)
- Patil, R. S., Fikes, R. E., Patel-Schneider, P. F., McKay, D., Gruber, T., & Neches, R. (1998). The DARPA knowledge sharing effort : Progress report. *Readings in agents*, 243. ↑[p160](#)
- Pearson, E. S. (1925). Bayes' theorem, examined in the light of experimental sampling. *Biometrika*, 17, 388-442.- ↑[p126](#)
- Pietsch, A.-S., & Lessmann, S. (2018). Topic modeling for analyzing open-ended survey responses. *Journal of Business Analytics*, 1(2), 93-116. ↑[p56](#)
- Pincemin, B. (2011). Sémantique interprétative et textométrie–Version abrégée. *Corpus*, 10, 259-269. ↑[p107](#)
- Pizelle, P., Hoffmann, J., Aubouy, M., & Verchère, C. (2014). *Innover par les usages*. Editions d'Innovations. ↑[p11;p12;p13](#)
- Poibeau, T. (2014). Le traitement automatique des langues pour les sciences sociales. *Réseaux*, 6(188), 25-51. ↑[p73](#)
- Polguere, A. (2009). Lexical systems : Graph models of natural language lexicons. *Language resources and evaluation*, 43(1), 41-55. ↑[p247](#)
- Pradet, Q. (2015). *Annotation en rôles sémantiques du français en domaine spécifique* [PhD Thesis]. Université Paris Diderot (Paris 7). ↑[p143;p144;p145;p148](#)
- Pradet, Q., De Chalendar, G., & Desormeaux, J. B. (2014). WoNeF, an improved, expanded and evaluated automatic French translation of WordNet. In H. Orav, C. Fellbaum, & P. Vossen (Éds.), *Proceedings of the Seventh Global Wordnet Conference* (p. 32-39). University of Tartu Press. ↑[p215](#)
- Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, 5(2), 221-232. ↑[p127](#)
- Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019). An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges. *2019 International Engineering Conference (IEC)*, 200-204. ↑[p123](#)
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza : A Python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*. ↑[p188](#)
- Qi, X., & Davison, B. D. (2009). Web page classification : Features and algorithms. *ACM computing surveys (CSUR)*, 41(2), Article 12. ↑[p121](#)

- Raja, R., Srivastavab, S., & Saumyac, S. (2021). *NSIT & IITDWD@ HASOC 2020 : Deep learning model for hate-speech identification in Indo-European languages*. FIRE '20, Forum for Information Retrieval Evaluation. ↑[p128](#)
- Rapley, T. (2011). Some pragmatics of data analysis. In D. Silverman (Éd.), *Qualitative research* (Vol. 3, p. 273-290). SAGE. ↑[p43](#)
- Ray Chowdhury, J., Caragea, C., & Caragea, D. (2019). Keyphrase extraction from disaster-related tweets. In L. Liu & R. White (Éds.), *WWW'19: The Web Conference* (p. 1555-1566). Association for Computing Machinery. ↑[p53](#)
- Richards, L. (2002). Qualitative computing—a methods revolution? *International Journal of Social Research Methodology*, 5(3), 263-276. ↑[p78](#)
- Richards, L., & Richards, T. (1994). From filing cabinet to computer. In A. Bryman & R. G. Burgess (Éds.), *Analyzing qualitative data* (p. 146-194). Routledge. ↑[p43](#)
- Rietz, T. (2021). *Designing AI-Based Systems for Qualitative Data Collection and Analysis* [PhD Thesis]. Karlsruhe Institut für Technologie (KIT). ↑[p61;p66;p70;p78;p226](#)
- Rietz, T., & Maedche, A. (2021). Cody : An AI-Based System to Semi-Automate Coding for Qualitative Research. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Article 394. ↑[p43;p66](#)
- Rihoux, B. (2017). Les boîtes à outils de l'analyse de données textuelles (ADT) : Des chantiers aussi prometteurs que périlleux. In P.-M. Daigneault & F. Pétry (Éds.), *L'analyse textuelle des idées, du discours et des pratiques politiques* (p. XVII-XXIII). Presses de l'Université de Laval. ↑[p35](#)
- Robert, A. D., & Bouillaguet, A. (2007). *L'analyse de contenu*. Presses universitaires de France. ↑[p38](#)
- Roy, N., & Garon, R. (2013). Hors thème Étude comparative des logiciels d'aide à l'analyse de données qualitatives : De l'approche automatique à l'approche manuelle. *Recherches qualitatives*, 32(1), 157-178. ↑[p46](#)
- Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. (2018). More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, 12(2-3), 140-157. ↑[p125](#)
- Sagot, B., & Fišer, D. (2008). Building a free French wordnet from multilingual resources. *OntoLex*. ↑[p215](#)
- Sahgal, D., & Ramesh, A. (2002). On Road Vehicle Detection Using Gabor Wavelet Features with Various Classification Techniques. *Proceedings of the 14th International Conference on Digital Signal Processing Proceedings. DSP 2002*. DSP 2020, Santorini, Greece. ↑[p127](#)
- Sai, S., & Sharma, Y. (2020). Siva@ HASOC-Dravidian-CodeMix-FIRE-2020 : Multilingual Offensive Speech Detection in Code-mixed and Romanized Text. *FIRE (Working Notes)*, 336-343. ↑[p128](#)
- Saldaña, J. (2009). An Introduction to codes and coding. In *The Coding manual for Qualitative Research*. SAGE. ↑[p40](#)
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd Edition). Sage. ↑[p43](#)
- Salguero, A., & Espinilla, M. (2018). A flexible text analyzer based on ontologies: An application for detecting discriminatory language. *Language Resources and Evaluation*, 52, 185-215. ↑[p164](#)
- Sangounpao, K., & Muenchaisri, P. (2019). Ontology-Based Naive Bayes Short Text Classification Method for a Small Dataset. *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 53-58. ↑[p129](#)

- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing*. ↑[p108;p149](#)
- Schreier, M. (2014). Qualitative content analysis. In U. Flick (Éd.), *The SAGE handbook of qualitative data analysis* (p. 170-183). SAGE. ↑[p38](#)
- Shaban, T. A., Hexter, L., & Choi, J. D. (2017). Event Analysis on the 2016 US Presidential Election using social media. *International conference on social informatics*, 201-217. ↑[p53](#)
- Shao, Y., Taylor, S., Marshall, N., Morioka, C., & Zeng-Treitler, Q. (2018). Clinical text classification with word embedding features vs. Bag-of-words features. *2018 IEEE International Conference on Big Data (Big Data)*, 2874-2878. ↑[p126](#)
- Sikelis, K., Tsekouras, G. E., & Kotis, K. (2021). Ontology-Based Feature Selection : A Survey. *Future Internet*, 13(6), 158. ↑[p164](#)
- Silberztein, M. (1990). Le dictionnaire électronique des mots composés. *Langue française*, 87, 71-83. ↑[pxi;p150](#)
- Silberztein, M. (2020). Les dictionnaires DEM et LVF au service des humanités numériques. *Linx [En ligne]*, 80. ↑[p143](#)
- Sinoara, R. A., Antunes, J., & Rezende, S. O. (2017). Text mining and semantics : A systematic mapping study. *Journal of the Brazilian Computer Society*, 23(1), 1-20. ↑[p52](#)
- Slimani, T. (2015). Ontology development : A comparing study on tools, languages and formalisms. *Indian Journal of Science and Technology*, 8(24), 1-12. ↑[p161163](#)
- Somodevilla, M. J., Vilariño Ayala, D., & Pineda, I. (2018). An Overview on Ontology Learning Tasks. *Computación y Sistemas [Online]*, 22(1), 137-146. ↑[p163](#)
- Soumya George, K., & Joseph, S. (2014). Text classification by augmenting bag of words (BOW) representation with co-occurrence feature. *IOSR Journal of Computer Engineering (IOSR-JCE)*, 16(1), 34-38. ↑[p124](#)
- Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1), 11-21. ↑[p124](#)
- Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer : A computer approach to content analysis*. MIT Press. ↑[p47](#)
- Suominen, A., Toivanen, H., & Seppänen, M. (2017). Firms' knowledge profiles : Mapping patent data with unsupervised learning. *Technological Forecasting and Social Change*, 115, 131-142. ↑[p55](#)
- Swier, R. S., & Stevenson, S. (2005). Exploiting a verb lexicon in automatic semantic role labelling. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 883-890. ↑[p148](#)
- Swier, R. S., & Stevenson, S. (2004). Unsupervised semantic role labelling. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 95-102. ↑[p148](#)
- Tan, A.-H. (1999). Text mining : The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, 8, 65-70. ↑[p44](#)
- Tauveron, M. (2012). *Variation diastématique du sens lexical Pour une représentation non-hiérarchique du lexique*. Atelier DIA du français. La Dia-Variation en français actuel Des corpus aux ouvrages de référence (Dictionnaires/Grammaires), Gand, Belgique. ↑[p27](#)
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing. ↑[p81](#)
- Tolone, E. (2009). Les tables du Lexique-Grammaire au format TAL. *MajecSTIC 2009*. ↑[p141](#)

- Tolone, E. (2011). Maintenance du Lexique-Grammaire : Formules définitives et arbre de classement. *Traitement Automatique des Langues*, 52(3), 153-190. ↑[p141](#)
- Tolone, E., Voyatzi, S., & Leclère, C. (2010). Constructions définitives des tables du Lexique-Grammaire. In P. L. Popović, C. Krstev, D. Vitas, & G. Pavlović-Lažetić, (Éds.), *Actes du 29ème Colloque international sur le Lexique et la Grammaire (LGC'10)* (p. 321-331). ↑[p141](#)
- van den Eynde, K., & Mertens, P. (2003). La valence : L'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, 13, 63-104. ↑[p142](#)
- van der Eynde, K., & Blanche-Benveniste, C. (1978). Syntaxe et mécanismes descriptifs : Présentation de l'approche pronominale. *Cahiers de Lexicologie Paris*, 32(1), 3-27. ↑[p142](#)
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology : Toward a Unified View. *MIS Quarterly*, 27(3), 425-478. ↑[p10](#)
- Verleysen, M., & François, D. (2005). The curse of dimensionality in data mining and time series prediction. In J. Cabestany, A. Prieto, & F. Sandoval (Éds.), *International work-conference on artificial neural networks* (p. 758-770). Springer. ↑[p122](#)
- Villatoro-Tello, E., Ramírez-de-la-Rosa, G., Kumar, S., Parida, S., & Motlíček, P. (2020). Idiap and UAM Participation at MEX-A3T Evaluation Campaign. *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)*, 252-257. ↑[p128](#)
- Villon de Benveniste, G. (2013). Innovation de rupture : Concept fondamental ou slogan creux ? *Entreprendre & Innover*, 18(2), 7-13. ↑[p9](#)
- Voynnet Fourboul, C., & Point, S. (2001). Le processus de (dé) codage des données qualitatives en gestion des ressources humaines. *Actes du XIIe congrès de l'AGRH*, 2. ↑[p37;p38](#)
- Voynnet-Fourboul, C. (2012). Ce que «analyse de données qualitatives» veut dire. *Revue internationale de Psychosociologie*, XVIII(44), 71-88. ↑[p35;p39](#)
- Wallach, H. (2016). Computational Social Science : Toward a Collaborative Future. In M. H. Alvarez (Éd.), *Computational Social Science: Discovery and Prediction* (p. 307-316). Cambridge University Press. Draft retrieved at <https://www.microsoft.com/en-us/research/publication/computational-social-science-toward-a-collaborative-future/> ↑[p73](#)
- Wan, X., & Xiao, J. (2008). Single Document Keyphrase Extraction Using Neighborhood Knowledge. *AAAI'08: Proceedings of the 23rd national conference on Artificial intelligence*, 2, 855-860. ↑[p53](#)
- Wang, C., Pastore, F., Goknil, A., & Briand, L. (2020). Automatic generation of acceptance test cases from use case specifications : An nlp-based approach. *IEEE Transactions on Software Engineering*. ↑[p58](#)
- Wang, J., Wang, Z., Zhang, D., & Yan, J. (2017). Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, 2916-2921. ↑[p129](#)
- Wanlin, P. (2007). L'analyse de contenu comme méthode d'analyse qualitative d'entretiens : Une comparaison entre les traitements manuels et l'utilisation de logiciels. *Recherches qualitatives*, 3(3), 243-272. ↑[p49](#)
- Weichbroth, P., & Baj-Rogowska, A. (2019). Do online reviews reveal mobile application usability and user experience? The case of WhatsApp. *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 747-754. ↑[p55;p58](#)

- Wiedemann, G. (2013). Opening up to big data : Computer-assisted analysis of textual data in social sciences. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 14(2), Article 23. [↑p45;p47;p48;p52;p73](#)
- Wiedemann, G. (2016). *Text mining for qualitative data analysis in the social sciences. A study on democratic discours in Germany*. (Vol. 1). Springer. [↑p42;p46;p47;p50;p51;p52;p53](#)
- Woods, M., Macklin, R., & Lewis, G. K. (2016). Researcher reflexivity : Exploring the impacts of CAQDAS use. *International Journal of Social Research Methodology*, 19(4), 385-403. [↑p45](#)
- Xia, T., & Chen, X. (2020). A discrete hidden Markov model for SMS spam detection. *Applied Sciences*, 10(14), Article 5011. [↑p121](#)
- Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1), 48-59. [↑p126](#)
- Yadav, V., & Bethard, S. (2019). A survey on recent advances in named entity recognition from deep learning models. In E. Bender, L. Derczynski, & P. Isabelle (Éds.), *Proceedings of the 27th International Conference on Computational Linguistics* (p. 2145-2158). Association for Computational Linguistics. [↑p54](#)
- Yan, J. L. S., McCracken, N., Zhou, S., & Crowston, K. (2014). Optimizing features in active machine learning for complex qualitative content analysis. *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 44-48. [↑p63](#)
- Yang, F.-J. (2019). An extended idea about decision trees. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, 349-354. [↑p127](#)
- Yang, Y., & Chute, C. G. (1994). An example-based mapping method for text catégorization and retrieval. *ACM Transactions on Information Systems*, 12(3), 252-277. [↑p127](#)
- Yogarajan, V., Gouk, H., Smith, T. C., Mayo, M., & Pfahringer, B. (2020). Comparing high dimensional word embeddings trained on medical text to bag-of-words for predicting medical codes. In P. Sitek, M. Petranik, M. KrÅ³tkiewicz, & C. Srinilta (Éds.), *Proceedings of 12th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2020) LNCS 12033* (p. 97-108). Springer. https://doi.org/10.1007/978-3-030-41964-6_9 [↑p264](#)
- Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *The Qualitative Report*, 16(3), 730-744. [↑p52;p73](#)
- Zharmagambetov, A. S., & Pak, A. A. (2015). Sentiment analysis of a document using deep learning approach and decision trees. *2015 Twelve international conference on electronics computer and computation (ICECCO)*, 1-4. [↑p125](#)
- Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A., Vongsangnak, W., & Shen, B. (2013). Biomedical text mining and its applications in cancer research. *Journal of biomedical informatics*, 46(2), 200-211. [↑p52](#)

ANNEXES

ANNEXE 1 : INTERFACES ET ARCHITECTURES

1. Interface et architecture de *Cody* (Rietz, 2021)

Les figures ci-dessous – ainsi que leurs légendes – sont des captures d’écran de la thèse issues de la thèse de Rietz (2021, pp. 66;68).

Architecture :

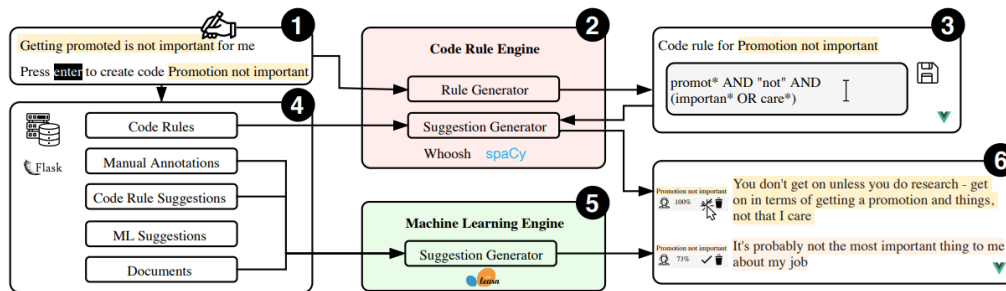


Figure 3.13: System architecture. (1) User makes an annotation, (2) code rule engine generates rule suggestion, (3) new rule is displayed for user review, (4) save triggers suggestion generator to search indexed document for occurrences, (4) and sends an update to suggestions in the database. (5) Machine learning engine retrains model and makes suggestions, (6) displayed for user review in the front end.

FIGURE EN ANNEXE 1 : Architecture de Cody

Vue d’ensemble de l’interface :

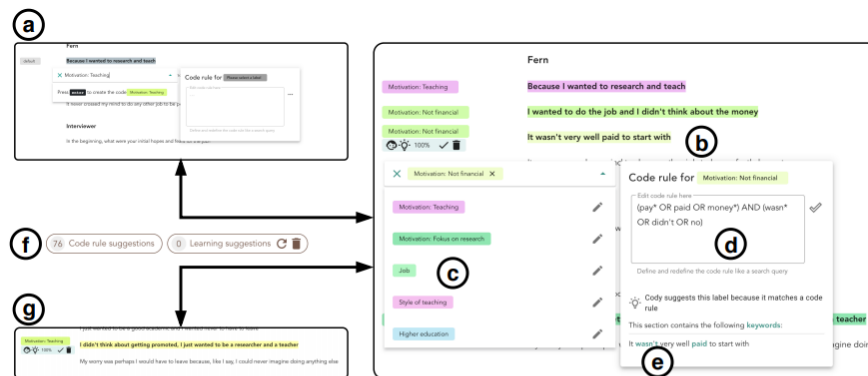


Figure 3.12: Coding workflow with Cody. Users make a new annotation and define a new code (a) which opens the code menu (b). Users may add codes to or delete codes from an annotation or edit a code (c). Cody suggests a possible code rule that users can edit (d). When clicking on suggestions to open the label menu, Cody shows explanations (e). Code rules are applied on saving to create suggestions and can be accepted/rejected by clicking the respective icons (g). The number of available suggestions is shown in the menu bar (f), where users can trigger ML model retraining (refresh icon) or delete all ML-based suggestions (trashcan icon).

FIGURE EN ANNEXE 2 : Vue d’ensemble de l’interface de Cody

Focalisation sur les modules de codage :

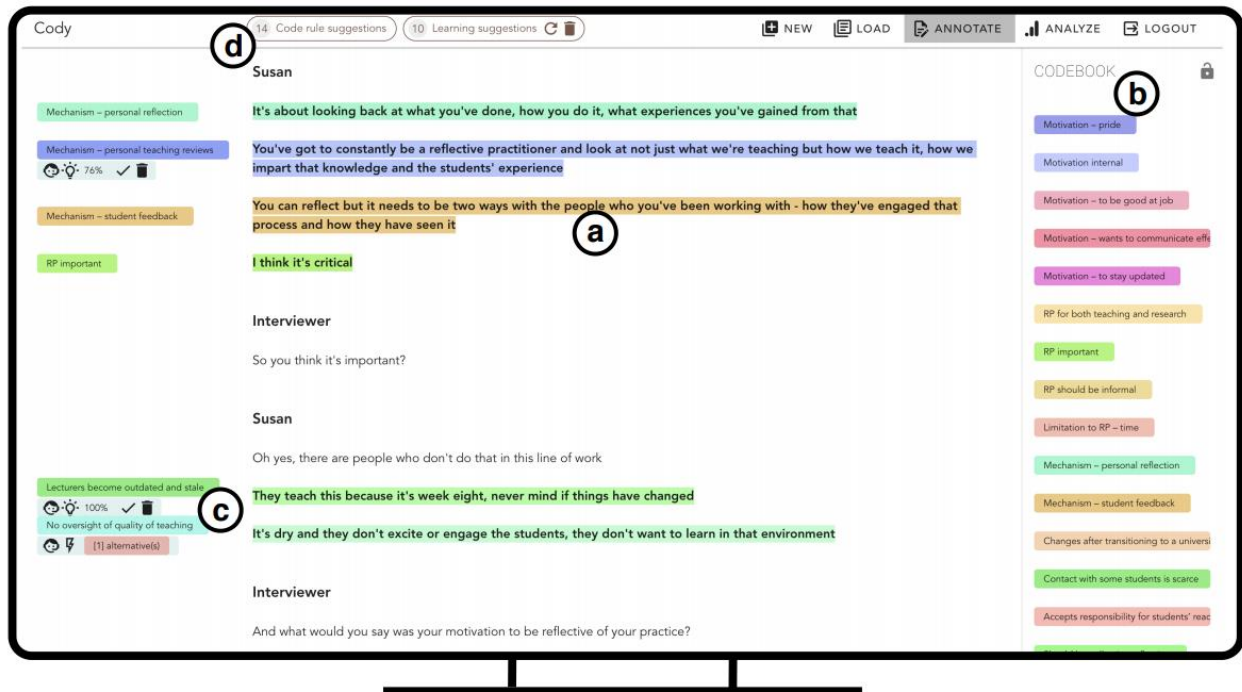


Figure 3.11: Final user interface of Cody. (a) main annotation view, (b) codebook sortable via drag-and-drop, (c) Code suggestion with confidence and accept/reject buttons. Below, Cody highlights multiple alternative suggestions for a section, (d) Number of rule- and ML-based suggestions

FIGURE EN ANNEXE 3 : Focalisation sur les modules de codage de Cody

2. Architecture du système proposé

Légende :

- C_x : Classe de l'ontologie
- R_x : Règle
- CLS_x : Catégorie lexico-sémantique
- ➔x.X : Critères à attribuer

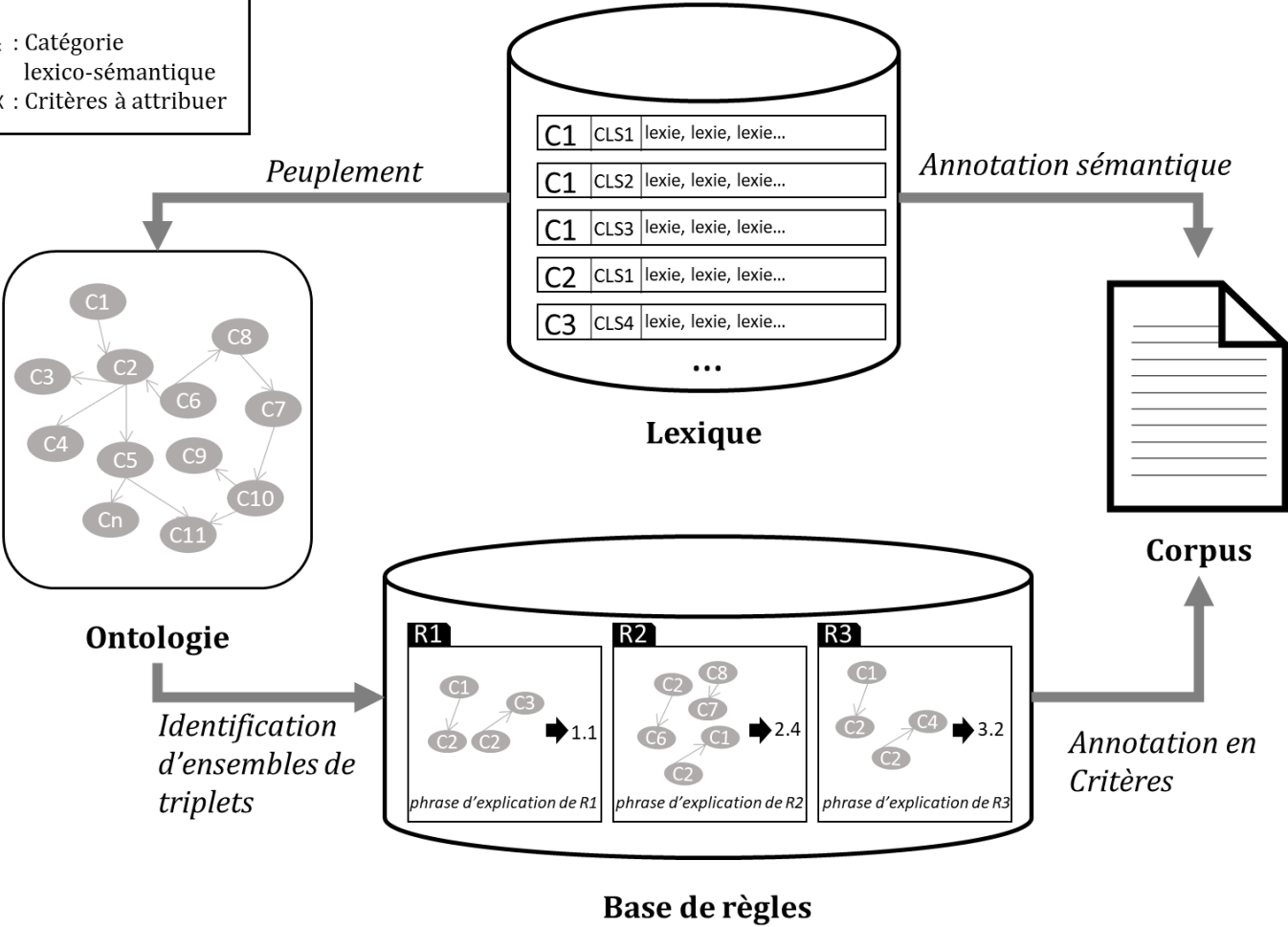


FIGURE EN ANNEXE 4 : Architecture proposée

3. Interface de la plateforme d'analyse de l'entreprise Ixiade

The screenshot displays a window titled "DOLOSPORT_MalDos_CDE" with the subtitle "En cours de traitement". A "Zoom" button is located in the top right corner. The main text area contains a question: "Dans quelles situations utiliseriez-vous cette solution et comment (à quels moments, dans quels objectifs,...) ?". Below this, there are two user comments: one from Doriane and one from Christophe. The Doriane comment is highlighted with orange boxes, and the Christophe comment is highlighted with blue boxes. Below the text is a "Verbatims" section with a radio button for "Classification automatique" (selected) and a text box containing a quote from Doriane. The "Certitude" is 0. A tooltip for the "Certitude" value shows "2.1 - (29) : L'utilisateur s'exprime sur la facilité/difficulté d'utilisation du concept." Below this are "Tags" for "Simplicité utilisation (2.1)" and a "Ajouter +" button. At the bottom of the verbatim box are buttons for "Choisir comme contenu illustratif", "Supprimer", and "Inclassifiable". Below the verbatim box is a radio button for "Classification automatique corrigée" and a "Valider" button.

FIGURE EN ANNEXE 5 : Interface plateforme d'analyse Ixiade – Codage sur texte entier

Textes Verbatims Classification par verbatims

CAUTIC®

Intérêt spontané (1.1) Compréhension technique (1.2) Comparaison existant (1.3) Intérêt fonctions (1.4) Usage courant (1.5)

Simplicité utilisation (2.1) Concurrence habitudes (2.2) Comparaison pratiques (2.3) Résolution problèmes (2.4)

Organisation personnelle (2.5) Cibles (3.1) Influence rôle (3.2) Adéquation valeurs (3.3) Imaginaire d'appropriation (3.4)

Extensions d'usage (3.5) Arrivée sur le marché (4.1) Adéquation aux relations (4.2) Adéquation place filère/famille (4.3)

Adéquation organisation/évolutions (4.4) Prix consenti (4.5) Inclassifiable (Inclassifiable)

FMC

Frein (F) Motivation (M) Condition (C)

Statut

Classification automatique corrigée Classification analyste Classification analyste corrigée Classification automatique

Contenu illustratif

Contenu illustratif

Participant

DOLOSPORT_MalDos_CDE_oiQAIBV DOLOSPORT_PasMalDos_JSJ_SCA_CmDL08W

Selectionnez un participant v Selectionnez un participant v

Classification automatique

« Est-ce que ceci décharge rapidement le téléphone ? » - Participant : MVT v

Certitude : 0 ⓘ DOLOSPORT_PasMalDos_JSJ_SCA_CmDL08W

Tags Compréhension technique (1.2) x Frein (F) x Ajouter +

Choisir comme contenu illustratif Supprimer Inclassifiable

Classification automatique

« Par contre comment cela va fonctionner (le son est-il intégré au vêtement ? » - Participant : SophieNJI v

Certitude : 0 ⓘ DOLOSPORT_PasMalDos_JSJ_SCA_CmDL08W

Tags Compréhension technique (1.2) x Ajouter +

Choisir comme contenu illustratif Supprimer Inclassifiable

Classification automatique corrigée

« Le fonctionnement pour moi n'est pas très clair. » - Participant : emma v

Certitude : 0 ⓘ DOLOSPORT_PasMalDos_JSJ_SCA_CmDL08W

Tags Compréhension technique (1.2) x Ajouter +

Choisir comme contenu illustratif Supprimer Inclassifiable

FIGURE EN ANNEXE 6 : Interface plateforme d'analyse Ixiade – Verbatim filtrés

ANNEXE 2 : CORPUS

4. Volumétrie du corpus non filtré

Etude	Analyse	Illustratif	Total	Pourcentage
EC1_Elec	315	119	434	7,99%
EC2_Elec	286	97	383	7,05%
EC3_Elec	180	94	274	5,04%
EC4_Elec	262	121	383	7,05%
EC5_Elec	931	159	1090	20,06%
EC6_Elec	225	93	318	5,85%
EC7_Elec	74	114	188	3,46%
EC8_Elec	51	116	167	3,07%
Total électrique	2324	913	3237	59,58%
EC1_Non Elec (électroménager)	650	57	707	13,01%
EC2_Non Elec (médecine)	147	67	214	3,94%
EC3_Non Elec (gérontologie)	214	63	277	5,10%
EC4_Non Elec (pâtisserie)	295	183	478	8,80%
ENC1_Non Elec (médecine)	93	41	134	2,47%
ENC2_Non Elec (médecine)	56	23	79	1,45%
ENC3_Non Elec (automatisme)	0	43	43	0,79%
ENC4_Non Elec (médecine)	46	167	213	3,92%
ENC5_Non Elec (pâtisserie)	0	51	51	0,94%
Total Non Electrique	1501	695	2196	40,42%

TABLEAU EN ANNEXE 1 : Répartition des verbatim du corpus non filtré par domaines

↑p91

		Verbatim analyse	Verbatim illustratifs	Total			Verbatim analyse	Verbatim illustratifs	Total
NIVEAU 1	1.1	160	55	215	NIVEAU 3	3.1	187	56	243
	1.2	191	75	266		3.2	107	38	145
	1.3	210	73	283		3.3	163	64	227
	1.4	312	196	508		3.4	119	110	229
	1.5	148	56	204		3.5	144	66	210
	Non précisé	22	28	50		Non précisé	19	42	61
	Total Niveau	1043	483	1526		Total Niveau	739	376	1115
NIVEAU 2	2.1	125	57	182	NIVEAU 4	4.1	178	81	259
	2.2	286	83	369		4.2	84	39	123
	2.3	368	133	501		4.3	45	36	81
	2.4	301	99	400		4.4	139	47	186
	2.5	212	55	267		4.5	161	54	215
	Non précisé	56	49	105		Non précisé	88	16	104
	Total Niveau	1348	476	1824		Total Niveau	695	273	968

TABLEAU EN ANNEXE 2 : Répartition des verbatim du corpus non filtré par Niveaux et/ou Critères attribués

	Femmes	Hommes	Inconnu
Répartition	25,00%	62,30%	12,70%
Professionnels	14,75%	67,76%	17,49%
Particuliers	52,17%	47,83%	0,00%
Moyenne d'âge	44,97	42,11	-

TABEAU EN ANNEXE 3 : Participants des études du corpus non filtré : ratios hommes/femmes, professionnels/particuliers et moyenne d'âge

↑p91

5. Eléments chiffrés et illustratifs du processus de filtrage du corpus

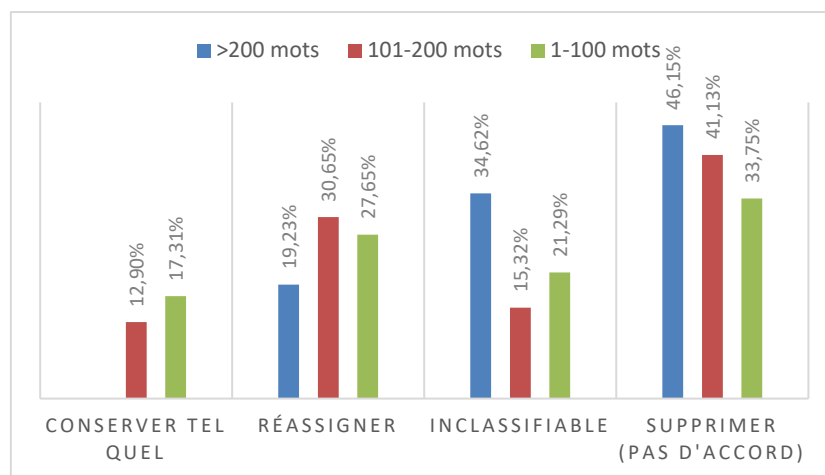


FIGURE EN ANNEXE 7 : Décisions prises à la suite du filtrage par accord inter-juge par rapport au nombre de mots par verbatim

↑p96

		Niveaux d'origine (tels qu'attribués dans les archives)			
		1	2	3	4
Critères attribués suite à l'évaluation inter-juge	1	0,39	0,16	0,11	0,09
	2	0,21	0,38	0,18	0,13
	3	0,04	0,04	0,30	0,09
	4	0,03	0,04	0,04	0,30
	Inclassifiable	0,18	0,23	0,15	0,21
	Pas d'accord	0,15	0,16	0,22	0,19

Les valeurs indiquées correspondent au pourcentage des verbatim par rapport au total par niveau dans le corpus non filtré

↑p100

TABEAU EN ANNEXE 4 : Matrice de confusion entre les classifications des verbatim en Niveaux dans les archives et celles obtenues par évaluation inter-juge

Critères d'origine (tels qu'attribués dans les archives)

	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5	Aucun	
1.1	0,20	0,12	0,05	0,05	0,03	0,04	0,01	0,01	0,03	0,05	0,01	0,02	0,12	0,03	0,01	0,01	0,07				0,02	0,01
1.2	0,04	0,22	0,04	0,01	0,02		0,03	0,02	0,02	0,02										0,02		0,01
1.3		0,04	0,29	0,03		0,04	0,03	0,04		0,01			0,01	0,04	0,01	0,03						0,03
1.4	0,08	0,07	0,01	0,14	0,02	0,02	0,02	0,02	0,03	0,02		0,02	0,01		0,04		0,03				0,02	0,02
1.5					0,02		0,01		0,01	0,02	0,03											0,01
2.1	0,01	0,01	0,01	0,03	0,03	0,14	0,02	0,03		0,03	0,01		0,03				0,03	0,04	0,05			0,01
2.2	0,03	0,04	0,01	0,02			0,03	0,01	0,03	0,01		0,05	0,01		0,03				0,04			0,01
2.3	0,04	0,04	0,07	0,04	0,03	0,02	0,07	0,14	0,03	0,05	0,04	0,02	0,03	0,04		0,03		0,03		0,02		0,03
2.4	0,03	0,01	0,02	0,08	0,03	0,04	0,04	0,05	0,29	0,05	0,04	0,02	0,04	0,01	0,01		0,03	0,04	0,02			0,02
2.5	0,04		0,01	0,01			0,03	0,02	0,03	0,03	0,01	0,02	0,01		0,03	0,01		0,04	0,02	0,05		0,05
3.1	0,06		0,01	0,05	0,02	0,04	0,05	0,02		0,02	0,37	0,12	0,05	0,03	0,04	0,01		0,04	0,04	0,05		0,08
3.2												0,05	0,01								0,02	
3.3			0,01				0,01					0,02	0,09	0,01		0,03	0,03	0,04	0,02	0,02	0,01	
3.4	0,01				0,02						0,03		0,01	0,57								
3.5					0,03		0,01			0,01	0,01			0,01	0,04							
4.1					0,03	0,02							0,06			0,20	0,03	0,04				0,02
4.2							0,01		0,01			0,02	0,01				0,17	0,04			0,02	0,01
4.3					0,02								0,01					0,04	0,02			
4.4	0,01				0,02		0,01		0,01	0,01				0,01		0,06				0,09		
4.5	0,03		0,01				0,01	0,01	0,02						0,06		0,04				0,47	0,03
Inclassifiable	0,06	0,18	0,26	0,18	0,23	0,22	0,27	0,25	0,26	0,21	0,03	0,23	0,21	0,09	0,30	0,26	0,10	0,15	0,23	0,16	0,32	
Pas d'accord	0,35	0,26	0,21	0,34	0,48	0,40	0,39	0,37	0,26	0,42	0,41	0,40	0,28	0,18	0,42	0,33	0,47	0,52	0,43	0,18	0,37	

Les valeurs indiquées correspondent au pourcentage des verbatim par rapport au total par critère dans le corpus non filtré

TABLEAU EN ANNEXE 5 : Matrice de confusion entre les classifications des verbatim en Critères dans les archives et celles obtenues par évaluation inter-juge

Aide à l'interprétation des deux tableaux suivants :

Soit le tableau :

	A	B	C
A			
B			x
C			

Et les codeurs : a, b et c

On considère la liste de toutes les annotations de toutes les paires codeur1-codeur2 telles que :

codeur1	codeur2
a	b
b	a
a	c
c	a
b	c
c	b

Annotation_{c1} - Annotation_{c2}
 Annotation_{c1} - Annotation_{c2}
 Annotation_{c1} - Annotation_{c2}
 ...
 Annotation_{c1} - Annotation_{c2}

$$X = (\text{nombre de couples [Annotation}_{c1} = C - \text{Annotation}_{c2} = B]} + \text{nombre de couples [Annotation}_{c1} = B - \text{Annotation}_{c2} = C]}) / (\text{nombre de Annotation}_{c1} = B + \text{nombre de Annotation}_{c2} = B)$$

En d'autres termes, les chiffres indiqués correspondent au pourcentage de fois où la classe indiquée en ligne est incluse dans une relation d'accord/désaccord inter-juge binaire avec la classe indiquée en colonne.

	1	2	3	4	Inclassifiable
1	0,47	0,25	0,08	0,05	0,15
2		0,49	0,07	0,06	0,14
3			0,46	0,06	0,11
4				0,41	0,10
Inclassifiable					0,83

TABLEAU EN ANNEXE 6 : Matrice de confusion de l'accord inter-juge sur les Niveaux dans le processus de filtrage du corpus

	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5	Inclassifiable	
1.1	0,14	0,08	0,04	0,07	0,02	0,02	0,05	0,05	0,04	0,02	0,03	0,01	0,04	0,02	0,01	0,02	0,00	0,01	0,01	0,01	0,01	0,04
1.2		0,26	0,06	0,04	0,02	0,03	0,02	0,04	0,02	0,02	0,01	0,00	0,01	0,00	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,04
1.3			0,32	0,04	0,03	0,02	0,03	0,06	0,02	0,03	0,02	0,00	0,01	0,02	0,02	0,01	0,00	0,00	0,02	0,01	0,01	0,05
1.4				0,23	0,03	0,02	0,07	0,05	0,04	0,03	0,01	0,00	0,02	0,01	0,01	0,01	0,00	0,00	0,01	0,01	0,01	0,05
1.5					0,06	0,05	0,05	0,04	0,03	0,05	0,01	0,01	0,01	0,01	0,01	0,02	0,00	0,01	0,01	0,02	0,02	0,01
2.1						0,28	0,04	0,05	0,03	0,04	0,01	0,01	0,02	0,00	0,00	0,01	0,01	0,00	0,01	0,00	0,01	0,01
2.2							0,12	0,07	0,06	0,03	0,02	0,00	0,03	0,00	0,03	0,01	0,01	0,00	0,02	0,01	0,01	0,03
2.3								0,19	0,10	0,06	0,02	0,01	0,02	0,01	0,02	0,01	0,02	0,01	0,04	0,02	0,02	0,07
2.4									0,30	0,06	0,02	0,02	0,03	0,01	0,01	0,01	0,02	0,01	0,02	0,01	0,02	0,05
2.5										0,18	0,02	0,01	0,02	0,01	0,04	0,01	0,02	0,02	0,05	0,01	0,01	0,02
3.1											0,49	0,02	0,02	0,00	0,01	0,01	0,02	0,02	0,03	0,01	0,04	0,01
3.2												0,07	0,03	0,00	0,00	0,00	0,02	0,06	0,02	0,00	0,01	0,01
3.3													0,22	0,03	0,01	0,00	0,02	0,02	0,02	0,00	0,03	0,03
3.4														0,53	0,02	0,01	0,00	0,00	0,01	0,00	0,03	0,03
3.5															0,15	0,01	0,00	0,00	0,01	0,00	0,01	0,01
4.1																0,36	0,01	0,00	0,03	0,01	0,02	0,02
4.2																	0,16	0,02	0,03	0,05	0,02	0,02
4.3																		0,06	0,05	0,01	0,01	0,01
4.4																			0,10	0,03	0,02	0,01
4.5																				0,55	0,01	0,01
Inclassifiable																						0,33

TABLEAU EN ANNEXE 7 : Matrice de confusion de l'accord inter-juge sur les Critères dans le processus de filtrage du corpus

Critères dans les archives et conservés par accord inter-juge unanime	Exemples	Commentaires
1.1 – L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?	<i>Le concept d'alarmes est pas mal, il faut voir comment c'est géré mais c'est pas mal.</i>	Commentaire général
1.2 – Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?	<i>On est obligé de garder ce HMI, avec notre table de variables, et on vient taper dans cette table de variables avec ce smart phone ?</i>	Question sur un point précis d'utilisation
1.3 – L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?	<i>Ça existe dans la gestion des interphonies, les techniciens en agences, ils programment les badges des locataires à distance. Avant ils allaient sur site pour programmer/déprogrammer les badges.</i>	Comparaison explicite avec l'existant
1.4 – L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?	<i>Le bac à glaçons en haut, c'est bien pensé aussi.</i>	Mention explicite d'un aspect du Concept + qualification positive
1.5 – L'usage du concept est-il envisageable comme un système d'usage courant ?	[Aucun verbatim]	
2.1 – Le concept est-il simple à utiliser ?	<i>Moi ça me paraît complexe.</i>	Mention explicite de la complexité
2.2 – Le concept complèterait-il sans les concurrencer les habitudes existantes ?	[Aucun verbatim]	
2.3 – La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?	<i>On a toujours fait comme ça, donc pourquoi... Si ce n'est au niveau hygiène. Je vous le disais tout à l'heure. Déjà par rapport à ça. Et puis bon, là, il y a juste un petit trou qui est ouvert, que quand on ouvre le seau, il peut y avoir des trucs qui tombent dedans, des morceaux d'amandes effilées, ou quoi ou qu'est-ce... J'invente tout et n'importe quoi. Mais au niveau hygiène, je pense que c'est mieux, mais après, à côté de ça... Je vous dis, il a toujours été en seau, donc pourquoi changer ? Je ne suis pas contre le changement, mais je me pose la question de pourquoi ils veulent changer le conditionnement.</i>	Remise en question explicite du Concept/un aspect du Concept par rapport aux pratiques existantes

2.4 – Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?	<i>Du point de vue médical, non, ça ne résoudrait pas de problème parce qu'il faut aller très doucement pour injecter. Et là, ça va quand même vite.</i>	Mention explicite d'un problème et de sa résolution
2.5 – L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ?	[Aucun verbatim]	
3.1 – L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?	<i>On est plus obligé d'être vraiment bricoleur, c'est à la portée de tout le monde.</i>	Mention explicite d'un type de personnes et de la possibilité pour ces personnes d'utiliser le Concept
3.2 – Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?	[Aucun verbatim]	
3.3 – Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?	<i>Le but c'est toujours de gagner du temps. Donc il faut qu'il y ait une personne avec un IPAD accroché au cou et qui regarde en permanence..... c'est anti-social ça</i>	Mention d'une problématique sociale induite par le concept
3.4 – Le concept suscite-t-il un imaginaire d'appropriation ?	<i>Je pense à une oie parce que dès que quelque chose se passe mal dans son environnement, elle se met à gueuler.</i>	Mention d'un nom d'animal
3.5 – Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?	[Aucun verbatim]	
4.1 – L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?	<i>Qu'est ce que je changerais... euh... moi je considère que ça ne répond pas aux tendances globales du marché mais bon... euh... c'est difficile de vous dire ce que je changerais parce que si la cible, on peut parler de la cible clientèle, consommateur final, mais si la cible... si la cible est... tout vient de la cible hein.</i>	Mention explicite du marché, des tendances, de la cible, mis en perspective avec le Concept
4.2 – Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?	[Aucun verbatim]	

4.3 – Le concept est-il adapté à la place qu’occupent les professionnels dans la filière économique / aux relations familiales de l’utilisateur et à leur évolution ?	[Aucun verbatim]	
4.4 – Le concept est-il adapté aux types d’organisation et à leur capacité d’évolution / aux manières de vivre de l’utilisateur et à leur évolution ?	[Aucun verbatim]	
4.5 – Quel prix l’utilisateur consent-il à payer pour ce concept ?	<i>L’investissement déjà a été fait, si tu veux te connecter sur un process qui existe, c’est pour ça que je dis l’application il ne faut pas qu’elle coûte très chère sinon les gens ne le feront pas.</i>	Mention de conditions de coût

TABLEAU EN ANNEXE 8 : Exemples commentés de verbatim qui suscitent un fort accord inter-juge dans le processus de filtrage

Critères dans les archives	Exemples considérés unanimement comme inclassifiables	Commentaires
1.1 – L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?	<i>Seuils de courant, ou...</i>	Pas assez de contexte
1.2 – Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?	<i>Parce que là, ça fait 80 litres, là.</i>	Pas assez de contexte
1.3 – L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?	<i>Comme je disais à Monsieur, je suis en train de faire un plateau de bureau et tous les luminaires font pratiquement tout tout seuls.</i>	Pas assez de contexte : Le fait qu'il s'agisse d'une comparaison avec des techniques existantes n'est pas explicite
1.4 – L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?	<i>Le fait de suivre la maintenance pour débogage d'un programme n'est pas évident en [gamme de produits]. Interface de débogage quoi.</i>	Nécessite une interprétation pragmatique : on ne sait pas si le suivi de la maintenance pour débogage d'un programme est ou non une fonction du concept
1.5 – L'usage du concept est-il envisageable comme un système d'usage courant ?	<i>Oui, sans souci... au contraire !</i>	Pas assez de contexte : manque la question
2.1 – Le concept est-il simple à utiliser ?	<i>Oui, quand je dis « main d'œuvre moins qualifiée »... j'ai pas dit, heu</i>	Pas assez de contexte + nécessite une interprétation pragmatique : on ne connaît pas le rapport du Concept avec la main d'œuvre
2.2 – Le concept compléterait-il sans les concurrencer les habitudes existantes ?	<i>On n'est pas dans le débrouillage.</i>	Pas assez de contexte + nécessite une interprétation pragmatique : on ne sait pas quel est le rapport entre le Concept et le fait d'être débrouillage, et on ne sait pas à quoi cette phrase se réfère (on n'est pas renvoyé à ce que contient le Concept ? Aux pratiques existantes ? Aux pratiques d'autres personnes ?)
2.3 – La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?	<i>quand on est à poil, on n'a pas besoin de mesurer la température de la peau pour savoir si on met un teeshirt ou une doudoune, on prend directement la doudoune</i>	Pas assez de contexte + nécessite une interprétation pragmatique : on ne comprend pas la métaphore vive car pas de comparé

2.4 – Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?	<i>dans les data centers mutualisés, c'est obligatoire, car le client est facturé non seulement au rack utilisé, mais aussi à l'énergie consommée.</i>	Nécessite une interprétation pragmatique : on ne peut pas identifier en quoi consiste le problème et sa résolution sans connaître la nature du Concept
2.5 – L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ?	<i>oui, mais on peut être confronté à des milliers d'infos</i>	Pas assez de contexte : la surcharge d'informations est-elle un problème existant ou qui serait induit par le Concept ?
3.1 – L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?	<i>faut tout refaire quand même</i>	Pas assez de contexte + nécessite une interprétation pragmatique : on ne sait pas à quoi se réfère le <i>tout</i> ni à quel moment (avant ou après le Concept ?) cette phrase se réfère.
3.2 – Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?	<i>Comme c'est un travail qui est moins technique</i>	Pas assez ce contexte
3.3 – Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?	<i>Ça n'en est pas une.</i>	Pas assez de contexte : manque la question
3.4 – Le concept suscite-t-il un imaginaire d'appropriation ?	[Aucun verbatim]	
3.5 – Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?	<i>Je ne vois pas trop... [...] Le cacao en poudre ?</i>	Nécessite une interprétation pragmatique : il faut connaître la nature du Concept pour savoir qu'il s'agit d'une extension d'usage
4.1 – L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?	<i>Ce sont des aspects stupides, peut-être, mais qui ont une... une importance, parce que c'est... c'est la première chose qui se voit, quoi.</i>	Pas assez de contexte
4.2 – Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?	<i>le client lui-même ne sait pas ce qu'on met dans l'armoire, sauf s'il demande du matériel d'une marque spéciale, ça nous est arrivé une fois, 2 fois même. Mais sinon le client lui-même ne sait pas ce qu'on met dans l'armoire.</i>	Nécessite une interprétation pragmatique : il faut connaître la nature du Concept pour savoir que le fait évoqué lui est relatif

4.3 – Le concept est-il adapté à la place qu’occupent les professionnels dans la filière économique / aux relations familiales de l’utilisateur et à leur évolution ?	<i>ce que je vois souvent, entre des systèmes, euh... ben ça intéresse plusieurs services, ces choses là, et il faut que tout le monde y trouve son utilité, quoi, parce que...</i>	Pas assez de contexte : on ne connaît pas l’antécédent de l’anaphore <i>ces choses là</i>
4.4 – Le concept est-il adapté aux types d’organisation et à leur capacité d’évolution / aux manières de vivre de l’utilisateur et à leur évolution ?	<i>L’idéal, voilà. Après, c’est la partie automatisme, quoi, qui gère ça.</i>	Pas assez de contexte + nécessite une interprétation pragmatique : on ne peut pas résoudre l’anaphore <i>ça</i> , et il faut connaître le lien entre le Concept et ce qui est désigné par <i>partie automatisme</i>
4.5 – Quel prix l’utilisateur consent-il à payer pour ce concept ?	<i>2 max 2 max plusieurs Jusqu’à 4 : ?? oui pour moi, c’est mon besoin. Moi aussi.</i>	Manque de contexte : on ne sait pas à quoi la question se réfère

TABLEAU EN ANNEXE 9 : Exemples commentés de verbatim étant unanimement considérés comme inclassifiables pendant le processus de filtrage

Critères dans les archives	Exemples qui ne donnent pas lieu à un accord inter-juge satisfaisant	Classifications données et raisons possibles
1.1 – L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?	<i>Et puis c'est vrai, on voit, ce que je vois là, on voit où ça veut en venir, c'est vraiment petite extension, petite rénovation.</i>	1.1 : Commentaire général sur le Concept 1.5 : Fréquence d'utilisation (petite rénovation -> occasionnel) 2.3 : ? 3.1 : <i>petite rénovation</i> = secteur d'activité = cible
1.2 – Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?	<i>Il n'y a pas le disjoncteur magnétothermique qui serait intégré au... contacteur.</i>	1.1 : Commentaire général sur le Concept 1.3 : Ferait référence à des techniques connues 1.4 : Exprime un choix/désir de fonction (?)
1.3 – L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?	<i>C'est plus attractif que ce que je connais : ce qui n'est pas négligeable quand on fait une maladie comme la miene.</i>	1.1 : Intérêt général pour le Concept 2.4 : Mention d'une aide à un problème (i.e., la <i>maladie</i>) 3.1 : Sous-entend que la personne se sent concernée par le Concept
1.4 – L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?	<i>c'est mieux à distance qu'en local. En local, je ne vois pas l'intérêt, il y a de moins en moins de personne qui se déplacent.</i>	1.4 : Expression d'une préférence sur un aspect du Concept 2.2 : Sous-entend une comparaison avec des pratiques actuelles (?) 2.5 : Projection dans une utilisation courante
1.5 – L'usage du concept est-il envisageable comme un système d'usage courant ?	<i>On pourrait faire une petite variante un peu moins compliquée. Aujourd'hui, on parle de domotique, on pourrait intégrer... C'est une question de coût aussi parce que les gens dans une villa...</i>	1.1 : ? 2.1 : Expression d'une complexité liée au Concept 3.1 : Expression d'un avis sur les catégories de personnes potentiellement visées/exclues de la cible du Concept

2.1 – Le concept est-il simple à utiliser ?	<i>Le problème, c'est que physiquement, on n'a pas essayé. Mais ça a pas l'air compliqué</i>	1.2 : Expression d'une complexité liée au Concept (interprétée ici comme une complexité de compréhension du Concept) 1.3 : Expression de l'impossibilité d'une comparaison avec le connu 1.5 : Projection dans un usage (accepté a priori) (?)
2.2 – Le concept compléterait-il sans les concurrencer les habitudes existantes ?	<i>là on parle du modernisme, tout se fait aussi avec des appareils connectés donc iPhone par exemple. Ah ce n'était pas un iPhone dans la démo.... Il faut quand même avoir un Smartphone</i>	1.2 : Expression d'un manque de compréhension de la démonstration du Concept 2.3 : Comparaison de l'usage actuel des smartphone (?) 4.1 : Expression du modernisme (tendances du marché, air du temps) (?) 4.4 : Expression d'un besoin d'équipement dans l'organisation (?)
2.3 – La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?	<i>On va dire compatible avec les réseaux à 80%. Le reste c'est des postes d'arrivée à 20KV.</i>	1.5 : Expression du fait que le Concept ne couvre pas tous les besoins rencontrés dans l'usage courant 2.2 : Expression d'un conflit avec les pratiques actuelles 2.5 : (Non) projection dans l'usage (?)
2.4 – Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?	<i>Les tests sont plus rapides à faire, parce que... tout est intégré, donc on n'a pas à le re-tester. Alors que nous, le câblage, il faut le re-tester.</i>	2.3 : Comparaison avec les pratiques (alors que nous, le câblage [...]) 2.5 : Projection dans l'usage (?)
2.5 – L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ?	<i>Il faut un enregistrement derrière et alors, les gens seront motivés. Ils seront plus responsables.</i>	2.3 : ? 3.5 : Expression d'une idée nouvelle concernant le Concept 4.4 : Sujet des responsabilités dans une organisation abordé

<p>3.1 – L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?</p>	<p><i>Je pense que c'est mieux d'avoir un retour avec alarmes. Après, pas par un centre de contrôle mais directement chez le client.</i></p>	<p>1.3 : Comparaison implicite avec d'autres systèmes dotés d'alarmes (?) 2.3 : Comparaison implicite avec les habitudes actuelles concernant les alarmes (?) 2.5 : Projection dans une organisation potentielle 4.4 : Avis sur la répartition des responsabilités dans l'organisation</p>
<p>3.2 – Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?</p>	<p><i>moi je disais que ça a un côté pratique, c'est une autre façon de voir l'électricité. Même moi je me vois le faire.</i></p>	<p>1.1 : Commentaire général sur le Concept 1.5 : Projection dans l'usage 3.1 : Identification de soi-même comme potentiel utilisateur du Concept 3.2 : Expression implicite d'une évolution des capacités personnelles d'actions</p>
<p>3.3 – Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?</p>	<p><i>Comme une nouveauté sur le marché, comme quelque chose qui peut être facile d'emploi, facile à mettre en œuvre et beaucoup moins onéreux qu'un PC, on va dire la mobilité, pour faire de petites modifications.</i></p>	<p>1.4 : Identification d'une caractéristique intéressante du concept (la mobilité) ? 1.5 : Projection dans l'usage (?) 2.3 : Comparaison avec pratiques actuelles (utilisation d'un PC) 2.4 : Expression du Concept comme alternative à un problème (le prix élevé du PC)</p>

3.4 – Le concept suscite-t-il un imaginaire d'appropriation ?	<i>Parce que ça me permet de faire une solution personnalisable, avec toutes les contraintes que j'ai émises pendant l'échange, hein.</i>	1.4 : Expression de l'aspect positif du fait de pouvoir choisir les fonctions (solution personnalisable) (?) 1.5 : Projection dans l'usage (?) 2.4 : Interprétation de <i>contraintes</i> comme celles rencontrées par l'utilisateur dans sa vie quotidienne (? – ambiguïté : nous l'interprétons personnellement comme des contraintes provoquées par le Concept, qui auraient été évoquées plus tôt pendant l'entretien)
3.5 – Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?	<i>Si c'est pas réutilisable, ça fait un peu gadget quand même. Ça se justifie pour une piqûre une fois mais peut-être pas pour une piqûre hebdomadaire.</i>	1.5 : Expression d'une restriction liée à la fréquence d'utilisation 2.2 : Expression d'un manque d'intérêt du Concept par rapport à des pratiques 2.3 : idem (?) 2.4 : Expression de l'inutilité potentielle du Concept
4.1 – L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?	<i>Tout le monde parle de transition énergétique, d'économie d'énergie mais il va falloir qu'on rentre comme dans l'industrie dans des processus d'évaluation de la qualité de ce qu'on fait et des sous qu'on met. Donc ça en fait partie.</i>	1.2 : (?) 2.2 : Comparaison avec les pratiques relatives à un secteur 4.1 : Mention de tendances
4.2 – Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?	<i>si les capteurs doivent être disposés en surface d'une paroi, ne doivent pas être revêtu par un revêtement ... c'est-à-dire qu'il faut que l'on interdise au locataire ce type de revêtement...</i>	2.3 : (?) 2.4 : Identification d'un problème provoqué par le Concept 2.5 : Projection dans l'usage (?)

4.3 – Le concept est-il adapté à la place qu’occupent les professionnels dans la filière économique / aux relations familiales de l’utilisateur et à leur évolution ?	<i>Ça va nous enlever du boulot.</i>	1.1 : Commentaire général sur le Concept 3.2 : Expression d’un impact sur le rôle (i.e., le rôle de l’utilisateur impliquerait moins de tâches) 4.3 : Mention implicite de changements dans la filière économique
4.4 – Le concept est-il adapté aux types d’organisation et à leur capacité d’évolution / aux manières de vivre de l’utilisateur et à leur évolution ?	<i>Aussi un aspect de valorisation d’un bien, ces choses peuvent ajouter une valeur pour une vente ou une location.</i>	2.4 : (?) 3.2 : Mention implicite de l’impact sur le rôle de la valorisation du bien (?)
4.5 – Quel prix l’utilisateur consent-il à payer pour ce concept ?	<i>Pour l’instant, tel que c’est là, c’est bien pour le moteur, mais ce n’est pas utile dans le suivi de pannes telles qu’on les a pour l’instant. On va voir comment le produit va arriver au niveau financier.</i>	1.4 : Identification de fonctions intéressantes ou non 2.3 : Comparaison avec les besoins liés aux problèmes 4.5 : Mention du coût potentiel du Concept

TABLEAU EN ANNEXE 10 : Exemples commentés de verbatim qui provoquent un désaccord important pendant le processus de filtrage

ANNEXE 3 : ANALYSE LEXICOMÉTRIQUE

Légende : **Gras** : champ lexical attendu par le Critère – **Bleu** : courant dans la langue – **Violet** : expression d’une opinion/d’un jugement général, non spécifique au champ lexical du Critère – **Orange** : jargon métier/technique/technologique – **Vert** : jargon commercial

6. Exemple de données extraites

2.4 – Le concept <i>résout-il des problèmes gênants rencontrés par l'utilisateur ?</i> Champs lexicaux attendus : problèmes, résolution						
Plus fréquents (disp. Min = 3 – freq. Min = 5)		Plus sur-représentés pour le Critère (freq min = 2 – disp. Min = 2 – rép min = 0,5)			Plus spécifiques	
Freq.	Ngram	% répartition	Freq.	Ngram	LogLike	Ngram
41	faire	1,00	5	être DET gain	116,99	problème
41	pouvoir	1,00	4	cela ne résoudre	97,94	niveau
29	problème	1,00	4	cela ne résoudre pas	86,39	résoudre
25	aller	1,00	4	être DET gain de	69,48	câblage
24	temps	1,00	4	faire gagner	68,44	maintenance
21	tout	1,00	4	ne résoudre	64,18	gain
18	avec	1,00	4	ne résoudre pas	62,28	vis
18	peu	1,00	4	résoudre pas	58,94	praliner
15	bien	1,00	4	risque de	43,54	problématique
15	niveau	1,00	3	ce être DET gain	38,94	lésion
14	comme	1,00	3	de risque de	36,54	contacteur
14	on pouvoir	1,00	3	nous faire gagner	34,82	temps
13	au niveau	1,00	3	supporter	34,72	produit
13	falloir	1,00	2	cela apporter	34,48	place
13	même	1,00	2	cela faciliter	31,26	risque
13	résoudre	1,00	2	cela résoudre	30,71	superviseur

13	voir	1,00	2	cela se faire	30,68	redémarrage
12	dire	1,00	2	cela sembler	30,16	concept
12	mettre	1,00	2	de changer	30,00	gagner
12	penser	1,00	2	de risque de se	29,03	source
12	très	1,00	2	DET multitude		
11	bon	1,00	2	DET multitude de		
11	gagner	1,00	2	DET problème du		
11	je penser	1,00	2	du cas		
11	moins	1,00	2	être plus précis		
10	DET problème	1,00	2	faire gagner du		
10	il falloir	1,00	2	faire gagner du temps		
10	mal	1,00	2	faire mal		
9	arriver	1,00	2	gagner vachement		
9	cela pouvoir	1,00	2	gagner vachement de		
9	chose	1,00	2	là je pouvoir		
9	comme cela	1,00	2	là on pouvoir		
9	devoir	1,00	2	lampe		
9	vrai	1,00	2	moins de perte		
8	après	1,00	2	multitude		
8	aussi	1,00	2	multitude de		
8	changer	1,00	2	ne résoudre pas de		
8	de temps	1,00	2	nous faire gagner du		
8	on faire	1,00	2	on faire du		
8	passer	1,00	2	on perdre		
8	permettre	1,00	2	on revenir		
8	prendre	1,00	2	parce que actuellement		
8	quand	1,00	2	permettre de faire		
7	au niveau du	1,00	2	problème du		
7	ce être vrai	1,00	2	que actuellement		
7	de DET place	1,00	2	que au niveau		
7	de problème	1,00	2	que au niveau de		

7	DET place	1,00	2	régulièrement
7	du temps	1,00	2	résoudre pas de
7	être vrai	1,00	2	résoudre pas de problème
7	moins de	1,00	2	risque de se
7	niveau du	1,00	2	standard
7	petit	1,00	2	stock
7	pouvoir être	1,00	2	vachement de
7	problème de	1,00	2	visser
7	risque	1,00	2	voir du
7	toujours	0,93	13	résoudre
7	voilà	0,78	7	DET gain
7	vouloir	0,78	7	risque
6	beaucoup	0,75	3	être vrai que cela
6	besoin	0,75	3	fil
6	ce être vrai que	0,75	3	le praliner
6	être vrai que	0,75	3	on pouvoir être
6	je penser que	0,75	3	problématique de
6	penser que	0,75	3	régler
6	permettre de	0,75	3	vrai que cela
6	poser	0,73	8	de temps
6	produit	0,71	5	DET gain de
6	savoir	0,70	7	de DET place
6	vrai que	0,70	7	vis
5	actuellement	0,67	4	aider
5	apporter	0,67	4	de risque
5	cela aller	0,67	4	pas de problème
5	cela permettre	0,67	4	référence
5	côté	0,67	2	anticiper
5	du problème	0,67	2	avoir moins de risque
5	ne pouvoir	0,67	2	cela être bien
		0,67	2	cela prendre moins

0,67	2	cela prendre moins de
0,67	2	changer de
0,67	2	chose sur
0,67	2	de problème mais
0,67	2	DET problématique
0,67	2	difficulté
0,67	2	médical
0,67	2	mobile
0,67	2	moins de risque
0,67	2	ne arriver
0,67	2	on avoir pouvoir
0,67	2	penser que cela aller
0,67	2	perdre du
0,67	2	plus précis
0,67	2	pour aller
0,67	2	prendre moins
0,67	2	prendre moins de
0,67	2	problème mais
0,67	2	quelque chose sur
0,67	2	tromper
0,64	7	de problème
0,63	5	DET vis
0,60	3	être DET problème
0,60	3	faciliter
0,57	4	gagner du
0,57	4	gagner du temps
0,56	5	DET peu plus
0,56	5	du problème
0,56	5	gain de
0,53	29	problème
0,50	7	gain

0,50	4	avoir moins
0,50	4	avoir moins de
0,50	4	éviter
0,50	4	problématique
0,50	3	cela prendre
0,50	3	de perte
0,50	3	fermer
0,50	3	prévoir
0,50	3	se faire
0,50	2	avoir pas de problème
0,50	2	avoir peur
0,50	2	ce être DET problème
0,50	2	cela éviter
0,50	2	de petit
0,50	2	dire si
0,50	2	je avoir peur
0,50	2	situation
0,50	2	supérieur
0,50	2	vachement

TABLEAU EN ANNEXE 11 : Exemple de n-grams extraits par analyse textométrique

7. Analyse par catégories morphosyntaxiques

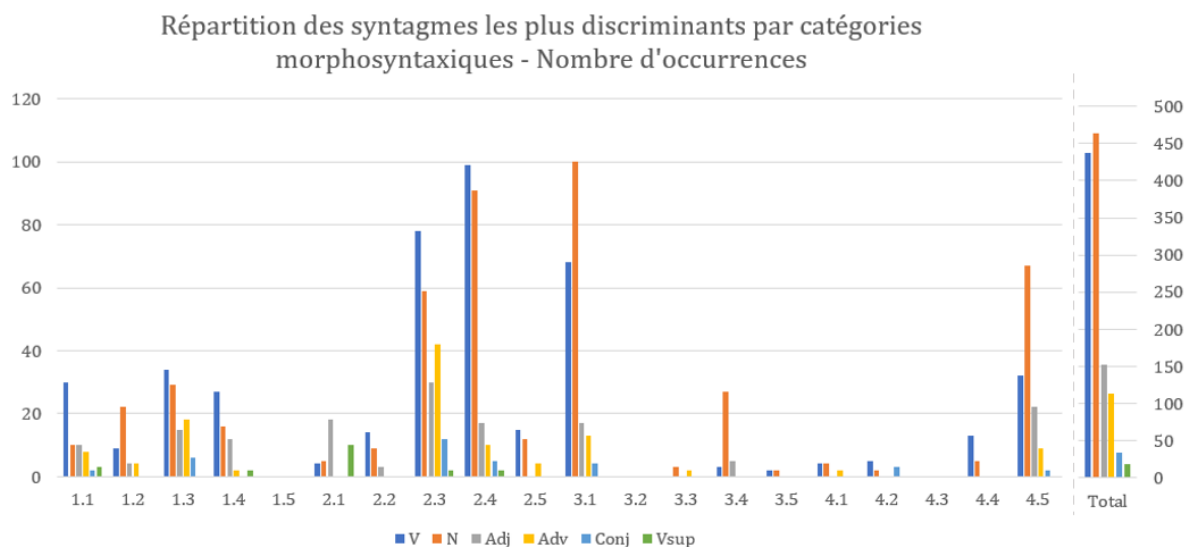


FIGURE EN ANNEXE 8 : Répartition des syntagmes les plus discriminants par catégories morphosyntaxiques – Nombre d'occurrences

↑p110;p119

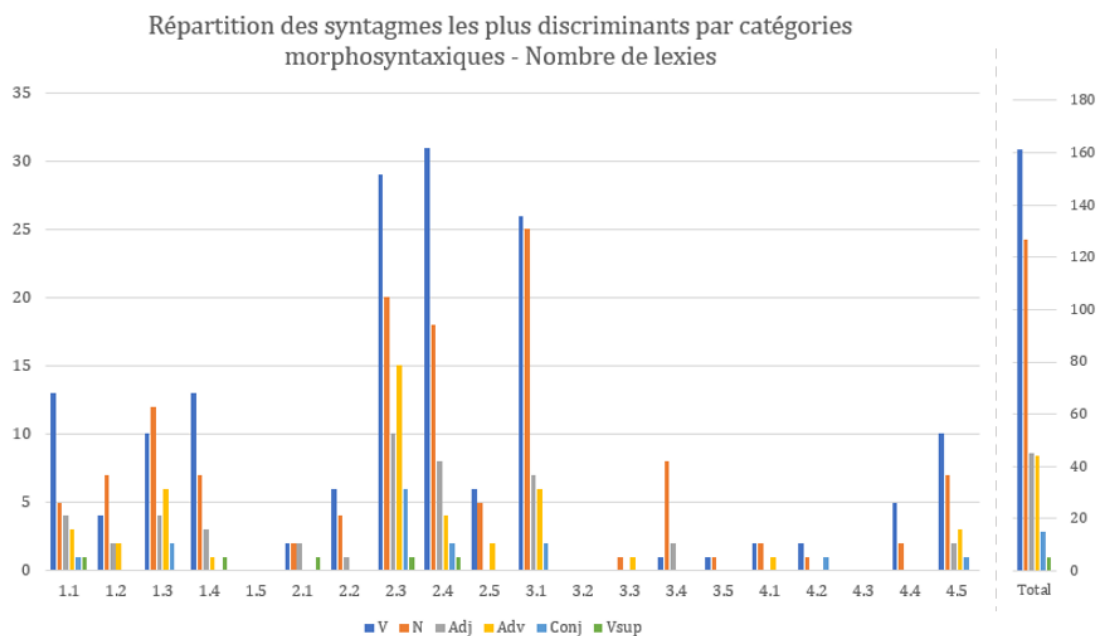


FIGURE EN ANNEXE 9 : Répartition des syntagmes les plus discriminants par catégories morphosyntaxiques – Nombre de lexies

↑p110;p119

ANNEXE 4 : TESTS DE CLASSIFIEURS

8. Résultats pour les approches par sacs de mots

Légende :

- Col. *Cl.*: *C* = Critères seuls, *C+I* = Critères + Inclassifiables
- *Lemm.* = Lemmatisation
- Col. *Mots outils* : *Remp. Cat* = mots outils remplacés par l'étiquette de leur catégorie morphosyntaxique
- Col. *N-grams* : tailles des *n-grams*, 1 = unigrams, 2 = bigrams, 1-2 = unigrams + bigrams
- Classifieurs : *NB* = Naïve Bayes, *SVM* = Support Machine Vector, *KNN*= k-Nearest Neighbors, *LR* = Logistic Regression, *RF* = Random Forest, *MLP* = Multilayer Perceptron

Classification	Classes	Mots Lemm.	Mots outils	n- grams	NB			SVM			KNN			LR			RF			MLP		
					p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1
Monolabel	C	non	Conserver	1	0,26	0,23	0,18	0,24	0,22	0,22	0,26	0,14	0,15	0,28	0,28	0,26	0,35	0,36	0,32	0,22	0,21	0,19
Monolabel	C	non	Conserver	1-2	0,36	0,24	0,18	0,28	0,26	0,25	0,18	0,09	0,08	0,26	0,28	0,26	0,33	0,32	0,28	0,21	0,16	0,17
Monolabel	C	non	Conserver	2	0,26	0,26	0,22	0,19	0,2	0,17	0,05	0,09	0,03	0,19	0,23	0,2	0,17	0,16	0,12	0,23	0,12	0,14
Monolabel	C	non	Remp. Cat	1	0,27	0,23	0,18	0,32	0,29	0,29	0,21	0,14	0,15	0,33	0,3	0,29	0,26	0,26	0,25	0,3	0,24	0,25
Monolabel	C	non	Remp. Cat	1-2	0,32	0,22	0,19	0,28	0,24	0,25	0,25	0,15	0,11	0,3	0,33	0,3	0,46	0,36	0,32	0,27	0,12	0,16
Monolabel	C	non	Remp. Cat	2	0,21	0,17	0,15	0,16	0,19	0,16	0,04	0,1	0,05	0,22	0,23	0,2	0,22	0,2	0,17	0,37	0,19	0,2
Monolabel	C	non	Supprimer	1	0,31	0,22	0,18	0,27	0,26	0,25	0,24	0,19	0,19	0,31	0,3	0,29	0,22	0,22	0,19	0,22	0,22	0,22
Monolabel	C	non	Supprimer	1-2	0,32	0,22	0,18	0,26	0,24	0,23	0,27	0,14	0,11	0,31	0,28	0,27	0,25	0,23	0,19	0,3	0,18	0,2
Monolabel	C	non	Supprimer	2	0,27	0,22	0,19	0,23	0,22	0,19	0,07	0,07	0,03	0,22	0,23	0,21	0,29	0,22	0,2	0,2	0,11	0,12
Monolabel	C	oui	Conserver	1	0,35	0,29	0,25	0,22	0,22	0,22	0,2	0,13	0,12	0,3	0,32	0,31	0,32	0,34	0,3	0,28	0,21	0,23
Monolabel	C	oui	Conserver	1-2	0,27	0,2	0,17	0,24	0,23	0,22	0,24	0,17	0,15	0,41	0,36	0,36	0,31	0,3	0,27	0,33	0,26	0,27
Monolabel	C	oui	Conserver	2	0,35	0,23	0,17	0,28	0,25	0,22	0,06	0,09	0,04	0,33	0,3	0,3	0,28	0,26	0,24	0,25	0,21	0,21
Monolabel	C	oui	Remp. Cat	1	0,33	0,18	0,14	0,29	0,27	0,27	0,24	0,2	0,19	0,29	0,27	0,27	0,29	0,34	0,3	0,26	0,23	0,24
Monolabel	C	oui	Remp. Cat	1-2	0,27	0,3	0,23	0,22	0,2	0,19	0,04	0,09	0,03	0,2	0,22	0,19	0,25	0,25	0,22	0,21	0,1	0,11

Monolabel	C	oui	Remp. Cat	1-2	0,22	0,2	0,14	0,29	0,28	0,28	0,24	0,15	0,14	0,31	0,35	0,33	0,28	0,3	0,26	0,23	0,19	0,2
Monolabel	C	oui	Supprimer	1	0,31	0,22	0,17	0,26	0,24	0,24	0,17	0,15	0,14	0,36	0,35	0,35	0,23	0,3	0,24	0,24	0,2	0,21
Monolabel	C	oui	Supprimer	1-2	0,4	0,26	0,2	0,3	0,27	0,27	0,19	0,11	0,1	0,3	0,34	0,31	0,3	0,35	0,3	0,32	0,26	0,27
Monolabel	C	oui	Supprimer	2	0,36	0,22	0,21	0,28	0,24	0,24	0,03	0,11	0,04	0,29	0,28	0,28	0,32	0,26	0,25	0,29	0,2	0,21
Monolabel	C+I	non	Conserver	1	0,17	0,16	0,12	0,24	0,22	0,21	0,12	0,08	0,08	0,24	0,23	0,23	0,29	0,26	0,22	0,23	0,21	0,21
Monolabel	C+I	non	Conserver	1-2	0,31	0,25	0,23	0,21	0,24	0,2	0,13	0,11	0,05	0,21	0,2	0,18	0,22	0,2	0,16	0,31	0,18	0,21
Monolabel	C+I	non	Conserver	2	0,19	0,14	0,13	0,26	0,27	0,25	0,13	0,15	0,1	0,26	0,26	0,23	0,31	0,27	0,23	0,28	0,19	0,21
Monolabel	C+I	non	Remp. Cat	1	0,28	0,27	0,2	0,27	0,24	0,24	0,12	0,1	0,09	0,34	0,34	0,33	0,38	0,33	0,31	0,19	0,13	0,14
Monolabel	C+I	non	Remp. Cat	1-2	0,18	0,17	0,15	0,31	0,24	0,21	0,14	0,1	0,05	0,24	0,24	0,22	0,27	0,24	0,19	0,22	0,18	0,19
Monolabel	C+I	non	Remp. Cat	2	0,4	0,19	0,14	0,27	0,25	0,24	0,17	0,15	0,12	0,22	0,26	0,23	0,37	0,35	0,29	0,17	0,16	0,16
Monolabel	C+I	non	Supprimer	1	0,3	0,21	0,16	0,24	0,23	0,21	0,19	0,17	0,17	0,25	0,27	0,24	0,3	0,27	0,22	0,15	0,14	0,14
Monolabel	C+I	non	Supprimer	1-2	0,19	0,2	0,15	0,17	0,23	0,17	0,1	0,06	0,02	0,25	0,24	0,22	0,13	0,16	0,12	0,42	0,18	0,21
Monolabel	C+I	non	Supprimer	2	0,42	0,21	0,19	0,26	0,27	0,25	0,2	0,12	0,11	0,27	0,29	0,27	0,42	0,3	0,28	0,29	0,17	0,17
Monolabel	C+I	oui	Conserver	1	0,34	0,21	0,17	0,3	0,25	0,25	0,25	0,17	0,17	0,28	0,29	0,27	0,37	0,35	0,32	0,24	0,19	0,2
Monolabel	C+I	oui	Conserver	1-2	0,23	0,18	0,15	0,29	0,29	0,26	0,04	0,07	0,03	0,26	0,24	0,23	0,21	0,24	0,2	0,3	0,2	0,22
Monolabel	C+I	oui	Conserver	2	0,22	0,24	0,18	0,26	0,28	0,26	0,31	0,18	0,19	0,3	0,3	0,29	0,3	0,29	0,26	0,27	0,18	0,19
Monolabel	C+I	oui	Remp. Cat	1	0,3	0,21	0,18	0,29	0,29	0,29	0,22	0,14	0,14	0,23	0,22	0,21	0,27	0,29	0,25	0,2	0,18	0,17
Monolabel	C+I	oui	Remp. Cat	1-2	0,28	0,21	0,18	0,28	0,25	0,23	0,13	0,07	0,05	0,18	0,2	0,18	0,29	0,23	0,19	0,29	0,17	0,2
Monolabel	C+I	oui	Remp. Cat	2	0,32	0,19	0,14	0,27	0,29	0,26	0,16	0,14	0,13	0,25	0,27	0,26	0,31	0,3	0,27	0,27	0,18	0,2
Monolabel	C+I	oui	Supprimer	1	0,37	0,26	0,25	0,36	0,27	0,28	0,14	0,12	0,11	0,28	0,3	0,28	0,27	0,29	0,26	0,18	0,14	0,15
Monolabel	C+I	oui	Supprimer	2	0,3	0,16	0,12	0,32	0,32	0,3	0,14	0,12	0,11	0,28	0,28	0,26	0,28	0,26	0,24	0,21	0,2	0,19
Monolabel	C+I	oui	Supprimer	2	0,22	0,22	0,19	0,27	0,29	0,25	0,21	0,11	0,09	0,26	0,22	0,2	0,33	0,24	0,21	0,36	0,18	0,19
Multilabel	C	non	Conserver	1	0,19	0,17	0,17	0,35	0,30	0,31	0,45	0,02	0,04	0,33	0,26	0,27	0,60	0,04	0,07	0,34	0,24	0,28
Multilabel	C	non	Conserver	1-2	0,24	0,08	0,12	0,36	0,18	0,23	0,51	0,05	0,08	0,42	0,18	0,24	0,71	0,02	0,04	0,37	0,21	0,26
Multilabel	C	non	Conserver	1-2	0,24	0,13	0,17	0,34	0,12	0,17	0,27	0,07	0,10	0,45	0,13	0,17	0,37	0,03	0,05	0,39	0,16	0,21
Multilabel	C	non	Remp. Cat	1	0,19	0,17	0,18	0,33	0,27	0,29	0,42	0,03	0,05	0,34	0,24	0,27	0,56	0,03	0,05	0,32	0,24	0,27
Multilabel	C	non	Remp. Cat	1-2	0,21	0,11	0,14	0,35	0,12	0,18	0,38	0,07	0,09	0,39	0,14	0,19	0,34	0,02	0,04	0,36	0,15	0,19
Multilabel	C	non	Remp. Cat	2	0,21	0,10	0,13	0,33	0,16	0,20	0,40	0,06	0,08	0,38	0,16	0,21	0,51	0,04	0,07	0,35	0,19	0,23
Multilabel	C	non	Supprimer	1	0,20	0,20	0,20	0,34	0,24	0,28	0,45	0,03	0,05	0,34	0,27	0,29	0,69	0,04	0,07	0,34	0,24	0,27
Multilabel	C	non	Supprimer	1-2	0,25	0,08	0,12	0,39	0,18	0,24	0,46	0,04	0,06	0,36	0,15	0,21	0,50	0,03	0,06	0,38	0,18	0,24
Multilabel	C	non	Supprimer	1-2	0,20	0,12	0,14	0,38	0,13	0,19	0,35	0,08	0,11	0,37	0,15	0,19	0,43	0,04	0,06	0,34	0,14	0,19

Multilabel	C	oui	Conserver	1	0,18	0,22	0,20	0,31	0,27	0,28	0,33	0,04	0,06	0,38	0,33	0,34	0,68	0,03	0,05	0,34	0,26	0,3
Multilabel	C	oui	Conserver	2	0,25	0,09	0,12	0,40	0,23	0,28	0,49	0,06	0,09	0,38	0,18	0,24	0,51	0,03	0,05	0,42	0,21	0,27
Multilabel	C	oui	Conserver	2	0,24	0,11	0,15	0,37	0,17	0,22	0,44	0,07	0,09	0,42	0,16	0,21	0,51	0,03	0,05	0,38	0,17	0,23
Multilabel	C	oui	Remp. Cat	1	0,20	0,26	0,22	0,32	0,25	0,28	0,53	0,03	0,05	0,32	0,25	0,28	0,47	0,05	0,09	0,36	0,28	0,31
Multilabel	C	oui	Remp. Cat	1-2	0,22	0,09	0,12	0,36	0,20	0,25	0,39	0,05	0,08	0,39	0,20	0,25	0,64	0,02	0,04	0,39	0,19	0,25
Multilabel	C	oui	Remp. Cat	2	0,23	0,11	0,15	0,34	0,12	0,17	0,31	0,06	0,08	0,34	0,13	0,17	0,67	0,03	0,06	0,42	0,16	0,22
Multilabel	C	oui	Supprimer	1	0,19	0,21	0,20	0,33	0,28	0,30	0,54	0,05	0,07	0,29	0,25	0,26	0,47	0,02	0,04	0,32	0,26	0,29
Multilabel	C	oui	Supprimer	1-2	0,29	0,12	0,16	0,35	0,18	0,23	0,50	0,05	0,09	0,53	0,24	0,31	0,51	0,03	0,05	0,37	0,2	0,25
Multilabel	C	oui	Supprimer	2	0,25	0,13	0,16	0,38	0,18	0,23	0,25	0,08	0,10	0,34	0,14	0,18	0,49	0,04	0,07	0,4	0,17	0,22
Multilabel	C+I	non	Conserver	1	0,26	0,19	0,21	0,36	0,30	0,32	0,43	0,12	0,15	0,38	0,34	0,36	0,54	0,12	0,15	0,42	0,35	0,37
Multilabel	C+I	non	Conserver	1-2	0,25	0,18	0,21	0,42	0,25	0,28	0,36	0,14	0,17	0,43	0,22	0,26	0,55	0,18	0,18	0,4	0,24	0,26
Multilabel	C+I	non	Conserver	1-2	0,31	0,18	0,21	0,40	0,27	0,31	0,44	0,18	0,20	0,49	0,29	0,33	0,70	0,14	0,16	0,41	0,29	0,32
Multilabel	C+I	non	Remp. Cat	1	0,24	0,20	0,21	0,38	0,30	0,33	0,50	0,11	0,14	0,43	0,37	0,39	0,55	0,13	0,16	0,38	0,3	0,33
Multilabel	C+I	non	Remp. Cat	1-2	0,25	0,19	0,22	0,36	0,23	0,26	0,31	0,17	0,17	0,39	0,24	0,26	0,41	0,18	0,18	0,42	0,24	0,27
Multilabel	C+I	non	Remp. Cat	2	0,28	0,15	0,18	0,41	0,28	0,31	0,44	0,16	0,17	0,43	0,25	0,29	0,40	0,14	0,16	0,41	0,25	0,28
Multilabel	C+I	non	Supprimer	1	0,25	0,22	0,22	0,39	0,33	0,35	0,48	0,14	0,16	0,41	0,30	0,34	0,46	0,13	0,14	0,38	0,29	0,32
Multilabel	C+I	non	Supprimer	1-2	0,30	0,17	0,20	0,43	0,28	0,31	0,41	0,20	0,20	0,41	0,25	0,29	0,48	0,14	0,16	0,43	0,29	0,33
Multilabel	C+I	non	Supprimer	2	0,26	0,21	0,23	0,41	0,22	0,25	0,43	0,16	0,18	0,40	0,23	0,25	0,46	0,17	0,17	0,39	0,23	0,26
Multilabel	C+I	oui	Conserver	1	0,24	0,23	0,22	0,34	0,32	0,32	0,56	0,14	0,16	0,37	0,34	0,35	0,52	0,13	0,16	0,38	0,31	0,33
Multilabel	C+I	oui	Conserver	2	0,26	0,21	0,23	0,41	0,27	0,30	0,37	0,15	0,16	0,41	0,25	0,29	0,49	0,16	0,17	0,4	0,27	0,31
Multilabel	C+I	oui	Conserver	2	0,26	0,13	0,17	0,42	0,30	0,34	0,35	0,10	0,12	0,45	0,25	0,31	0,68	0,15	0,18	0,45	0,27	0,33
Multilabel	C+I	oui	Remp. Cat	1	0,26	0,28	0,26	0,37	0,34	0,35	0,55	0,14	0,16	0,37	0,35	0,35	0,59	0,13	0,15	0,38	0,3	0,32
Multilabel	C+I	oui	Remp. Cat	1-2	0,27	0,19	0,21	0,40	0,30	0,33	0,40	0,13	0,16	0,43	0,28	0,33	0,55	0,13	0,15	0,44	0,3	0,34
Multilabel	C+I	oui	Remp. Cat	1-2	0,26	0,19	0,22	0,42	0,26	0,30	0,35	0,13	0,15	0,44	0,26	0,29	0,50	0,16	0,18	0,38	0,23	0,26
Multilabel	C+I	oui	Supprimer	1	0,23	0,22	0,22	0,38	0,36	0,37	0,44	0,12	0,13	0,36	0,35	0,35	0,58	0,14	0,17	0,38	0,32	0,34
Multilabel	C+I	oui	Supprimer	1-2	0,25	0,20	0,22	0,39	0,25	0,28	0,44	0,19	0,19	0,39	0,23	0,26	0,53	0,16	0,17	0,37	0,22	0,26
Multilabel	C+I	oui	Supprimer	1-2	0,25	0,17	0,19	0,43	0,31	0,35	0,61	0,10	0,13	0,43	0,27	0,31	0,60	0,15	0,19	0,42	0,28	0,32

TABLEAU EN ANNEXE 12 : Résultats pour l'approche par sacs de mots

9. Résultats pour les approches par TF-IDF

Légende :

- Col. *Classes* : *C* = Critères seuls, *C+I* = Critères + Inclassifiables
- *Lemm.* = *Lemmatisation*
- Col. *Mots outils* : *Remp. Cat* = mots outils remplacés par l'étiquette de leur catégorie morphosyntaxique
- Col. *N-grams* : tailles des *n-grams*, 1 = unigrams, 2 = bigrams, 1-2 = unigrams + bigrams
- Classifieurs : *NB* = Naïve Bayes, *SVM* = Support Machine Vector, *KNN* = k-Nearest Neighbors, *LR* = Logistic Regression, *RF* = Random Forest, *MLP* = Multilayer Perceptron

Classification	Classes	Mots			n-grams	NB			SVM			KNN			LR			RF			MLP		
		Lemm.	outils			p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1
Monolabel	C	non	Conserver	1	0,14	0,17	0,11	0,33	0,34	0,29	0,37	0,26	0,27	0,37	0,34	0,33	0,29	0,34	0,29	0,41	0,3	0,31	
Monolabel	C	non	Conserver	1-2	0,11	0,19	0,12	0,18	0,24	0,17	0,19	0,17	0,17	0,30	0,30	0,27	0,24	0,24	0,20	0,3	0,15	0,18	
Monolabel	C	non	Conserver	2	0,09	0,20	0,09	0,32	0,28	0,24	0,25	0,21	0,20	0,35	0,38	0,34	0,33	0,30	0,27	0,37	0,22	0,23	
Monolabel	C	non	Remp. Cat	1	0,22	0,21	0,13	0,32	0,33	0,29	0,37	0,26	0,27	0,38	0,36	0,35	0,34	0,30	0,29	0,36	0,23	0,26	
Monolabel	C	non	Remp. Cat	1-2	0,24	0,22	0,12	0,25	0,23	0,19	0,27	0,20	0,21	0,36	0,35	0,32	0,25	0,22	0,20	0,34	0,28	0,28	
Monolabel	C	non	Remp. Cat	1-2	0,11	0,17	0,10	0,29	0,20	0,14	0,30	0,21	0,21	0,31	0,28	0,25	0,21	0,22	0,19	0,3	0,09	0,12	
Monolabel	C	non	Supprimer	1	0,20	0,24	0,15	0,34	0,34	0,29	0,38	0,23	0,25	0,37	0,36	0,36	0,27	0,28	0,24	0,25	0,22	0,23	
Monolabel	C	non	Supprimer	1-2	0,18	0,21	0,14	0,44	0,28	0,25	0,29	0,25	0,24	0,31	0,32	0,29	0,21	0,28	0,23	0,4	0,2	0,23	
Monolabel	C	non	Supprimer	2	0,16	0,18	0,11	0,32	0,30	0,25	0,24	0,18	0,17	0,29	0,29	0,26	0,27	0,21	0,19	0,39	0,15	0,15	
Monolabel	C	oui	Conserver	1	0,10	0,15	0,09	0,28	0,30	0,26	0,30	0,24	0,23	0,40	0,38	0,36	0,37	0,32	0,30	0,29	0,25	0,26	
Monolabel	C	oui	Conserver	1-2	0,20	0,26	0,18	0,43	0,34	0,33	0,31	0,26	0,24	0,43	0,40	0,38	0,29	0,31	0,28	0,23	0,24	0,2	
Monolabel	C	oui	Conserver	2	0,12	0,15	0,07	0,39	0,27	0,22	0,20	0,18	0,17	0,26	0,29	0,26	0,42	0,22	0,20	0,42	0,19	0,18	
Monolabel	C	oui	Remp. Cat	1	0,29	0,20	0,13	0,38	0,32	0,29	0,32	0,25	0,25	0,32	0,35	0,33	0,28	0,31	0,26	0,35	0,3	0,3	
Monolabel	C	oui	Remp. Cat	1-2	0,19	0,21	0,13	0,31	0,24	0,19	0,17	0,17	0,16	0,30	0,32	0,29	0,29	0,26	0,25	0,18	0,17	0,15	
Monolabel	C	oui	Remp. Cat	2	0,24	0,25	0,19	0,43	0,33	0,31	0,29	0,27	0,25	0,34	0,34	0,31	0,26	0,26	0,23	0,34	0,24	0,26	
Monolabel	C	oui	Supprimer	1	0,10	0,14	0,06	0,34	0,31	0,29	0,43	0,30	0,30	0,38	0,38	0,36	0,26	0,29	0,26	0,36	0,2	0,24	
Monolabel	C	oui	Supprimer	1-2	0,43	0,19	0,15	0,37	0,32	0,29	0,31	0,22	0,22	0,37	0,37	0,35	0,35	0,34	0,32	0,3	0,25	0,23	
Monolabel	C	oui	Supprimer	1-2	0,12	0,20	0,11	0,31	0,24	0,17	0,19	0,16	0,15	0,36	0,32	0,30	0,31	0,26	0,23	0,26	0,14	0,14	

Monolabel	C+I	non	Conserver	1	0,17	0,26	0,17	0,26	0,27	0,24	0,33	0,24	0,26	0,34	0,33	0,32	0,30	0,27	0,25	0,26	0,17	0,18
Monolabel	C+I	non	Conserver	1-2	0,18	0,16	0,11	0,15	0,24	0,17	0,24	0,17	0,18	0,30	0,30	0,28	0,34	0,23	0,21	0,5	0,09	0,12
Monolabel	C+I	non	Conserver	2	0,20	0,20	0,12	0,32	0,28	0,25	0,47	0,23	0,24	0,42	0,40	0,39	0,26	0,33	0,27	0,36	0,2	0,22
Monolabel	C+I	non	Remp. Cat	1	0,15	0,20	0,12	0,37	0,34	0,30	0,29	0,20	0,19	0,33	0,32	0,28	0,36	0,34	0,31	0,4	0,26	0,28
Monolabel	C+I	non	Remp. Cat	1-2	0,18	0,18	0,13	0,28	0,26	0,21	0,31	0,17	0,19	0,28	0,25	0,23	0,21	0,18	0,14	0,34	0,12	0,14
Monolabel	C+I	non	Remp. Cat	2	0,20	0,18	0,12	0,34	0,29	0,26	0,29	0,18	0,20	0,39	0,38	0,36	0,29	0,32	0,29	0,3	0,17	0,18
Monolabel	C+I	non	Supprimer	1	0,13	0,18	0,10	0,20	0,26	0,21	0,29	0,20	0,20	0,32	0,33	0,32	0,31	0,27	0,23	0,32	0,21	0,23
Monolabel	C+I	non	Supprimer	1-2	0,18	0,21	0,15	0,28	0,26	0,19	0,23	0,18	0,18	0,23	0,24	0,23	0,16	0,23	0,17	0,34	0,12	0,14
Monolabel	C+I	non	Supprimer	2	0,25	0,12	0,08	0,36	0,26	0,20	0,24	0,20	0,19	0,42	0,38	0,37	0,35	0,30	0,27	0,44	0,18	0,17
Monolabel	C+I	oui	Conserver	1	0,28	0,19	0,13	0,33	0,33	0,30	0,39	0,26	0,25	0,38	0,35	0,33	0,23	0,25	0,21	0,28	0,26	0,26
Monolabel	C+I	oui	Conserver	1-2	0,14	0,23	0,14	0,40	0,29	0,24	0,21	0,15	0,15	0,32	0,35	0,32	0,20	0,20	0,17	0,37	0,15	0,16
Monolabel	C+I	oui	Conserver	2	0,27	0,20	0,12	0,33	0,33	0,28	0,37	0,23	0,25	0,25	0,29	0,26	0,22	0,24	0,21	0,35	0,16	0,19
Monolabel	C+I	oui	Remp. Cat	1	0,14	0,23	0,15	0,31	0,29	0,26	0,37	0,23	0,23	0,40	0,36	0,36	0,35	0,32	0,28	0,37	0,26	0,27
Monolabel	C+I	oui	Remp. Cat	1-2	0,14	0,18	0,11	0,30	0,24	0,20	0,24	0,15	0,16	0,32	0,29	0,27	0,24	0,23	0,20	0,35	0,27	0,27
Monolabel	C+I	oui	Remp. Cat	2	0,14	0,16	0,12	0,32	0,29	0,25	0,22	0,17	0,17	0,32	0,32	0,30	0,39	0,32	0,29	0,28	0,15	0,16
Monolabel	C+I	oui	Supprimer	1	0,18	0,19	0,13	0,40	0,35	0,34	0,33	0,26	0,26	0,39	0,35	0,35	0,27	0,32	0,28	0,33	0,23	0,25
Monolabel	C+I	oui	Supprimer	1-2	0,12	0,23	0,15	0,34	0,23	0,20	0,22	0,18	0,17	0,25	0,27	0,25	0,19	0,21	0,17	0,29	0,11	0,14
Monolabel	C+I	oui	Supprimer	2	0,35	0,26	0,20	0,32	0,28	0,25	0,31	0,23	0,22	0,31	0,32	0,29	0,36	0,33	0,30	0,38	0,2	0,23
Multilabel	C	non	Conserver	1	0,20	0,18	0,19	0,74	0,05	0,09	0,59	0,12	0,16	0,42	0,20	0,25	0,57	0,02	0,03	0,4	0,21	0,27
Multilabel	C	non	Conserver	1-2	0,25	0,09	0,12	0,59	0,04	0,08	0,53	0,10	0,15	0,53	0,15	0,21	0,68	0,02	0,04	0,47	0,18	0,24
Multilabel	C	non	Conserver	1-2	0,23	0,11	0,15	0,74	0,02	0,04	0,45	0,08	0,13	0,42	0,10	0,15	0,54	0,03	0,06	0,39	0,15	0,2
Multilabel	C	non	Remp. Cat	1	0,21	0,19	0,20	0,81	0,04	0,08	0,44	0,13	0,18	0,43	0,20	0,26	0,73	0,02	0,04	0,46	0,22	0,28
Multilabel	C	non	Remp. Cat	1-2	0,22	0,14	0,17	0,53	0,02	0,04	0,38	0,06	0,10	0,35	0,08	0,12	0,55	0,02	0,03	0,44	0,16	0,2
Multilabel	C	non	Remp. Cat	2	0,24	0,09	0,12	0,66	0,04	0,06	0,55	0,09	0,14	0,56	0,12	0,18	0,56	0,02	0,04	0,45	0,22	0,27
Multilabel	C	non	Supprimer	1	0,23	0,18	0,20	0,84	0,06	0,10	0,44	0,10	0,15	0,43	0,21	0,27	0,71	0,02	0,03	0,37	0,2	0,24
Multilabel	C	non	Supprimer	1-2	0,21	0,08	0,11	0,83	0,04	0,08	0,57	0,11	0,16	0,50	0,15	0,22	0,45	0,03	0,05	0,5	0,16	0,22
Multilabel	C	non	Supprimer	1-2	0,22	0,12	0,15	0,70	0,03	0,06	0,49	0,11	0,15	0,54	0,08	0,12	0,49	0,02	0,04	0,32	0,12	0,15
Multilabel	C	oui	Conserver	1	0,20	0,25	0,22	0,73	0,06	0,11	0,51	0,12	0,16	0,41	0,24	0,29	0,41	0,04	0,07	0,4	0,25	0,29
Multilabel	C	oui	Conserver	2	0,26	0,10	0,14	0,89	0,06	0,11	0,44	0,08	0,12	0,56	0,17	0,24	0,45	0,03	0,05	0,47	0,23	0,28
Multilabel	C	oui	Conserver	2	0,25	0,15	0,18	0,58	0,03	0,06	0,42	0,09	0,13	0,39	0,08	0,13	0,41	0,03	0,06	0,45	0,17	0,23
Multilabel	C	oui	Remp. Cat	1	0,20	0,23	0,21	0,68	0,05	0,09	0,48	0,10	0,14	0,35	0,23	0,26	0,45	0,03	0,05	0,38	0,23	0,28

Multilabel	C	oui	Remp. Cat	1-2	0,23	0,10	0,13	0,81	0,06	0,10	0,44	0,09	0,13	0,42	0,14	0,19	0,69	0,02	0,04	0,47	0,22	0,29
Multilabel	C	oui	Remp. Cat	2	0,28	0,13	0,17	0,63	0,03	0,06	0,42	0,08	0,12	0,40	0,10	0,15	0,55	0,03	0,05	0,45	0,17	0,22
Multilabel	C	oui	Supprimer	1	0,18	0,20	0,19	0,68	0,04	0,07	0,40	0,11	0,15	0,36	0,22	0,26	0,48	0,05	0,08	0,39	0,23	0,28
Multilabel	C	oui	Supprimer	1-2	0,28	0,08	0,11	0,91	0,05	0,09	0,56	0,10	0,14	0,49	0,15	0,21	0,60	0,02	0,04	0,46	0,21	0,27
Multilabel	C	oui	Supprimer	2	0,24	0,12	0,16	0,65	0,03	0,06	0,51	0,09	0,13	0,33	0,09	0,14	0,64	0,03	0,05	0,39	0,16	0,21
Multilabel	C+I	non	Conserver	1	0,27	0,25	0,25	0,75	0,15	0,18	0,56	0,15	0,20	0,44	0,28	0,32	0,60	0,13	0,16	0,43	0,31	0,35
Multilabel	C+I	non	Conserver	1-2	0,24	0,19	0,21	0,65	0,13	0,17	0,54	0,13	0,16	0,53	0,17	0,21	0,53	0,15	0,16	0,48	0,21	0,26
Multilabel	C+I	non	Conserver	1-2	0,29	0,16	0,20	0,73	0,14	0,18	0,48	0,17	0,21	0,49	0,21	0,27	0,69	0,15	0,17	0,51	0,28	0,33
Multilabel	C+I	non	Remp. Cat	1	0,24	0,22	0,22	0,64	0,18	0,22	0,54	0,16	0,21	0,44	0,29	0,34	0,76	0,14	0,16	0,41	0,25	0,3
Multilabel	C+I	non	Remp. Cat	1-2	0,23	0,17	0,19	0,62	0,12	0,14	0,44	0,16	0,19	0,47	0,16	0,19	0,54	0,17	0,17	0,44	0,21	0,24
Multilabel	C+I	non	Remp. Cat	2	0,28	0,17	0,20	0,74	0,15	0,19	0,43	0,18	0,20	0,55	0,24	0,30	0,81	0,12	0,14	0,48	0,28	0,32
Multilabel	C+I	non	Supprimer	1	0,25	0,21	0,23	0,82	0,14	0,17	0,47	0,11	0,15	0,42	0,26	0,29	0,62	0,14	0,17	0,43	0,29	0,33
Multilabel	C+I	non	Supprimer	1-2	0,30	0,17	0,20	0,75	0,16	0,19	0,51	0,20	0,24	0,54	0,22	0,29	0,58	0,14	0,17	0,46	0,26	0,31
Multilabel	C+I	non	Supprimer	2	0,26	0,19	0,21	0,67	0,11	0,14	0,58	0,15	0,18	0,46	0,20	0,24	0,51	0,14	0,13	0,43	0,24	0,27
Multilabel	C+I	oui	Conserver	1	0,23	0,25	0,24	0,81	0,15	0,18	0,41	0,16	0,20	0,43	0,29	0,32	0,58	0,13	0,15	0,41	0,29	0,33
Multilabel	C+I	oui	Conserver	2	0,29	0,19	0,22	0,72	0,14	0,18	0,45	0,15	0,18	0,52	0,18	0,22	0,56	0,16	0,17	0,43	0,22	0,26
Multilabel	C+I	oui	Conserver	2	0,26	0,16	0,19	0,61	0,14	0,18	0,47	0,17	0,21	0,46	0,21	0,26	0,63	0,15	0,16	0,48	0,25	0,31
Multilabel	C+I	oui	Remp. Cat	1	0,25	0,30	0,27	0,78	0,17	0,20	0,44	0,21	0,23	0,40	0,29	0,33	0,58	0,13	0,15	0,41	0,3	0,34
Multilabel	C+I	oui	Remp. Cat	1-2	0,29	0,16	0,19	0,80	0,14	0,17	0,47	0,13	0,17	0,50	0,22	0,28	0,67	0,13	0,15	0,51	0,28	0,33
Multilabel	C+I	oui	Remp. Cat	1-2	0,28	0,20	0,23	0,68	0,13	0,16	0,51	0,18	0,19	0,47	0,20	0,25	0,53	0,14	0,17	0,45	0,24	0,29
Multilabel	C+I	oui	Supprimer	1	0,24	0,26	0,24	0,72	0,15	0,17	0,55	0,14	0,18	0,40	0,29	0,32	0,66	0,12	0,15	0,43	0,3	0,34
Multilabel	C+I	oui	Supprimer	1-2	0,28	0,19	0,23	0,52	0,11	0,13	0,39	0,17	0,21	0,52	0,21	0,25	0,55	0,14	0,16	0,45	0,28	0,32
Multilabel	C+I	oui	Supprimer	1-2	0,26	0,17	0,20	0,73	0,15	0,18	0,53	0,15	0,20	0,54	0,23	0,29	0,52	0,12	0,14	0,46	0,26	0,31

TABLEAU EN ANNEXE 13 : Résultats pour l'approche TF-IDF

10. Résultats pour les approches par *embedding*

Légende :

- Col. *Classes* : *C* = Critères seuls, *C+I* = Critères + Inclassifiables
- *Lemm.* = *Lemmatisation*
- Col. *Mots outils* : *Remp. Cat* = mots outils remplacés par l'étiquette de leur catégorie morphosyntaxique
- Col. *Entr.* : *Corpus* = modèle entraîné sur notre corpus, *Pré-entr.* : modèle pré-entraîné (*spaCy*)
- Classifieurs : *NB* = Naïve Bayes, *SVM* = Support Machine Vector, *KNN* = k-Nearest Neighbors, *LR* = Logistic Regression, *RF* = Random Forest, *MLP* = Multilayer Perceptron

Classification	Classes	Entr.	Lemm.	Mots outils	NB			SVM			KNN			LR			RF			MLP		
					p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1	p	r	f1
Monolabel	C	Corpus	non	Conserver	0,22	0,16	0,15	0,20	0,18	0,18	0,10	0,10	0,09	0,18	0,17	0,17	0,13	0,18	0,15	0,14	0,13	0,13
Monolabel	C	Corpus	non	Remp. Cat	0,19	0,13	0,12	0,23	0,18	0,19	0,14	0,11	0,11	0,21	0,18	0,19	0,21	0,23	0,19	0,20	0,18	0,18
Monolabel	C	Corpus	non	Supprimer	0,16	0,12	0,11	0,26	0,20	0,21	0,17	0,12	0,12	0,26	0,20	0,21	0,20	0,23	0,20	0,22	0,19	0,19
Monolabel	C	Corpus	oui	Conserver	0,16	0,12	0,10	0,24	0,20	0,20	0,13	0,10	0,10	0,23	0,20	0,21	0,19	0,23	0,20	0,19	0,18	0,18
Monolabel	C	Corpus	oui	Remp. Cat	0,12	0,12	0,10	0,17	0,15	0,15	0,12	0,11	0,10	0,22	0,20	0,19	0,17	0,18	0,15	0,17	0,15	0,16
Monolabel	C	Corpus	oui	Supprimer	0,12	0,12	0,10	0,20	0,18	0,18	0,11	0,10	0,09	0,19	0,18	0,17	0,14	0,18	0,14	0,17	0,17	0,16
Monolabel	C	Pré-entr.	Non	Conserver	0,16	0,15	0,15	0,23	0,21	0,21	0,13	0,13	0,11	0,27	0,26	0,26	0,17	0,22	0,18	0,16	0,16	0,15
Monolabel	C	Pré-entr.	Non	Supprimer	0,19	0,13	0,14	0,28	0,26	0,26	0,20	0,16	0,16	0,33	0,32	0,31	0,21	0,19	0,17	0,21	0,25	0,21
Monolabel	C	Pré-entr.	Oui	Conserver	0,29	0,15	0,18	0,34	0,29	0,30	0,22	0,15	0,14	0,29	0,28	0,27	0,24	0,22	0,19	0,18	0,21	0,18
Monolabel	C	Pré-entr.	Oui	Supprimer	0,19	0,13	0,14	0,28	0,26	0,26	0,20	0,16	0,16	0,33	0,32	0,31	0,21	0,19	0,17	0,21	0,25	0,21
Monolabel	C+I	Corpus	non	Conserver	0,24	0,15	0,15	0,19	0,18	0,17	0,15	0,12	0,12	0,21	0,18	0,18	0,24	0,23	0,21	0,22	0,21	0,21
Monolabel	C+I	Corpus	non	Remp. Cat	0,20	0,16	0,14	0,19	0,18	0,18	0,16	0,13	0,12	0,26	0,18	0,20	0,24	0,21	0,18	0,18	0,17	0,17
Monolabel	C+I	Corpus	non	Supprimer	0,21	0,14	0,13	0,17	0,17	0,16	0,19	0,14	0,14	0,24	0,21	0,22	0,19	0,19	0,17	0,20	0,20	0,19
Monolabel	C+I	Corpus	oui	Conserver	0,25	0,14	0,14	0,20	0,18	0,18	0,15	0,11	0,11	0,21	0,18	0,18	0,22	0,22	0,20	0,20	0,18	0,18
Monolabel	C+I	Corpus	oui	Remp. Cat	0,19	0,14	0,13	0,23	0,21	0,20	0,12	0,14	0,12	0,19	0,20	0,19	0,15	0,21	0,17	0,13	0,13	0,12
Monolabel	C+I	Corpus	oui	Supprimer	0,19	0,18	0,15	0,23	0,21	0,20	0,14	0,17	0,14	0,25	0,26	0,24	0,15	0,20	0,16	0,13	0,12	0,12
Monolabel	C+I	Pré-entr.	Non	Conserver	0,22	0,15	0,16	0,30	0,24	0,25	0,12	0,10	0,09	0,31	0,29	0,29	0,17	0,20	0,18	0,17	0,19	0,17
Monolabel	C+I	Pré-entr.	Non	Supprimer	0,22	0,15	0,16	0,30	0,24	0,25	0,12	0,10	0,09	0,31	0,29	0,29	0,17	0,20	0,18	0,17	0,19	0,17

Monolabel	C+I	Pré-entr.	Oui	Conserver	0,17	0,14	0,15	0,27	0,25	0,26	0,20	0,16	0,15	0,29	0,26	0,27	0,22	0,23	0,19	0,17	0,18	0,16
Monolabel	C+I	Pré-entr.	Oui	Supprimer	0,17	0,13	0,14	0,23	0,20	0,21	0,16	0,15	0,14	0,26	0,22	0,23	0,20	0,18	0,16	0,18	0,17	0,17
Multilabel	C	Corpus	non	Conserver	0,17	0,50	0,24	1,00	0,00	0,00	0,47	0,01	0,02	0,15	0,11	0,13	0,45	0,03	0,05	0,17	0,15	0,05
Multilabel	C	Corpus	non	Remp. Cat	0,17	0,44	0,23	1,00	0,00	0,00	0,62	0,03	0,05	0,16	0,11	0,13	0,29	0,03	0,04	0,18	0,19	0,04
Multilabel	C	Corpus	non	Supprimer	0,17	0,42	0,22	1,00	0,00	0,00	0,48	0,01	0,01	0,16	0,12	0,13	0,40	0,03	0,05	0,17	0,18	0,05
Multilabel	C	Corpus	oui	Conserver	0,19	0,55	0,24	1,00	0,00	0,00	0,46	0,00	0,01	0,18	0,10	0,13	0,40	0,03	0,05	0,17	0,17	0,05
Multilabel	C	Corpus	oui	Remp. Cat	0,18	0,48	0,23	1,00	0,00	0,00	0,58	0,01	0,01	0,18	0,12	0,14	0,32	0,02	0,03	0,16	0,17	0,03
Multilabel	C	Corpus	oui	Supprimer	0,17	0,54	0,23	1,00	0,00	0,00	0,45	0,01	0,02	0,13	0,10	0,11	0,43	0,02	0,04	0,15	0,13	0,04
Multilabel	C	Pré-entr.	Non	Conserver	0,19	0,58	0,28	0,29	0,19	0,22	0,40	0,03	0,05	0,27	0,31	0,28	0,60	0,03	0,05	0,30	0,27	0,28
Multilabel	C	Pré-entr.	Non	Supprimer	0,18	0,57	0,26	0,32	0,19	0,23	0,45	0,03	0,05	0,25	0,31	0,27	0,50	0,04	0,06	0,29	0,27	0,28
Multilabel	C	Pré-entr.	Oui	Conserver	0,19	0,63	0,28	0,30	0,19	0,23	0,53	0,03	0,05	0,25	0,31	0,27	0,39	0,03	0,05	0,30	0,30	0,30
Multilabel	C	Pré-entr.	Oui	Supprimer	0,19	0,59	0,28	0,27	0,17	0,20	0,49	0,04	0,06	0,29	0,34	0,31	0,54	0,02	0,04	0,30	0,28	0,28
Multilabel	C+I	Corpus	non	Conserver	0,25	0,49	0,30	0,91	0,12	0,11	0,52	0,09	0,10	0,22	0,19	0,20	0,34	0,13	0,15	0,22	0,23	0,15
Multilabel	C+I	Corpus	non	Remp. Cat	0,21	0,49	0,27	0,89	0,12	0,11	0,61	0,11	0,14	0,25	0,20	0,22	0,32	0,09	0,11	0,19	0,20	0,11
Multilabel	C+I	Corpus	non	Supprimer	0,21	0,41	0,20	0,89	0,12	0,11	0,51	0,09	0,11	0,24	0,20	0,21	0,39	0,11	0,14	0,23	0,22	0,14
Multilabel	C+I	Corpus	oui	Conserver	0,21	0,54	0,27	0,91	0,02	0,03	0,58	0,10	0,11	0,23	0,18	0,20	0,38	0,11	0,13	0,21	0,22	0,13
Multilabel	C+I	Corpus	oui	Conserver	0,23	0,45	0,22	0,93	0,01	0,01	0,46	0,06	0,08	0,21	0,17	0,18	0,38	0,11	0,13	0,21	0,21	0,13
Multilabel	C+I	Corpus	oui	Remp. Cat	0,25	0,43	0,24	0,89	0,01	0,02	0,53	0,13	0,13	0,23	0,19	0,20	0,30	0,10	0,11	0,21	0,21	0,11
Multilabel	C+I	Pré-entr.	Non	Conserver	0,27	0,59	0,33	0,30	0,23	0,25	0,54	0,13	0,16	0,29	0,32	0,30	0,39	0,13	0,14	0,35	0,33	0,34
Multilabel	C+I	Pré-entr.	Non	Supprimer	0,27	0,60	0,32	0,37	0,27	0,31	0,52	0,13	0,14	0,31	0,36	0,33	0,56	0,14	0,15	0,32	0,29	0,30
Multilabel	C+I	Pré-entr.	Oui	Conserver	0,27	0,57	0,31	0,36	0,25	0,29	0,49	0,12	0,13	0,30	0,34	0,31	0,53	0,12	0,13	0,33	0,32	0,32
Multilabel	C+I	Pré-entr.	Oui	Supprimer	0,27	0,58	0,31	0,38	0,24	0,29	0,57	0,12	0,13	0,28	0,31	0,29	0,55	0,13	0,16	0,32	0,33	0,32

TABLEAU EN ANNEXE 14 : Résultats pour l'approche par embedding

11. Performances en fonction des méthodes de vectorisation et des algorithmes

	C						C+I					
	Monolabel			Multilabel			Monolabel			Multilabel		
	p	r	f	p	r	f	p	r	f	p	r	f
MoyScore BOW	0,26	0,23	0,21	0,38	0,14	0,17	0,25	0,22	0,20	0,41	0,22	0,25
MoyScores TF-IDF	0,29	0,26	0,23	0,48	0,25	0,23	0,29	0,24	0,22	0,46	0,24	0,22
MoyScores W2V Pré-entraîné	0,23	0,21	0,20	0,34	0,24	0,19	0,21	0,19	0,19	0,38	0,29	0,25
MoyScores W2V Sur corpus	0,18	0,16	0,15	0,40	0,13	0,08	0,19	0,18	0,17	0,41	0,19	0,15

TABLEAU EN ANNEXE 15 : Performances par approches de vectorisation

↑p134

	C						C+I					
	Monolabel			Multilabel			Monolabel			Multilabel		
	p	r	f	p	r	f	p	r	f	p	r	f
MoyScores NB	0,23	0,20	0,15	0,40	0,24	0,16	0,23	0,19	0,15	0,41	0,27	0,20
MoyScores SVM	0,28	0,25	0,23	0,51	0,21	0,21	0,28	0,26	0,23	0,50	0,25	0,25
MoyScores KNN	0,21	0,16	0,15	0,41	0,11	0,12	0,21	0,15	0,15	0,42	0,15	0,16
MoyScores LR	0,30	0,30	0,28	0,41	0,21	0,23	0,28	0,28	0,26	0,41	0,26	0,27
MoyScores RF	0,27	0,26	0,23	0,44	0,16	0,16	0,27	0,26	0,22	0,44	0,22	0,21
MoyScores MLP	0,27	0,20	0,20	0,34	0,20	0,21	0,27	0,18	0,19	0,37	0,24	0,24

TABLEAU EN ANNEXE 16 : Performances par algorithmes

↑p134

12. Mise en parallèle des observations textométriques et des résultats des classifieurs

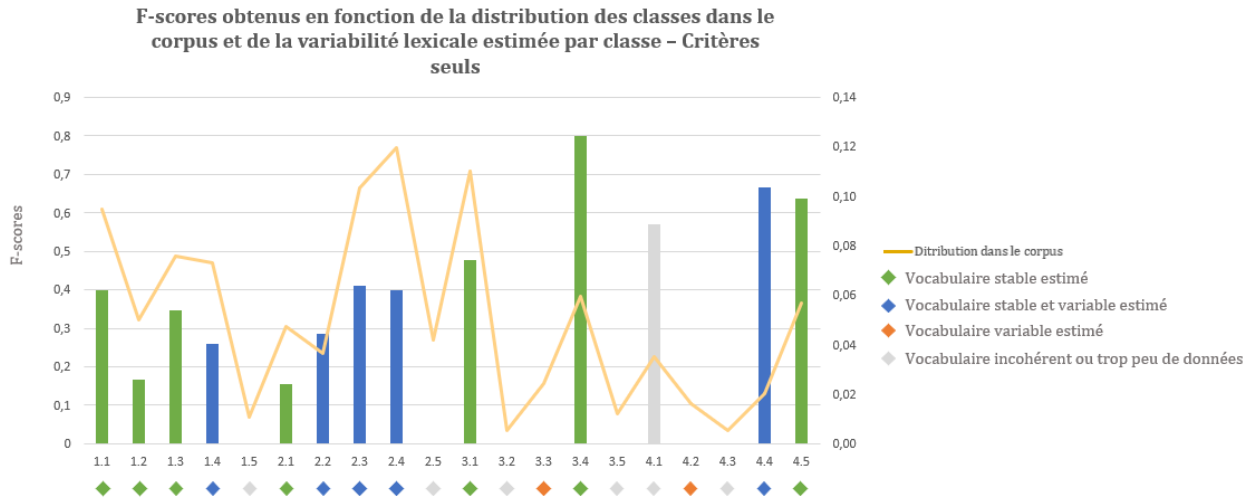


FIGURE EN ANNEXE 10 : F-scores obtenus en fonction de la distribution des classes dans le corpus et de la variabilité lexicale estimée par classe – Critères seuls

↑p136

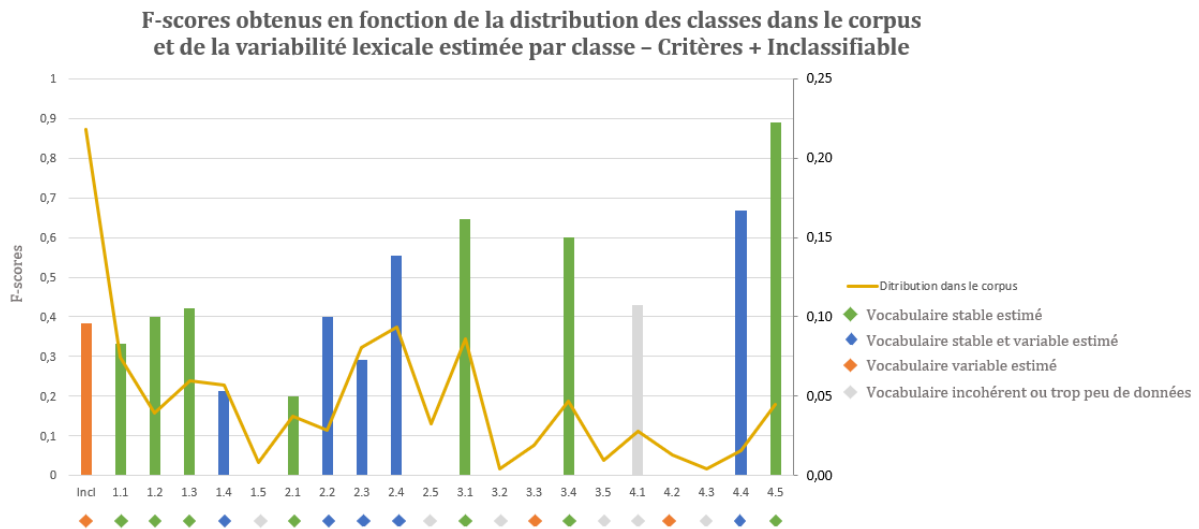


FIGURE EN ANNEXE 11 : F-scores obtenus en fonction de la distribution des classes dans le corpus et de la variabilité lexicale estimée par classe – Critères + Inclassifiable

↑p136

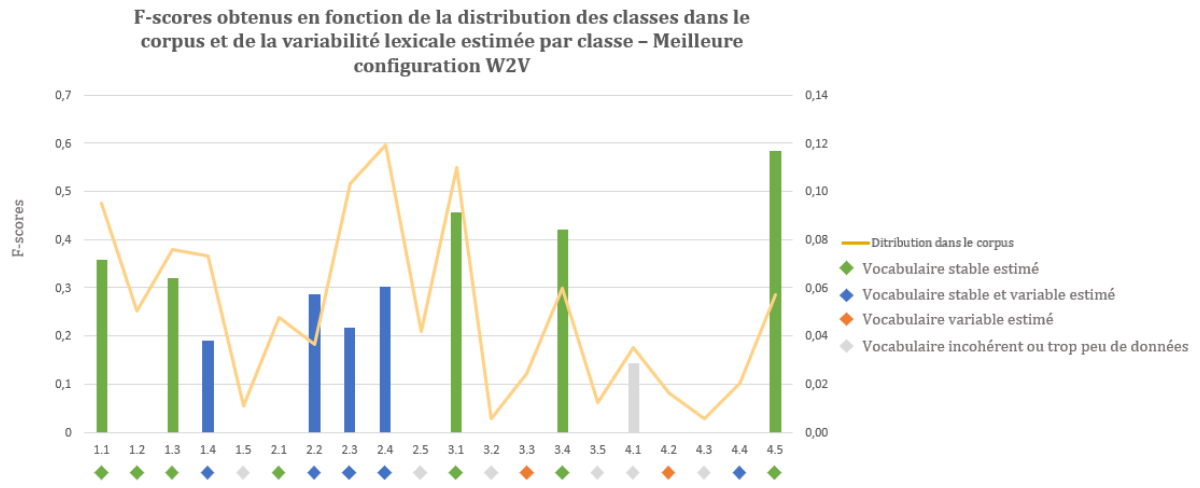


FIGURE EN ANNEXE 12 : F-scores obtenus en fonction de la distribution des classes dans le corpus et de la variabilité lexicale estimée par classe – Meilleure configuration W2V

ANNEXE 5 : PROJECTION DE VERBÆNET

13. Corpus annoté

Phrase : « ça peut donner une photographie pour dire que le R n'est pas bon mais ça n'aura pas bougé par rapport à la semaine d'avant. »

```
<verbatim id="3" etude="EC1_Elec" elec="oui" niveau="2" critere="2.2" origine="IW"
type="discours">
<part code="EC1_Elec_IW02_0">
  <NP type="GN" headLemma="cela" lemma="cela" headGender="" headNumber="s" len="2">
    <w wordId="1" form="ça" lemma="cela" cat="PRO">ça</w>
  </NP>
  <VERB headLemma="donner" lemma="donner" construction="semiAux"
    verbeAux="pouvoir" pattern="VERB VINF">
    <VERB type="simple" lemma="pouvoir">
      <w wordId="2" form="peut" lemma="pouvoir" cat="VER"
        conj="ind:pre:3s;">peut</w>
    </VERB>
    <VINF type="simple" lemma="donner">
      <w wordId="3" form="donner" lemma="donner" cat="VER" conj="inf">donner</w>
    </VINF>
  </VERB>
  <NP type="GN" headLemma="photographie" lemma="photographie" form="une
photographie" pattern="DET NOM" headGender="f" headNumber="s" len="2">
    <w wordId="4" form="une" lemma="un" cat="DET">une</w>
    <w wordId="5" form="photographie" lemma="photographie" cat="NOM" gender="f"
number="s">photographie</w>
  </NP>
  <PREP headLemma="pour">
    <w wordId="6" form="pour" lemma="pour" cat="PRP">pour</w>
  </PREP>
  <VINF type="simple" lemma="dire">
    <w wordId="7" form="dire" lemma="dire" cat="VER" conj="inf;">dire</w>
  </VINF>
  <PSUBJIND headLemma="être">
    <KON headLemma="que">
      <w wordId="8" form="que" lemma="que" cat="KON">que</w>
    </KON>
    <NP type="GN" headLemma="r" lemma="r" form="le r" pattern="DET NOM"
headGender="" headNumber="" len="2">
      <w wordId="9" form="le" lemma="le" cat="DET">le</w>
      <w wordId="10" form="r" lemma="r" cat="NOM">R</w>
    </NP>
    <VERB type="simple" lemma="être" neg="True">
      <w worded="11" form="n'" lemma="ne" cat="ADV">n'</w>
      <w wordId="12" form="est" lemma="être" cat="VER" conj="ind:pre:3s;">est</w>
      <w wordId="13" form="pas" lemma="pas" cat="ADV">pas</w>
    </VERB>
    <ADJ type="GN" headLemma="bon" headGender="m" headNumber="s" len="7">
      <w wordId="14" form="bon" lemma="bon" cat="ADJ" gender="m" number="s">bon</w>
    </ADJ>
  </PSUBJIND>
  <KON headLemma="mais">
    <w wordId="15" form="mais" lemma="mais" cat="KON">mais</w>
  </KON>
  <NP type="GN" headLemma="cela" lemma="cela" headGender="" headNumber="s" len="2">
    <w wordId="16" form="ça" lemma="cela" cat="PRO">ça</w>
  </NP>
  <VERB type="composed" lemma="bouger" neg="True">
    <w wordId="17" form="n'" lemma="ne" cat="ADV">n'</w>
```

```

<w wordId="18" form="aura" lemma="avoir" cat="VER" conj="ind:fut:3s;">aura</w>
<w wordId="19" form="pas" lemma="pas" cat="ADV">pas</w>
<w wordId="20" form="bougé" lemma="bouger" cat="VER" gender="m" number="s"
conj="par:pas;">bougé</w>
</VERB>
<PREP lemma="par rapport à">
  <w wordId="21" form="par" lemma="par" cat="PRP">par</w>
  <w wordId="22" form="rapport" lemma="rapport" cat="NOM" gender="m"
number="s">rapport</w>
  <w wordId="23" form="à" lemma="à" cat="PRP">à</w>
</PREP>
<NP type="GN" headLemma="semaine" lemma="semaine de avant" form="la semaine d'
avant" pattern="DET NOM PREP NOM" headGender="f" headNumber="f" len="4">
  <w wordId="24" form="la" lemma="le" cat="DET">la</w>
  <w wordId="25" form="semaine" lemma="semaine" cat="NOM" gender="f"
number="s">semaine</w>
  <w wordId="26" form="d'" lemma="de" cat="PRP">d'</w>
  <w wordId="27" form="avant" lemma="avant" cat="NOM" gender="m"
number="s">avant</w>
</NP>
<w wordId="28" form="." lemma="." cat="SENT">.</w>
</part>
</verbatim>

```

FIGURE EN ANNEXE 13 : Exemple d'étiquetage morphosyntaxique du corpus pour l'expérience Verbænet

↑p151

Phrase : « je pense que j'aurais un bel outil parce que c'est relativement puissant mais que je n'aurais ni le temps ni les moyens d'exploiter. »

Les cadres annotés avec Verbænet sont indiqués par les balises <VN>. Plusieurs éléments étiquetés avec cette balise et ayant la même valeur d'attribut ref appartiennent au même frame.

```

<verbatim id"34" etude="EC1_Elec" elec="oui" niveau="2" critere="2.5"
origine="IW" type="discours">
  <part code="EC1_Elec_IW02_0">
    <VN cat="NP" ref="9" role="Agent">
      <NP type="GN" headLemma="je" lemma="je" headGender="" headNumber="s" len="2">
        <w wordId="181" form="je" lemma="je" cat="PRO">je</w>
      </NP>
    </VN>
    <VN cat="VERB" ref="9" class="consider-29.9">
      <VERB type="simple" lemma="penser">
        <w wordId="182" form="pense" lemma="penser" cat="VER"
conj="imp:pre:2s;ind:pre:1s;ind:pre:3s;sub:pre:1s;sub:pre:3s;">pense</w>
      </VERB>
    </VN>
    <VN cat="PSUBJIND" ref="9" role="Theme">
      <PSUBJIND headLemma="avoir">
        <KON headLemma="que">
          <w wordId="183" form="que" lemma="que" cat="KON">que</w>
        </KON>
        <VN cat="NP" ref="3-1" role="Agent">
          <NP type="GN" headLemma="je" lemma="je" headGender="" headNumber=""
len="2">
            <w wordId="184" form="j'" lemma="je" cat="PRO">j'</w>
          </NP>
        </VN>
        <VN cat="VERB" ref="3-1" class="hire-13.5.3">
          <VERB type="simple" lemma="avoir">

```



```

    <w wordId="185" form="aurais" lemma="avoir" cat="VER"
    conj="cnd:pre:1s;cnd:pre:2s; ">aurais</w>
  </VERB>
</VN>
<VN cat="NP" ref="3-1" role="Theme">
  <NP type="GN" headLemma="outil" lemma="beau outil" form="un bel outil"
  pattern="DET ADJ NOM" headGender="m" headNumber="s" len="3">
    <w wordId="186" form="un" lemma="un" cat="DET">un</w>
    <w wordId="187" form="bel" lemma="beau" cat="ADJ" gender="m"
    number="s">bel</w>
    <w wordId="188" form="outil" lemma="outil" cat="NOM" gender="m"
    number="s">outil</w>
  </NP>
</VN>
</PSUBJIND>
</VN>
<KON lemma="parce que">
  <w wordId="189" form="parce" lemma="parce" cat="KON">parce</w>
  <w wordId="190" form="que" lemma="que" cat="KON">que</w>
</KON>
<NP type="GN" headLemma="ce" lemma="ce" headGender="" headNumber="" len="2">
  <w wordId="191" form="c'" lemma="ce" cat="PRO">c'</w>
</NP>
<VERB type="simple" lemma="être">
  <w wordId="192" form="est" lemma="être" cat="VER" conj="ind:pre:3s;">est</w>
</VERB>
<ADJ type="GN" headLemma="puissant" headGender="m" headNumber="s" len="7">
  <w wordId="193" form="relativement" lemma="relativement"
  cat="ADV">relativement</w>
  <w wordId="194" form="puissant" lemma="puissant" cat="ADJ" gender="m"
  number="s">puissant</w>
</ADJ>
<KON headLemma="mais">
  <w wordId="195" form="mais" lemma="mais" cat="KON">mais</w>
</KON>
<PSUBJIND headLemma="avoir">
  <KON headLemma="que">
    <w wordId="196" form="que" lemma="que" cat="KON">que</w>
  </KON>
  <VN cat="NP" ref="2-1" role="Agent">
    <NP type="GN" headLemma="je" lemma="je" headGender="" headNumber="s"
    len="2">
      <w wordId="197" form="je" lemma="je" cat="PRO">je</w>
    </NP>
  </VN>
  <VN cat="VERB" ref="2-1" class="hire-13.5.3">
    <VERB type="simple" lemma="avoir" neg="True">
      <w wordId="198" form="n'" lemma="ne" cat="ADV">n'</w>
      <w wordId="199" form="aurais" lemma="avoir" cat="VER"
      conj="cnd:pre:1s;cnd:pre:2s; ">aurais</w>
    </VERB>
  </VN>
  <VN cat="NP" ref="2-1" role="Theme">
  <NP headLemma="temps" type="coord" pattern="pas/ni NP , ni NP de VINF">
    <KON headLemma="ni">
      <w wordId="200" form="ni" lemma="ni" cat="KON">ni</w>
    </KON>
    <NP type="GN" headLemma="temps" lemma="temps" form="le temps"
    pattern="DET NOM" headGender="m" headNumber="s" len="2">
      <w wordId="201" form="le" lemma="le" cat="DET">le</w>
      <w wordId="202" form="temps" lemma="temps" cat="NOM" gender="m"
      number="s">temps</w>
    </NP>
    <w wordId="203" form="," lemma="," cat="PUN">,</w>
    <KON headLemma="ni">
      <w wordId="204" form="ni" lemma="ni" cat="KON">ni</w>
    </KON>
    <NP type="GN" headLemma="moyen" lemma="moyen" form="les moyens"

```

```

pattern="DET NOM" headGender="m" headNumber="p" len="2">
  <w wordId="205" form="les" lemma="le" cat="DET">les</w>
  <w wordId="206" form="moyens" lemma="moyen" cat="NOM" gender="m"
number="p">moyens</w>
</NP>
<PREP headLemma="de">
  <w wordId="207" form="d'" lemma="de" cat="PRP">d'</w>
</PREP>
<VINF type="simple" lemma="exploiter">
  <w wordId="208" form="exploiter" lemma="exploiter" cat="VER"
conj="inf; ">exploiter</w>
</VINF>
</NP>
</VN>
</PSUBJIND>
<w wordId="209" form="." lemma="." cat="SENT">.</w>

```

FIGURE EN ANNEXE 14 : Extrait du corpus étiqueté avec Verbānet

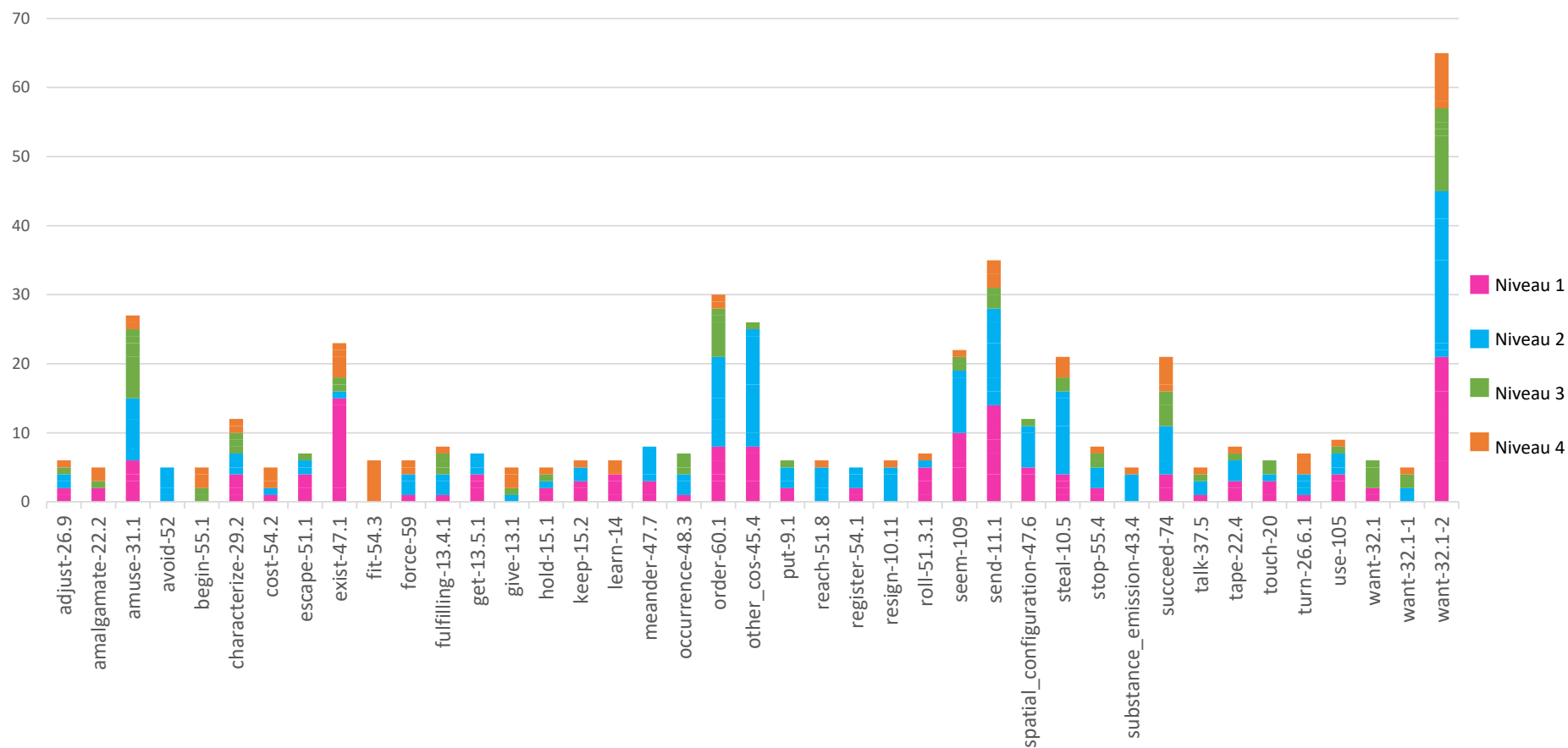


FIGURE EN ANNEXE 15 : Répartition des classes de Verbænet dans les différents Niveaux

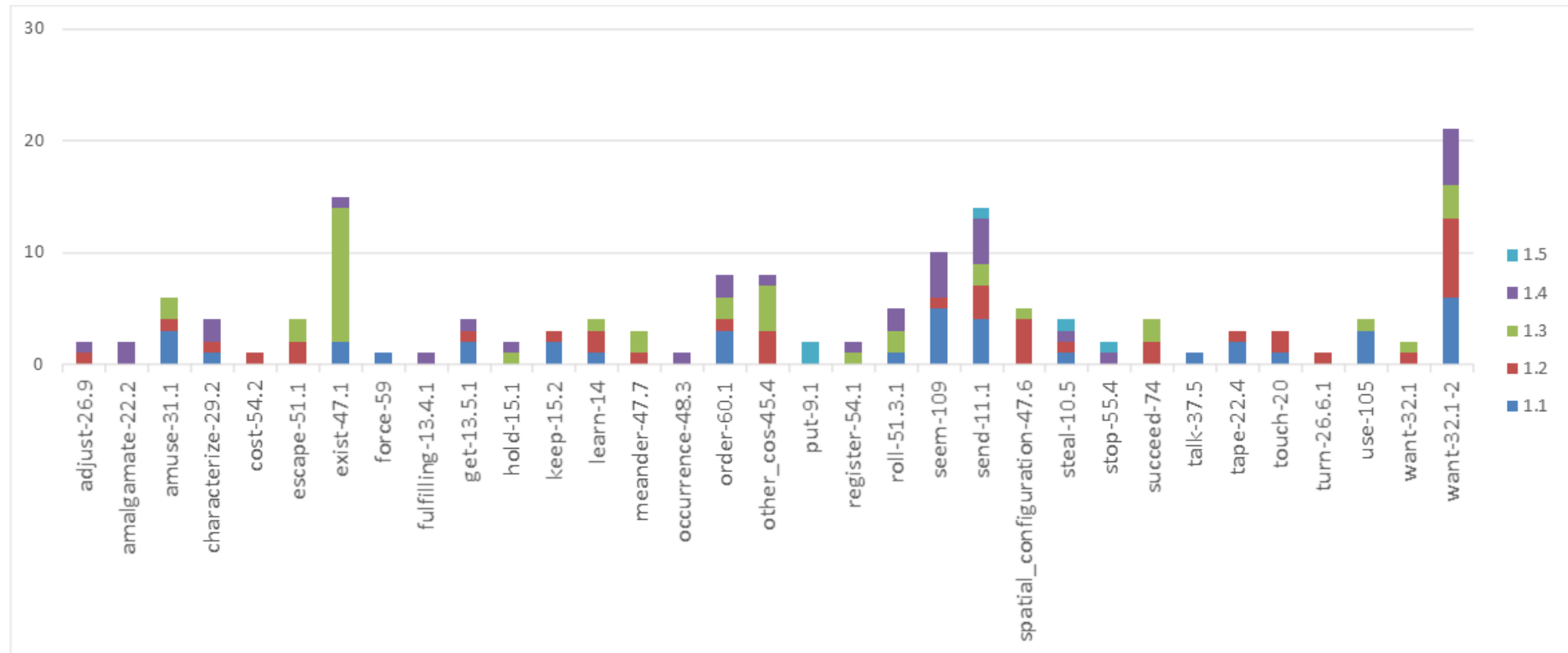


FIGURE EN ANNEXE 16 : Répartition des classes de Verbænet dans les différents Critères du Niveau 1

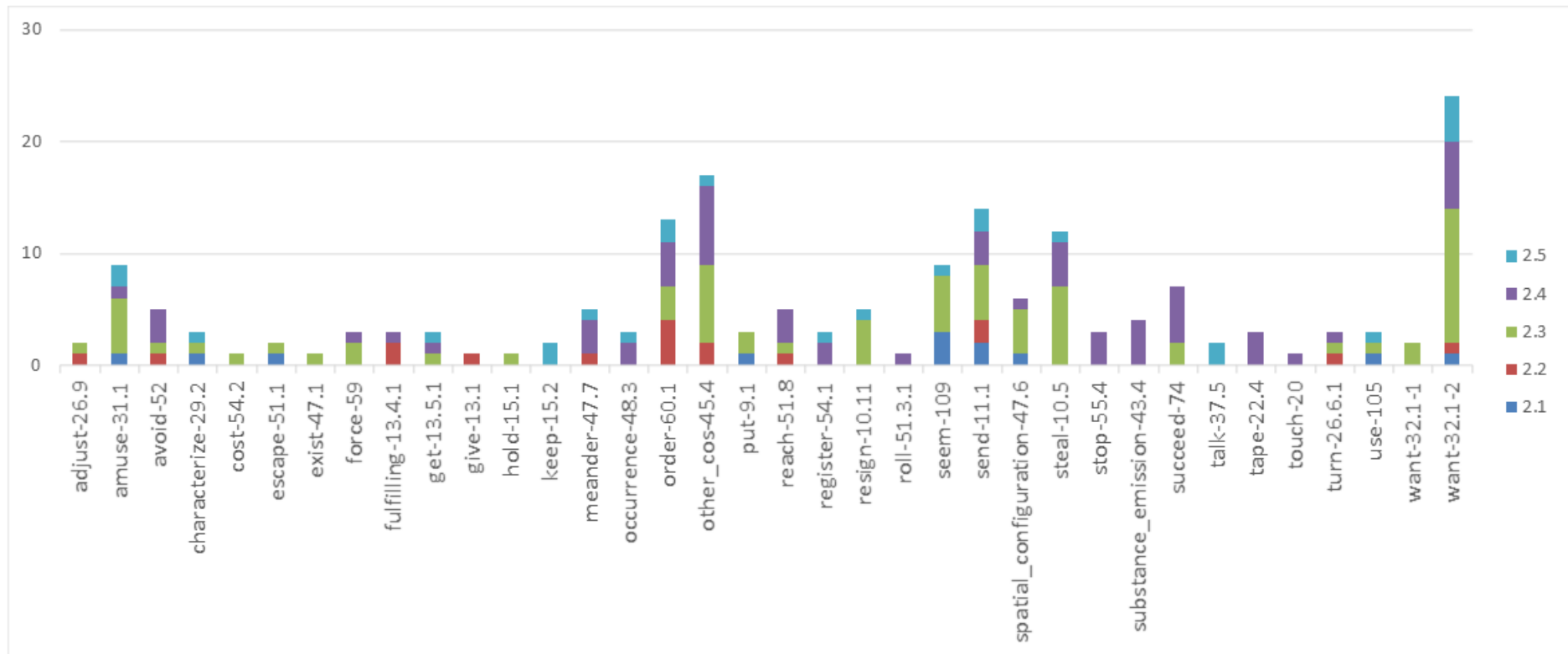


FIGURE EN ANNEXE 17 : Répartition des classes de Verbænet dans les différents Critères du Niveau 2

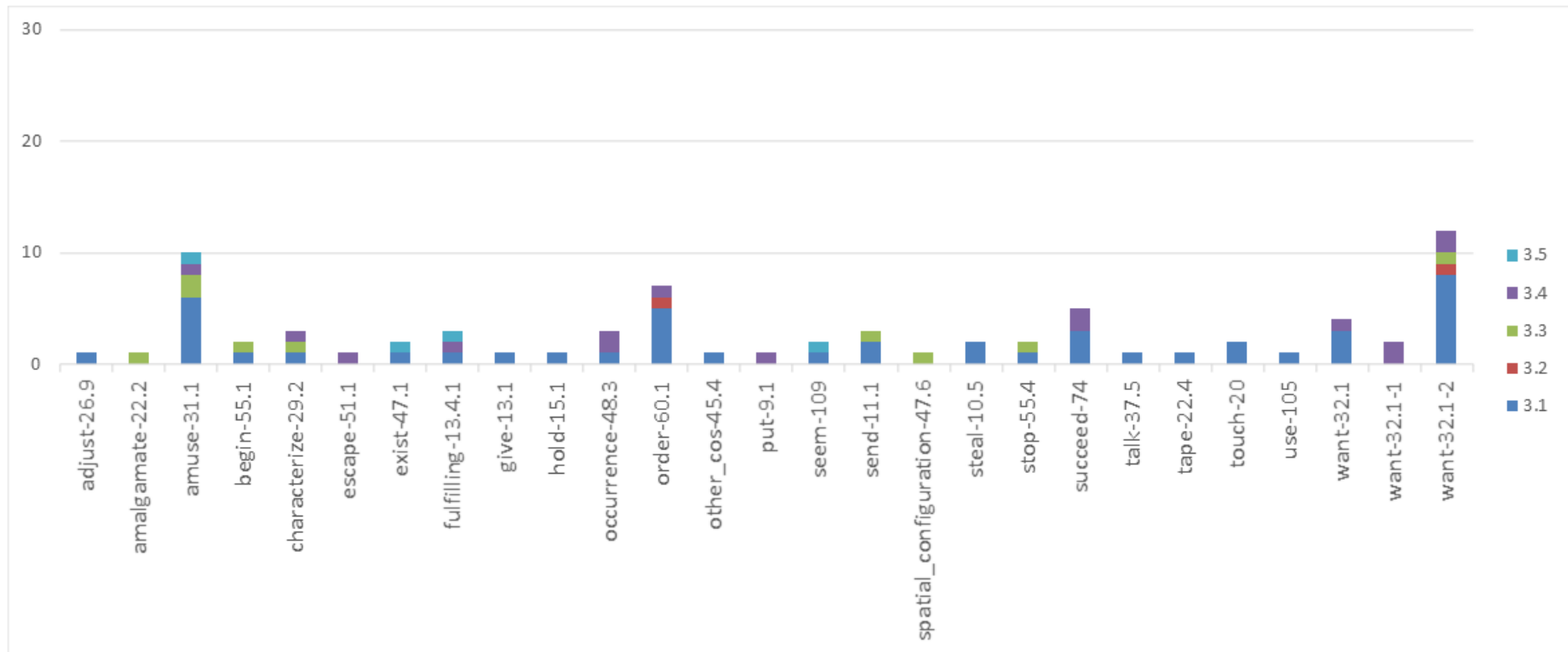


FIGURE EN ANNEXE 18 : Répartition des classes de Verbænet dans les différents Critères du Niveau 3

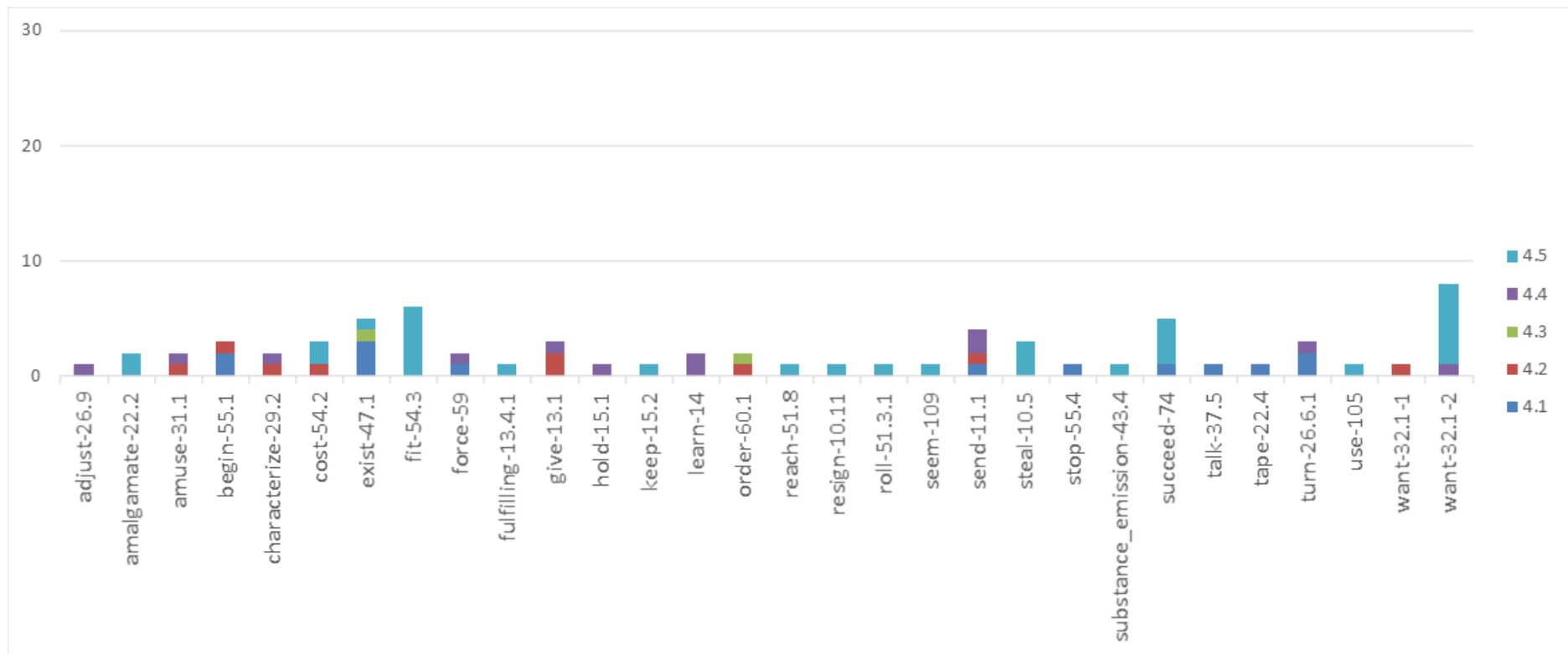


FIGURE EN ANNEXE 19 : Répartition des classes de Verbænet dans les différents Critères du Niveau 4

ANNEXE 6 : MODÉLISATION EXPERTE

14. Analyse de la collecte de verbatim types

Légende du tableau ci-dessous :

Vert : motivation

Bleu : condition

Rouge : Frein

Critère	Verbatim	Patrons	Indices courts	Type d'éléments
1.1 – L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?	<i>C'est incroyable, c'est le rêve, c'est exactement tout ce qu'on vient d'exprimer comme étant nos attentes holala, vous m'avez saoulé avec votre présentation, ça m'intéresse pas du tout</i>	<i>j'ai pas compris [moment de la présentation]</i>	<i>j'ai du mal je vois pas trop</i>	<ul style="list-style-type: none"> - Questions sur le concept en général - Impressions générales - Manifestation d'intérêt sur la globalité du concept et non des fonctions spécifiques
1.2 – Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?	<i>C'est super intelligent de faire le lien entre ma maison et mon téléphone oui, mais là, est-ce qu'il faut avoir une connexion Internet ? ça fait boîte noire quoi !</i>	<i>j'ai compris qu'il y avait besoin de [élément technique nécessaire au fonctionnement du concept]</i>	<i>boîte noire opaque comment, pourquoi, où</i>	<ul style="list-style-type: none"> - Question ou expression de la compréhension en mentionnant un aspect technique
1.3 – L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?	<i>par contre, là-dessus, je ne connais personne qui propose ça c'est une nouvelle manière de voir les choses ça existe déjà</i>	<i>ça me fait penser à [produit/service] c'est ce que fait [entreprise]</i>	<i>c'est comme</i>	<ul style="list-style-type: none"> - Comparaison avec des techniques utilisées par la concurrence

1.4 – L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?	<i>Si j'ai bien compris, je peux accéder à l'appareil photo et à mes mails dans un même appareil</i>	<i>moi, ce qui m'intéresse, c'est [une des fonctions du concept] j'aime bien pouvoir faire [action] grâce à [fonction du concept]</i>		- Nom ou description de fonctions associées à une marque d'intérêt
1.5 – L'usage du concept est-il envisageable comme un système d'usage courant ?	<i>Je l'utiliserais quand je prendrai ma douche je vais l'utiliser tous les jours</i>	<i>je pourrais l'utiliser pour faire [action] ah, et je pourrais l'utiliser aussi pour [contexte d'utilisation]</i>	<i>quotidien, occasionnellement, rarement</i>	- Différentes sphères d'utilisation (cercle privé et cercle professionnel, tâches et contextes divers) - Champ lexical de la fréquence
2.1 – Le concept est-il simple à utiliser ?	<i>ça a l'air super simple en pratique, comment je fais ?</i>	<i>je trouverais ça [adjectif relatif à la facilité/complexité]</i>	<i>intuitif en pratique usine à gaz</i>	- Verbes d'action à la première personne
2.2 – Le concept compléterait-il sans les concurrencer les habitudes existantes ?	<i>ça existe déjà</i>	<i>ça va me permettre de [faire quelque chose] je fais déjà [action] avec [produit/service concurrent]</i>	<i>améliorer, compléter, enrichir</i>	- Verbes d'action
2.3 – La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?	<i>ah bah ça va me faire gagner du temps on a l'habitude de faire par papier, maintenant c'est sur le téléphone ça va être encore plus chiant que quand je l'écrivais sur papier</i>			
2.4 – Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?	<i>j'aurais plus à prendre ma voiture bah comme ça, j'aurais plus à me priver de manger des frites tous les jours</i>	<i>je perdrais plus mon temps à chercher [quelque chose] j'aurais plus à m'emmerder avec [quelque chose]</i>	<i>j'aurais plus à je pourrai(s) Hallelujah</i>	- Conditionnel ou futur

2.5 – L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ?	<i>je ferais les choses différemment comment je vais m'organiser avec ma femme ? il va falloir que je parte de chez moi un peu plus tôt le matin</i>	<i>je vais pouvoir abandonner [action ou produit/service concurrent]</i>	<i>abandonner</i>	- Conditionnel ou futur - 314 éléments en rapport avec la notion du changement
3.1 – L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?	<i>c'est pas pour moi</i>	<i>ça s'adresse plutôt à [personnes / organisation / corps de métier] je vois ça plutôt pour [personnes / organisation / corps de métier]</i>	<i>moi, les autres ça s'adresse (plutôt) à jeunes, vieux, connectés</i>	- Caractéristiques personnelles
3.2 – Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?	<i>ça va influencer mes responsabilités ma situation ne va pas changer mon métier va devoir évoluer c'est la mort du métier</i>	<i>ça me permettrait d'[être quelque chose] qui est responsable de [malfaçon] ? ça va prendre la place de [corps de métier]</i>	<i>être mieux reconnu par changer mon mon s'occuper de</i>	
3.3 – Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?	<i>qu'est-ce que ça va faire de mes données ? ça m'obligerait à avoir les épaules dénudées mais c'est pas français</i>	<i>chez nous, on fait [pratique] ça m'obligerait à renoncer à [quelque chose]</i>	<i>chez nous</i>	- 314 éléments en lien avec la sécurité (des personnes et des données)
3.4 – Le concept suscite-t-il un imaginaire d'appropriation ?	<i>une girafe</i>	<i>je m'imagine faire [action]</i>	<i>je m'imagine, je me vois</i>	- Noms d'animaux (portrait chinois)
3.5 – Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?	<i>si c'était waterproof je pourrais les poser sur le lavabo</i>	<i>ça pourrait me servir aussi à [action] je le vois pas autrement que dans [contexte]</i>	<i>aussi à que à</i>	

4.1 – L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?	<i>ça va avec l'air du temps tout le monde fait ça ça existe pas déjà ? ça m'étonnerait qu'on vois ça un jour</i>		<i>digitalisation, connecté tendance actuellement sortir, arriver, lancement</i>	- Noms relatifs à l'air du temps / aux évolutions du présent (ex : <i>digital</i>)
4.2 – Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?	<i>ça perturbe les relations</i>	<i>je suis pas sûr que [personne / statut professionnel] accepte de [action]</i>	<i>adapter fournisseur, partenaire, client, sous-traitant famille ça va bloquer, ça va casser ça va cimenter avec mon [personne] communication</i>	
4.3 – Le concept est-il adapté à la place qu'occupe les professionnels dans la filière économique / aux relations familiales de l'utilisateur et à leur évolution ?	<i>ça correspond à mon rôle</i>			
4.4 – Le concept est-il adapté aux types d'organisation et à leur capacité d'évolution / aux manières de vivre de l'utilisateur et à leur évolution ?	<i>mais c'est n'importe quoi, ça va à contre-courant</i>		<i>dans mon entreprise, dans notre activité dans ma vie quotidienne synergie</i>	

4.5 – Quel prix l'utilisateur consent-il à payer pour ce concept ?	<i>il faut que ce soit abordable je ne paie pas plus de 100€ pour cet appareil-là</i>		<i>euros coûter cher accessible, abordable</i>	
--	---	--	--	--

TABLEAU EN ANNEXE 17 : Analyse des verbatim types

15. Ontologie

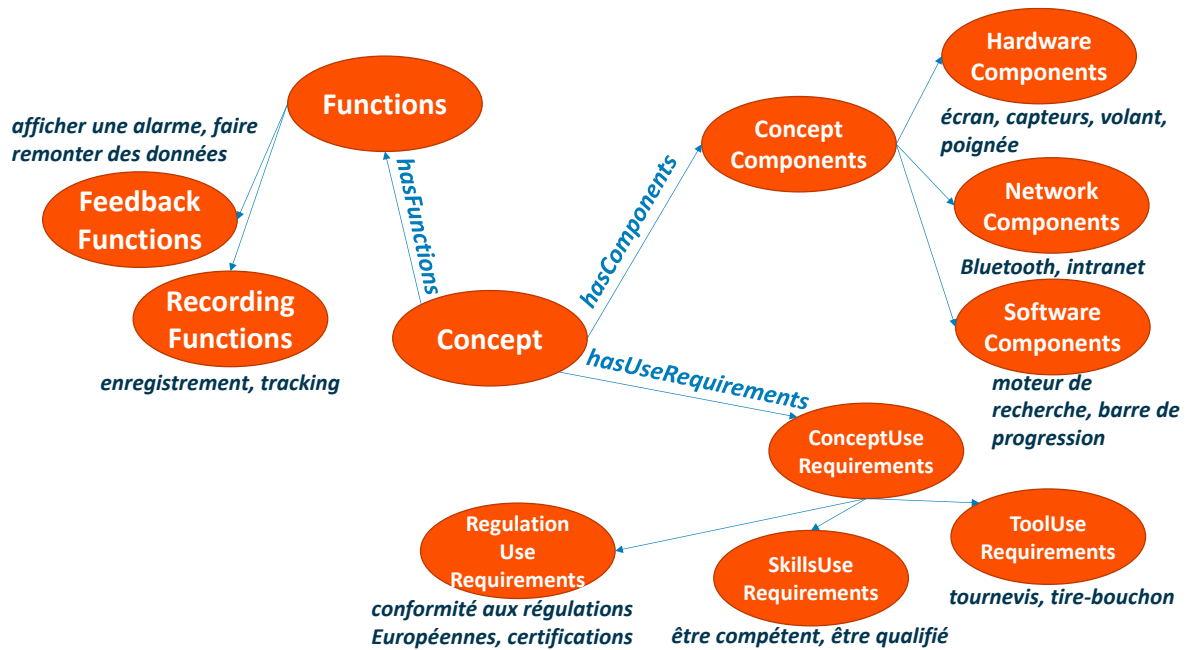


FIGURE EN ANNEXE 20 : Exemples de classes gravitant autour de la classe Concept

↑p173

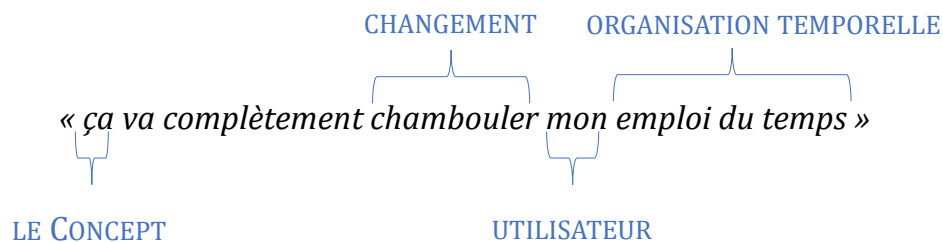
Considérons le verbatim stéréotypique :

« ça va complètement chambouler mon emploi du temps »

qui a été donné par les experts comme exemple pour le Critère :

2.5 – « L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ? »

Nous analysons premièrement les concepts que dénotent ce verbatim :



Nous faisons correspondre les concepts identifiés aux classes de l'ontologie existante :

LE CONCEPT —→ *Concept* (définition : Le concept qui est testé pendant l'entretien)
 UTILISATEUR —→ *User* (définition : L'interviewé / utilisateur potentiel du Concept)
 CHANGEMENT —→ *NeedOfChangePerception* (définition : Réaction de l'interviewé qui porte sur la nécessité d'un changement induit par le Concept)

Nous identifions le fait que le concept ORGANISATION TEMPORELLE n'est incarné par aucune des classes de notre ontologie. Nous créons alors une nouvelle que nous intégrons à notre ontologie :

ORGANISATION
TEMPORELLE —→ *UserTimeSchedule* (définition : Organisation temporelle/planning/emploi du temps de l'interviewé)

Nous cherchons, parmi les couples catégories lexico-sémantiques / classe ou propriétés dont nous disposons déjà, si certains peuvent être utilisés pour modéliser les différentes unités lexicales d'intérêt :

ça —→ [*C** → *Concept*]

chambouler —→ [*VbChanger* → *NeedOfChangePerception*]

mon —→ [*AdjPoss1erePersonne* → *User*]

* catégorie lexico-sémantique particulière qui regroupe les différentes façons potentielles de nommer le Concept

Nous ajoutons *chambouler*, rencontré ici pour la première fois, dans notre lexique.

Nous créons un nouveau couple pour la nouvelle classe :

emploi du temps —→ [*NomPlanning* → *UserTimeSchedule*]

et ajoutons cette nouvelle entrée dans le lexique.

Nous obtenons donc un patron de catégories lexico-sémantiques qui représente le verbatim :

[*C*]₍₁₎[*VbChanger*]₍₂₎[*AdjPoss1erePersonne*]₍₃₎[*NomPlanning*]₍₄₎

Nous cherchons ensuite à identifier l'ensemble de triplets de l'ontologie qui mobilisent les classes identifiées et qui représente au mieux le contenu du verbatim :

User₍₃₎*_hasReaction_**NeedOfChangePerception*₍₂₎
 + *NeedOfChangePerception*₍₂₎*_isAbout_**Concept*₍₁₎
 + *NeedOfChangePerception*₍₂₎*_isAbout_**UserTimeSchedule*₍₄₎

Nous décrivons cet ensemble de triplet par une phrase explication :

« L'utilisateur exprime sa perception d'une nécessité de changement induite par le concept et qui impacterait son organisation temporelle »

Nous obtenons ainsi la règle :

Patron	[C] [VbChanger] [AdjPoss1erePersonne] [NomPlanning]
Règle de classification par triplets	User_hasReaction_NeedOfChangePerception + NeedOfChangePerception_isAbout_Concept + NeedOfChangePerception_isAbout_UserTimeSchedule
Explication de la règle	L'utilisateur exprime sa perception d'une nécessité de changement induite par le concept et qui impacterait son organisation temporelle
Verbatim d'exemple	<i>ça va complètement chambouler mon emploi du temps</i>
Critère à attribuer	2.1

ENCADRÉ EN ANNEXE 1 : Exemple de création d'une règle à partir d'un verbatim

Exemples de sous-classes de *ReactionToConcept*

AdequacyPerception : réaction qui porte sur l'adéquation entre le Concept et un autre élément

BenefitsPerception : réaction au Concept qui consiste en l'identification d'un potentiel bénéfique apporté par le Concept

Comparison : réaction au Concept qui implique une comparaison entre le Concept et un autre élément

ConceptTargetPerception : réaction au Concept dans laquelle l'utilisateur s'exprime sur « A qui s'adresse le Concept ? »

DeletingPossibilityPerception : réaction au Concept qui implique que ce dernier offre la possibilité de supprimer quelque chose

EaseOfUsePerception : réaction au Concept qui porte sur sa facilité d'utilisation

Interest : réaction qui porte sur l'intérêt suscité chez l'interviewé par le Concept ou sa présentation

SafetyPerception : réaction au Concept qui porte la sécurité/le danger qui lui est lié

ENCADRÉ EN ANNEXE 2 : Exemples de sous-classes de *ReactionToConcept*

↑p179

Exemples de classes relatives aux caractéristiques du Concept

ConceptCompany : l'entreprise qui propose le Concept

ConceptManufacturingCountry : pays de fabrication du Concept

ConceptProofingParticularities : caractéristiques liées aux propriétés de résistance du Concept, par exemple, le fait d'être *waterproof* ou résistant à la foudre

ConceptUseContext : contexte d'utilisation du Concept comme « quand on prend l'avion »

ConceptUseRequirements : conditions préalables à l'utilisation du Concept, qui sont sous-divisées en conditions réglementaires et en conditions techniques, ces dernières étant divisées en conditions techniques matérielles et conditions techniques réseau

TargetMarket : le marché cible que le Concept adresse

Competiton : concurrence au Concept et l'entreprise qui le propose, divisée en entreprises concurrentes et produits/services concurrents

PotentialMarketReleaseSpecifics : spécificités de la potentielle arrivée sur le marché du Concept, comme sa date ou le fait que cette sortie sur le marché ait lieu ou non

ENCADRÉ EN ANNEXE 3 : Exemples de classes relatives aux caractéristiques du Concept

↑p182

Exemples de classes pouvant être ou non relatives au Concept

Functions : fonctions d'un objet technique / technologique, subdivisée en huit sous-classes telles que les fonctions de feedback ou les fonctions d'enregistrement

Costs : coûts, divisée en trois sous-classes pour les coûts faibles, moyens et élevés

Data : données manipulées par un système

ENCADRÉ EN ANNEXE 4 : Exemples de classes pouvant être ou non relatives au Concept

Exemples de classes relatives à l'utilisateur

Expectations : attentes de l'utilisateur antérieures au Concept

MaterialEnvironment : ensemble des éléments qui constituent l'environnement matériel de l'utilisateur, comme les bâtiments qu'il occupe ou les appareils qu'il utilise

Problems : problèmes rencontrés par l'utilisateur, subdivisés en sous-classes telles que *WasteOfTimeProblems*

UserAcquaintances : entourage de l'utilisateur, subdivisé entre l'entourage privé et professionnel

UserAcquaintancesActions : actions, tâches et procédures réalisées par les connaissances de l'utilisateur

UserExistingPractices : pratiques de l'utilisateur qui préexistent au Concept

UserIdentityComponents : les composants de l'identité de l'utilisateur, comme sa façon de s'habiller, son métier, son statut social, et ses valeurs éthiques et morales

UserKnowledge : les connaissances de l'utilisateur

UserLifePath : parcours de vie de l'utilisateur, dont une des sous-classes est son parcours professionnel

UserOrganization : organisation spatio-temporelle de l'utilisateur qui incluent par exemple son emploi du temps

UserSkills : compétences de l'utilisateur

UserTasksToAcheive : tâches que l'utilisateur doit accomplir, subdivisée en tâches autorisées, interdites, déplaisantes ou qui concernent un déplacement

UserMeansOfAcheivingTasks : moyens employés par l'utilisateur pour accomplir ses tâches

ENCADRÉ EN ANNEXE 5 : Exemples de classes relatives à l'utilisateur

Exemples de classes générales non intrinsèquement liées à l'utilisateur ou au Concept

Frequency : fréquences temporelles (par exemple, *tous les jours*), subdivisées en fréquences faibles et fréquences élevées

Jobs : métier ou corps de métier

Trends : modes, tendances

ENCADRÉ EN ANNEXE 6 : Exemples de classes générales non intrinsèquement liées à l'utilisateur ou au Concept

Exemples de classes liées à la présentation du Concept

PresentationContent : contenu de la présentation, subdivisé en trois classes, à savoir, les contenus qui sont des illustrations, ceux qui sont du texte et ceux qui sont des vidéos

PresentationParts : parties de la présentation du Concept, qui sont les sept questions de définitions du Concept, chacune étant reliée par la propriété *describes* à la sous-classe de *ConceptCharacteristics* correspondante (par exemple *HowDoesItWorkPart_describes_ConceptFunctionning*)

TextualMeaning : le sens d'un texte

ENCADRÉ EN ANNEXE 7 : Exemples de classes liées à la présentation du Concept

16. Règles de classification

Règles créées à partir de verbatim stéréotypiques				
Règles	Critère à attribuer	Phrase d'explication	Verbatim	Patron correspondant
Presentation_hasPresentationContent_TextualPresentationContent + Interviewer_showsPresentation_User + User_hasReaction_Understanding + TextualPresentationContent_hasTextualMeaning_TextualMeaning	1.1 - L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?	L'utilisateur s'exprime sur sa compréhension d'un élément textuel contenu dans la présentation du Concept qui lui montre l'interviewer.	<i>il y avait un acronyme dans votre présentation je comprends même pas ce qu'il veut dire quoi</i>	[NomTypeTexte] [PrepDans] [AdjPoss2emePersonne] [NomPrésentation] [PronPers1erePersonne] [VbComprendre] [PropSens]
User_hasReaction_Understanding + Understanding_isAbout_Concept + Understanding_isAbout_DataProcessingFunctions	1.2 - Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?	L'utilisateur s'exprime sur sa compréhension d'une fonction de traitement des données du Concept.	<i>par contre j'ai pas trop bien compris comment ça récupérait les données</i>	[PronPers1erePersonne] [VbComprendre] [AdvComment] [C] [GvAccèsData]
User_hasReaction_Comparison + Comparison_isAbout_CompetitionKnowledge + DifficultyToReactionModulation_modulatesReaction_Comparison	1.3 - L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?	L'utilisateur évoque sa difficulté à comparer le Concept à l'existant sur la base de ses connaissances de la concurrence.	<i>j'ai du mal à me raccrocher à des choses que je connais</i>	[PronPers1erePersonne] [VbAvoirDifficulté] [VbFaireComparaison] [GnChosesConnues]
User_hasReaction_Interest + Interest_isAbout_Functions	1.4 - L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?	L'utilisateur s'exprime sur son intérêt pour une ou plusieurs fonctions du Concept.	<i>moi ce qui m'intéresse c'est [une des fonctions du concept]</i>	[PronGénérique] [qui] [PronCOI1erePersonne] [VbIntéresser] [Cest] [NomFonction]
User_hasReaction_UseProjection + UseProjection_isAbout_Concept + UseProjection_isAbout_ConceptUseContext	1.5 - L'usage du concept est-il envisageable comme un système d'usage courant ?	L'utilisateur exprime une projection dans l'usage du Concept qui implique un contexte spécifique d'utilisation.	<i>je l'utiliserais quand je prendrai ma douche</i>	[PronPers1erePersonne] [PronCOD] [VbUtiliser] [quand] [PronPers1erePersonne] [GvActionDomestique]

User_hasReaction_EaseOfUsePerception + EaseOfUsePerception_isAbout_ Concept	2.1 - Le concept est-il simple à utiliser ?	L'utilisateur s'exprime sur sa perception de la facilité/difficulté d'utilisation du Concept.	<i>c'est une usine à gaz</i>	[Cest] [NomObjetComplexe]
User_hasReaction_ NewPossibilityPerception + NewPossibilityPerception_isAbout_ Concept + NewPossibilityPerception_isAbout_ UserTasksToAcheive	2.2 - Le concept compléterait-il sans les concurrencer les habitudes existantes ?	L'utilisateur s'exprime une perception de nouvelle possibilité portée par le Concept qui porte sur une des tâches qu'il doit accomplir.	<i>ça va me permettre de [faire quelque chose]</i>	[C] [PronCO11erePersonne] [VbPermettre] [GvAction]
User_hasReaction_ PreferencesExpression + PreferencesExpression_isAbout_ UserMeansOfAcheivingTasks	2.3 - La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?	L'utilisateur exprime une préférence qui porte sur un moyen dont il dispose pour réaliser une tâche.	<i>c'est mieux de faire sur téléphone</i>	[Cest] [AdvMieux] [de] [GvUtiliserMoyenDeCommunication]
User_hasReaction_ DeletingPossibilityPerception + DeletingPossibilityPerception_isAbout_ UserUnpleasantTasksToAcheive	2.4 - Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?	L'utilisateur exprime sa perception d'une possibilité de suppression qui porte sur une tâche désagréable qu'il doit accomplir.	<i>bah comme ça j'aurais plus à me priver de manger des frites tous les jours</i>	[PronPers1erePersonne] [VbPlusDevoir] [PronCOD1erePersonne] [GvPriverDeQuelqueChose]
User_hasReaction_ NeedOfChangePerception +NeedOfChangePerception_isAbout_ UserMotionTasksToAcheive +NeedOfChangePerception_isAbout_ UserTimeSchedule	2.5 - L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ?	L'utilisateur s'exprime sur une nécessité de changement qu'implique le Concept et qui porte sur ses déplacements et son organisation temporelle.	<i>il va falloir que je parte de chez moi un peu plus tôt le matin</i>	[VbFalloirBesoin] [que] [PronPers1erePersonne] [GvDéplacement] [AdvMoment]Journée]

ConceptTargetPerception_isAbout_Concept + ConceptTargetPerception_isAbout_TargetMarket	3.1 - L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?	L'utilisateur s'exprime sur sa perception de la cible du Concept en identifiant un métier.	<i>ça s'adresse plutôt à [nom métier]</i>	[C] [VbSAdresserA] [NomMétier]
User_hasReaction_ThreatPerception + ThreatPerception_isAbout_UserJob	3.2 - Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?	L'utilisateur exprime sa perception d'une menace portée par le Concept sur son métier.	<i>ils vont nous voler notre métier</i>	[PronPers3emePersonne] [VbAller] [PronCOD1erePersonne] [VbVolerDérober] [NomMétierUtilisateur]
DataProcessingFunctions_dealsWith_UserData + User_hasReaction_Question + Question_isAbout_DataProcessingFunctions	3.3 - Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?	L'utilisateur pose une question sur le traitement de ses données par le Concept.	<i>qu'est-ce que ça va faire de mes données ?</i>	[PronInterrogatif] [PronGénérique] [VbFaireDe] [AdjPoss1erePersonne] [NomData]
User_hasReaction_ImaginationEvocation + ImaginationEvocation_isAbout_UserTasksToAcheive	3.4 - Le concept suscite-t-il un imaginaire d'appropriation ?	L'utilisateur s'exprime sur le fait qu'il s'imagine utiliser le Concept pour une des tâches qu'il doit réaliser.	<i>je m'imagine faire [action]</i>	[PronPers1erePersonne] [PronCOD1erePersonne] [VbImaginer] [GvAction]
User_hasReaction_NewPossibilityPerception + NewPossibilityPerception_isAbout_Concept + ChangesConditionsReactionModulation_modulatesReaction_NewPossibilityPerception + ChangesConditionsReactionModulation_wantedChangesConcern_ConceptProofingParticularities	3.5 - Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?	L'utilisateur s'exprime sur nouvelle possibilité apportée par le Concept et qui est conditionnée par un changement qui devrait être effectué sur les propriétés de résistance du Concept.	<i>si c'était waterproof je pourrais les poser sur</i>	[si] [Cest] [AdjCaractéristiqueRésistance] [PronPers1erePersonne] [VbPouvoir] [PronCOD] [GVDéplacementObjet]

Question_isAbout_Concept + Question_isAbout_ PotentialMarketReleaseHappening	4.1 - L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?	L'utilisateur pose une question sur l'arrivée effective du Concept sur le marché.	<i>comment ça se fait que ce soit pas déjà sur le marché ?</i>	[AdvPourquoi] [Cest] [VbAvoirSurMarché]
InterferencePerception_isAbout_Concept + InterferencePerception_isAbout_ UserSocialInteractions	4.2 - Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?	L'utilisateur s'exprime sur sa perception d'une interférence portée par le Concept sur ses interactions sociales.	<i>ça générerait la communication</i>	[C] [VbGênerInterférer] [NomCommunication]
User_hasReaction_AdequacyPerception + AdequacyPerception_isAbout_Concept + AdequacyPerception_isAbout_ UserStatus	4.3 - Le concept est-il adapté à la place qu'occupent les professionnels dans la filière économique / aux relations familiales de l'utilisateur et à leur évolution ?	L'utilisateur s'exprime sur une perception d'adéquation entre le Concept et son propre statut.	<i>ça correspond à mon rôle</i>	[C] [VbCorrespondreRépondre] [AdjPoss1erePersonne] [NomPlacePosition]
AdequacyPerception_isAbout_Concept + AdequacyPerception_isAbout_Trends	4.4 - Le concept est-il adapté aux types d'organisation et à leur capacité d'évolution / aux manières de vivre de l'utilisateur et à leur évolution ?	L'utilisateur s'exprime sur une perception d'adéquation entre le Concept et les tendances.	<i>mais c'est n'importe quoi ça va à contre-courant</i>	[C] [VbAllerAContresens]
User_hasReaction_ValuePerception + GeneralConditionsReactionModulation_ modulatesReaction_ValuePerception + ValuePerception_isAbout_Concept + ValuePerception_isAbout_LowCosts	4.5 - Quel prix l'utilisateur consent-il à payer pour ce concept ?	L'utilisateur s'exprime sur sa perception d'une valeur pour le Concept qui est conditionnée par un coût bas.	<i>Il faut que ce soit abordable</i>	[VbFalloirBesoin] [que] [Cest] [AdjAbordableFinancière] [ment]

TABLEAU EN ANNEXE 18 : Exemples de règles construites sur des verbatim stéréotypiques

Règles créées à partir de verbatim réels				
Règles	Critère à attribuer	Phrase d'explication	Verbatim	Patron correspondant
User_hasReaction_Opinion + User_hasReaction_ ProgressPerception + Opinion_isAbout_ Concept + ProgressPerception_isAbout_ Concept	1.1 - L'utilisateur comprend-il le concept et le trouve-t-il intéressant a priori ?	L'utilisateur donne une opinion générale et s'exprime sur sa perception du progrès porté par le Concept.	<i>Moi, ce qui me plaît, c'est la nouveauté.</i>	[Pron1erePersonne] [VbPlaire] [Cest] [NomChoseNouvelle]
User_hasReaction_Question + Question_isAbout_ MaterialTechnicalUseRequirements	1.2 - Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ?	L'utilisateur pose une question sur les condition matérielles prérequis pour l'utilisation du Concept.	<i>Est-ce qu'il faudra refaire toute l'installation électrique ?</i>	[PronInterrogatif] [VbFalloirBesoin] [GvAdaptationMatérielle]
Concept_isSimilarTo_ CompetitiveProductsServices + NonTargetMarket_ isOccupiedByCompetition_ CompetitiveProductsServices	1.3 - L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ?	L'utilisateur indique qu'un marché différent de celui visé par le Concept est occupé par un produit ou service qui lui est similaire.	<i>Ça existe sur des gros parcs Hlm</i>	[C] [GvExister] [AdvNomTypeSites]
User_hasReaction_Opinion + Opinion_isAbout_ ConceptComponents	1.4 - L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ?	L'utilisateur exprime une opinion générale sur un composant du Concept.	<i>moi j' ai aimé les prises rondes</i>	[PronPers1erePersonne] [VbApprécier] [NomComposant]
User_hasReaction_UseProjection + UserProjection_isAbout_ UserTasksToAcheive	1.5 - L'usage du concept est-il envisageable comme un système d'usage courant ?	L'utilisateur exprime une projection dans l'utilisation du Concept pour une de ses tâches à accomplir.	<i>Ce n'est pas possible ça, on ne peut pas faire de l'éclairage avec ça, en tout cas.</i>	[VbPouvoir] [GvAction] [avec] [C]
User_hasReaction_Comparison + Comparison_isAbout_ ToolsUseRequirements	2.1 - Le concept est-il simple à utiliser ?	L'utilisateur fait une comparaison qui porte sur un prérequis d'utilisation concernant des outils.	<i>Il n'y a plus besoin de tournevis ou de pince</i>	[VbPlusFalloirBesoin] [NomOutils]

User_hasReaction_DeletingPossibilityPerception + DeletingPossibilityPreception_isAbout_Costs	2.2 - Le concept compléterait-il sans les concurrencer les habitudes existantes ?	L'utilisateur exprime une perception de possibilité de suppression par le Concept qui concerne les coûts.	<i>si ça permet d'éviter de payer des choses</i>	[si][C][VbPermettre][VbEviter][VbPayer]
User_hasReaction_Comparison + Comparison_isAbout_HardwareComponents + User_hasReaction_ThreatPerception + ThreatPerception_isAbout_HardwareComponents	2.3 - La comparaison avec les pratiques existantes valide-t-elle et rend-elle crédibles les pratiques nouvelles proposées ?	L'utilisateur fait une comparaison dans laquelle il identifie une menace portée par le Concept et qui concerne les composants <i>hardware</i> .	<i>Actuellement il faut une vis mais là les clips plastiques ça va pêter</i>	[AdvActuellement][VbFalloirBesoin][NomOutils][AdvConjMais][AdvLàMaintenant][NomComposant][PronGénérique][VbEndommager]
User_hasProblems_FaultyWorkProblems + FaultyWorkProblems_hasFrequency_LowFrequency	2.4 - Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ?	Le participant s'exprime sur la fréquence basse d'un problème que le Concept pourrait résoudre.	<i>j' ai rarement vu des lampes en panne</i>	[PronPers1erePersonne][VbAvoir][AdvFréquenceFaible][VbVoir][NomChoseEnPanne]
User_hasReaction_AdequacyPerception + AdequacyPerception_isAbout_UserTimeSchedule	2.5 - L'utilisateur construit-il une organisation personnelle de ses pratiques nouvelles avec le concept ?	Le participant s'exprime sur une adéquation entre le Concept et son organisation temporelle.	<i>je n' aurais ni le temps ni les moyens d'exploiter.</i>	[PronPers1erePersonne][VbAvoirVbAvoirRencontrer][NomTemps][VbExploiter]
User_hasReaction_ConceptTargetPerception + ConceptTargetPreception_isAbout_TargetMarketPersonalProfile	3.1 - L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ?	Le participant exprime une perception de cible du Concept qui concerne un type de personne.	<i>Je disais avant que c'était une solution que moi je verrais pour l'aménagement d'une cuisine pour une personne seule.</i>	[C][que][Pron1erePersonne][VbVoir][PrepPour][GnAction][PrepPour][GnTypePersonne]

User_hasReaction_ImprovementPossibility Perception + ImprovementPossibility Perception_isAbout_Recognition + ImprovementPossibility Perception_isAbout_UserCoworkers	3.2 - Avec le concept, l'utilisateur peut-il agir/jouer personnellement sur son rôle ?	L'utilisation exprime sa perception d'une amélioration portée par le Concept qui concerne la notion de reconnaissance par les pairs et ses collègues.	<i>ce serait gratifiant alors pour le... pour la personne à l'atelier, hein.</i>	[Cest] [AdjGratifiant] [PrepPour] [NomMétier]
User_hasReaction_ThreatPerception + ThreatPerception_isAbout_ConceptIndirectConsequences	3.3 - Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ?	Le participant s'exprime sur une perception de menace induite par une conséquence indirecte du Concept.	<i>Le côté espionnage aussi.</i>	[NomCaractéristiqueDéranganteValeurs]
User_hasReaction_ImaginationEvocation	3.4 - Le concept suscite-t-il un imaginaire d'appropriation ?	Le participant s'exprime sur une évocation suscitée par le concept dans son imagination.	<i>la première image qui me vient, c'est l'histoire du caméléon</i>	[GvImaginerEvoquer] [Cest] [NomAnimal]
User_hasReaction_UseProjection + UserProjection_isAbout_UserTasksToAcheive + NewIdeaReactionModulation_modulatesReaction_UseProjection	3.5 - Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ?	Le participant se projette dans l'usage du Concept pour une des tâches qu'il doit réaliser en imaginant une nouvelle idée.	<i>Il y a aussi toute sorte d'interventions de l'appareil urinaire qui seraient susceptibles d'être faites avec ce système.</i>	[aussi] [NomAction] [qui] [Copule] [AdjPossible] [GvFaireRéaliser] [avec] [C]
User_hasReaction_AdequacyPerception + AdequacyPerception_isAbout_Concept + AdequacyPerception_isAbout_Trends	4.1 - L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ?	Le participant s'exprime sur une perception d'adéquation entre le Concept et les tendances.	<i>il est tout à fait en adéquation avec l'évolution de la chirurgie robotique qu'on va avoir dans les années qui viennent, c'est clair</i>	[C] [VbEtreEnAdéquation] [NomEvolution]

User_hasReaction_Skepticism + Skepticism_isAbout_ UserAcquaintancesActions + Skepticism_isAbout_ UserExternalCoworker	4.2 - Le concept est-il adapté aux relations clients-fournisseurs des professionnels / aux formes de sociabilité de l'utilisateur et à leurs évolutions ?	L'utilisateur exprime une réaction de scepticisme qui concerne les actions de ses connaissances et ses collaborateurs externes.	<i>J'ai du mal à voir l'intervention d'un diagnostiqueur externe.</i>	[GvDifficultéAImaginer] [NomAction] [NomMétierExterne]
User_hasReaction_ObligationPerception + ObligationPerception_isAbout_SkillUseRequirements + ObligationPerception_isAbout_UserCoworker	4.3 - Le concept est-il adapté à la place qu'occupent les professionnels dans la filière économique / aux relations familiales de l'utilisateur et à leur évolution ?	L'utilisateur exprime une perception d'obligation portée par le Concept qui concerne des connaissances prérequis à l'utilisation et ses collègues.	<i>Donc la compétence, elle doit être... elle doit être chez l'automaticien ou au Bureau d'Études.</i>	[NomCompétences] [VbDevoirYAvoir] [chez] [NomMétier]
User_hasReaction_AdequacyPerception + AdequacyPerception_isAbout_Concept + AdequacyPerception_isAbout_TargetMarket + User_hasReaction_EaseOfUsePerception + GeneralConditionModulation_modulatesReaction_AdequacyPerception + GeneralConditionModulation_modulatesReaction_EaseOfUsePerception	4.4 - Le concept est-il adapté aux types d'organisation et à leur capacité d'évolution / aux manières de vivre de l'utilisateur et à leur évolution ?	L'utilisation exprime une perception de facilité d'utilisation du Concept et une adéquation entre le Concept et un son marché cible, avec des conditions.	<i>Il faut quand même que ce soit une solution assez ouverte pour que ce soit facilement utilisable par l'industriel.</i>	[VbFalloirBesoin] [Copule] [C] [GnAspectConcept] [PrepPour] [AdjIntuitif] [par] [AdvNomDomaine]
User_hasReaction_Question + Question_isAbout_Costs	4.5 - Quel prix l'utilisateur consent-il à payer pour ce concept ?	L'utilisateur pose une question sur le prix du Concept.	<i>Le prix, on le sait : c'est déjà 400 ou 500 € ?!</i>	[NomSomme] [Cest] [déjà] [NOM] [NomMonnaie] [INTER]

TABLEAU EN ANNEXE 19 : Exemples de règles construites sur des verbatim réels

17. Augmentation du lexique

Catégorie	Déclencheur	Phrase de désambiguïsation	Mono. WoNeF	Poly. WoNeF
AdjAgréable	agréable	<i>C'est super fun votre truc !</i>	plaisant, confortable, concordant, doux, chaleureux, engageant, gai, amusant, excellent, sympathique	qui fait plaisir
AdjStupide	stupide	<i>C'est un peu stupide de ne pas mettre un suivi GPS là-dedans.</i>	bête, gauche, idiot, maladroit, sot, inintelligent, délirant, zinzin, barjot, fada, allumé, imbécile, grossier, niais, insensé	sans intelligence, dépourvu d'intelligence
AdvActuellement	actuellement	<i>Ce n'est pas ce que l'on fait actuellement, mais pourquoi pas.</i>	maintenant, aujourd'hui, présentement	en ce moment, à présent, ces jours-ci, de nos jours
AdvRapidité	rapidement	<i>Ça va me permettre de faire les vérifications plus rapidement.</i>	promptement, vélocement, vite, diligemment, fissa	à tire-d'aile
VbEvoquer	évoquer	<i>Ça m'évoque plutôt une idée de détente.</i>	rappeler, imaginer, souvenir_se, rappeler_se, remémorer_se, susciter, provoquer, renvoyer	faire apparaître
VbCraindre	craindre	<i>Je crains de l'abimer vu comme il a l'air fragile.</i>	redouter, appréhender	
NomCible	cible	<i>Je ne pense pas être la cible de votre concept.</i>		
NomCompétences	compétence	<i>Ça ne fait pas partie de mes compétences.</i>	capacité, aptitude, savoir-faire, expertise, maîtrise	sphère d'attributions
AdjFacile	facile	<i>Je pense que ça sera assez facile à utiliser, à première vue.</i>	aisé, simple, élémentaire, simplet	peu exigeant, peu astreignant
AdjCoûteux	cher	<i>C'est beaucoup trop cher pour nous.</i>	cher, dispendieux, chérot, onéreux	qui coûte cher

AdvDavantage	davantage	<i>Il faut ajouter davantage de fonctions.</i>		de plus en plus
AdvDifféremment	différemment	<i>Ça me permettrait de faire les choses différemment.</i>	autrement	d'une manière différente
VbAbandonner	abandonner	<i>Je vais pouvoir abandonner mes cahiers.</i>	lâcher, délaisser, détacher_se, éliminer, jeter, larguer, écarter, défausser_se, enterrer, renoncer,	renoncer à, mettre de côté, défausser_se de, débarrasser_se de, défaire_se de, faire abstraction de, départir_se de,
VbCimenterConsolider	consolider	<i>Ça va consolider les relations.</i>	consolider_se, affermir, maintenir,	

TABLEAU EN ANNEXE 20 : UL extraites à partir de WoNef

Catégorie	Déclencheur	Phrase de désambiguïisation	Mono. WOLF	Poly. Wolf
AdjAgréable	agréable	<i>C'est super fun votre truc !</i>	plaisant, drôle,	
AdjStupide	stupide	<i>C'est un peu stupide de ne pas mettre un suivi GPS là-dedans.</i>	idiot, absurde, imbécile, insensé, aberrant, ridicule, niais,	
AdvActuellement	actuellement	<i>Ce n'est pas ce que l'on fait actuellement, mais pourquoi pas.</i>	aujourd'hui, maintenant, présentement, désormais,	de nos jours, de notre temps,
AdvRapidité	rapidement	<i>Ça va me permettre de faire les vérifications plus rapidement.</i>	diligemment, vite, illico, promptement, hâtivement,	à la hâte, à la va-vite,
VbEvoquer	évoquer	<i>Ça m'évoque plutôt une idée de détente.</i>	suggérer, provoquer, inspirer, susciter,	
VbCraindre	craindre	<i>Je crains de l'abimer vu comme il a l'air fragile.</i>	redouter, inquiéter, préoccuper,	
NomCible	cible	<i>Je ne pense pas être la cible de votre concept.</i>		
NomCompétences	compétence	<i>Ça ne fait pas partie de mes compétences.</i>	habileté, expertise, ressort, savoir-faire,	
AdjFacile	facile	<i>Je pense que ça sera assez facile à utiliser, à première vue.</i>	primaire, simple, élémentaire,	
AdjCoûteux	cher	<i>C'est beaucoup trop cher pour nous.</i>	cher, dispendieux, chérot,	
AdvDavantage	davantage	<i>Il faut ajouter davantage de fonctions.</i>		
AdvDifféremment	différemment	<i>Ça me permettrait de faire les choses différemment.</i>	autrement,	
VbAbandonner	abandonner	<i>Je vais pouvoir abandonner mes cahiers.</i>	délaisser, dégager, renoncer, laisser,	
VbCimenterConsolider	consolider	<i>Ça va consolider les relations.</i>	renforcer,	

TABLEAU EN ANNEXE 21 : UL extraites à partir de WOLF

Catégorie	Déclencheur	Phrase de désambiguïsation	Mono. JeuxDeMots	Poly. JeuxDeMots
AdjAgréable	agréable	<i>C'est super fun votre truc !</i>	gai, plaisant, charmant, favorable, sympathique, attirant, accommodant, confortable, accueillant, vivable, réjouissant, bien, chouette, commode, bon, enchanteur, doux, engageant, amusant, convenable, captivant, léger, joyeux, fameux, sympa, récréatif,	pas chiant,
AdjStupide	stupide	<i>C'est un peu stupide de ne pas mettre un suivi GPS là-dedans.</i>	absurde, aberrant, balourd, bête, couillon, crétin, déraisonnable, idiot, imbécile, incohérent, inepte, inintelligent, insensé, nigaud, sot, niais, nouille, débile, saugrenu, illogique, benêt,	
AdvActuellement	actuellement	<i>Ce n'est pas ce que l'on fait actuellement, mais pourquoi pas.</i>	aujourd'hui, présentement, maintenant, præsentement,	de nos jour, en ce moment, à présent, à l' époque actuel, pour l' instant, par le temps qui courent, à l' heure qu' il est, à l' heure actuel, à notre époque,
AdvRapidité	rapidement	<i>Ça va me permettre de faire les vérifications plus rapidement.</i>	précipitamment, prestement, vite, dare-dare, hâtivement, presto, promptement, fissa, diligemment, daredare, rapidos, vitement, expéditivement, viteuf,	à bride abattre, aussi sec, illico presto, à grand train, en moins de deux, à grand vitesse, en vitesse, en deux temps trois mouvement, en un tournemain, à fond de train, à fond le ballon, à le hâte, à le va-vite, à tout vitesse, à grand pas, en hâte, entre deux porte, à tout allure, sur le pouce, d' un seul jet, en un instant, en un moment, en un tour de main, à tire-d'aile, très vite, de façon rapide, entre deux porte, à fond le caisse, d' un seul coup, à l'emporte-pièce, à vitesse grand v, en grand vitesse, en moins de temps qu' il ne faut pour le dire, d' un coup, en deux coup de cuillère à pot,

VbEvoquer	évoquer	<i>Ça m'évoque plutôt une idée de détente.</i>	rappeler, suggérer, apparaître, éveiller, rêver, imaginer, songer, retracer, remémorer, figurer_se, susciter, repasser, souvenir_se, remémorer_se,	
VbCraindre	craindre	<i>Je crains de l'abimer vu comme il a l'air fragile.</i>	appréhender, redouter, trembler, alarmer_se, pétocher, méfier_se,	avoir peur, être effrayé, être épouvanté, avoir le trouille, avoir le chocotte,
NomCible	cible	<i>Je ne pense pas être la cible de votre concept.</i>		
NomCompétences	compétence	<i>Ça ne fait pas partie de mes compétences.</i>	aptitude, capacité, ressort, savoir-faire, bagage, qualification, expérience,	
AdjFacile	facile	<i>Je pense que ça sera assez facile à utiliser, à première vue.</i>	accessible, limpide, accommodant, simple, aisé, élémentaire, enfantin, intelligible, évident, bateau,	bête comme chou,
AdjCoûteux	cher	<i>C'est beaucoup trop cher pour nous.</i>	cher, onéreux, ruineux, inabordable, dispendieux,	hors de prix,
AdvDavantage	davantage	<i>Il faut ajouter davantage de fonctions.</i>	encore, plus,	de plus beau,
AdvDifféremment	différemment	<i>Ça me permettrait de faire les choses différemment.</i>	autrement, diversement,	
VbAbandonner	abandonner	<i>Je vais pouvoir abandonner mes cahiers.</i>	débarrasser_se, lâcher, défaire_se, céder, laisser, départir_se, larguer, négliger, délaissé, enterrer, renoncer, détacher_se, jeter, séparer_se, balancer, supprimer, désister_se, abolir,	laisser tomber, faire cesser, tirer un trait, renoncer à,
VbCimenterConsolider	consolider	<i>Ça va consolider les relations.</i>	soutenir, solidifier, affermir, renforcer, ancrer, fortifier, cimenter, stabiliser, fixer, assurer, resserrer, enraciner, enforcir,	

18. Expérimentation en conditions réelles



FIGURE EN ANNEXE 21 : Concept testé dans le cadre de l'étude de l'expérimentation en conditions réelles

Q1	Nous allons vous présenter la solution DOLOSPORT au travers d'une vidéo (que vous pourrez relancer si besoin). Merci de lire attentivement les informations contenues. Une fois la vidéo visionnée, décrivez en quelques lignes ce que vous avez spontanément pensé de la solution et sélectionnez les émoticônes qui représentent le plus ce que vous avez ressenti à la découverte de celle-ci.
Q2	Qu'est-ce qui vous a semblé facile ou difficile à comprendre ? / Quelles informations supplémentaires auriez-vous souhaité avoir ?
Q3	Dans quelles situations utiliseriez-vous cette solution et comment (à quels moments, dans quels objectifs, ...) ?
Q4	Ce concept résout-il des problèmes que vous rencontrez aujourd'hui ? Si oui, lesquels ?
Q5	Parmi ces trois fonctions de la solution : 1) Conseil d'exercices adaptés à vos douleurs, 2) Conseils pour corriger la posture durant l'exercice et 3) Remise sur les mutuelles de santé, laquelle ou lesquelles vous intéressent le plus et pourquoi ?
Q6	Si vous deviez associer trois adjectifs à cette solution, lesquels choisiriez-vous ? Pourquoi ?
Q7	En quoi la solution DOLOSPORT est-elle en phase ou non avec les tendances actuelles ?
Q8	Quel est le prix minimum en-dessous duquel la solution ne vous semblerait pas crédible et quel est le prix maximum au-dessus duquel la solution vous semblerait trop chère ? Merci d'indiquer une fourchette de prix.


TABLEAU EN ANNEXE 23 : Questions posées aux participants lors de l'étude test

1. Nous allons vous présenter la solution DOLOSPORT au travers d'une vidéo (que vous pourrez relancer si besoin). Merci de lire attentivement les informations contenues.

Une fois la vidéo visionnée, décrivez en quelques lignes ce que vous avez spontanément pensé de la solution et sélectionnez les émoticônes qui représentent le plus ce que vous avez ressenti à la découverte de celle-ci.

[Participant1] 

Je trouve cela très innovateur. Là où les vêtements de sport connectés actuels sont axés sur la performance des sportifs, ici, on aura l'occasion pour des amateurs de se rassurer que les mouvements que nous effectuons sont justes. Cela nous évitera bien des mauvaises surprises (douleurs et autres...) au lendemain d'une activité.

[Participant2] 

Très bonne idée mais je pense que le nom du projet ne devrait pas comporter la notion de douleur (dolo)

[Participant3] 

Le vêtement guide ... Il est réalisé sur mesure ? (je ne pense pas) Il faut un mobile pendant les exercices ? (a priori oui) L'argument de réduction du coût de la mutuelle ... (illusoire)

[Animatrice1]


Pour quelle.s raison.s trouvez-vous cet argument illusoire ?

[Participant3]

il me semble qu'il faille principalement cibler la communication sur le produit lui-même, cet argument de la mutuelle me paraît très secondaire, et probablement pas de nature à motiver un achat. Avant que le vêtement soit reconnu comme ayant des vertus thérapeutiques, et donc soit susceptible d'être remboursé (en partie) par l'assurance maladie et les mutuelles, il faudrait se focaliser d'abord sur ses vertus intrinsèques, ses bienfaits, sa fiabilité, ergonomie, etc ...

[Participant5]  2★

Cette solution s'appuie sur une approche "objet connecté", clairement dans l'air du temps. Pour autant, je reste circonspect sur une telle technologie. Il est vrai qu'elle peut apporter des informations intéressantes sur notre posture et les origines de nos douleurs. Mais, elle ne remplacera jamais le conseil de professionnel et de coach sportif. L'utilité d'une telle solution ne me paraît pas acquise pour le moment.

[Participant6] 

Grand doute sur la spécificité du sensor de posture, est-ce que cela est adaptée aussi aux gens qui ont des malformations (e.g. scoliose)? Est-ce que ceci décharge rapidement le téléphone? Comment ça 'bénéficie' pour la mutuelle? Petit remarque sur la présentation: la texte va assez rapide et c'est donc difficile de tout lire.

[Participant7] 

Solution simple, à priori facile d'utilisation. Peut soulager les douleurs et corriger les postures. Plutôt sceptique par rapport au lien entre utilisation du produit & remise sur mutuelle

[Participant8]

je suis tout à fait d'accord avec cette remarque sur la remise mutuelle

[Participant9] 

Solution pratique pour savoir si nos postures lors des mouvements sont bien adaptés car c'est toujours une question que l'on se pose au moment d'un exercice non accompagné de quelqu'un d'expérimenter. Mais l'individualisation de tel outil me fait me poser des question sur l'impact sur notre environnement avec les ressources naturelles que cela nécessite pour être fabriqué.

ENCADRÉ EN ANNEXE 8 : Extraits de posts collectés pour l'étude test

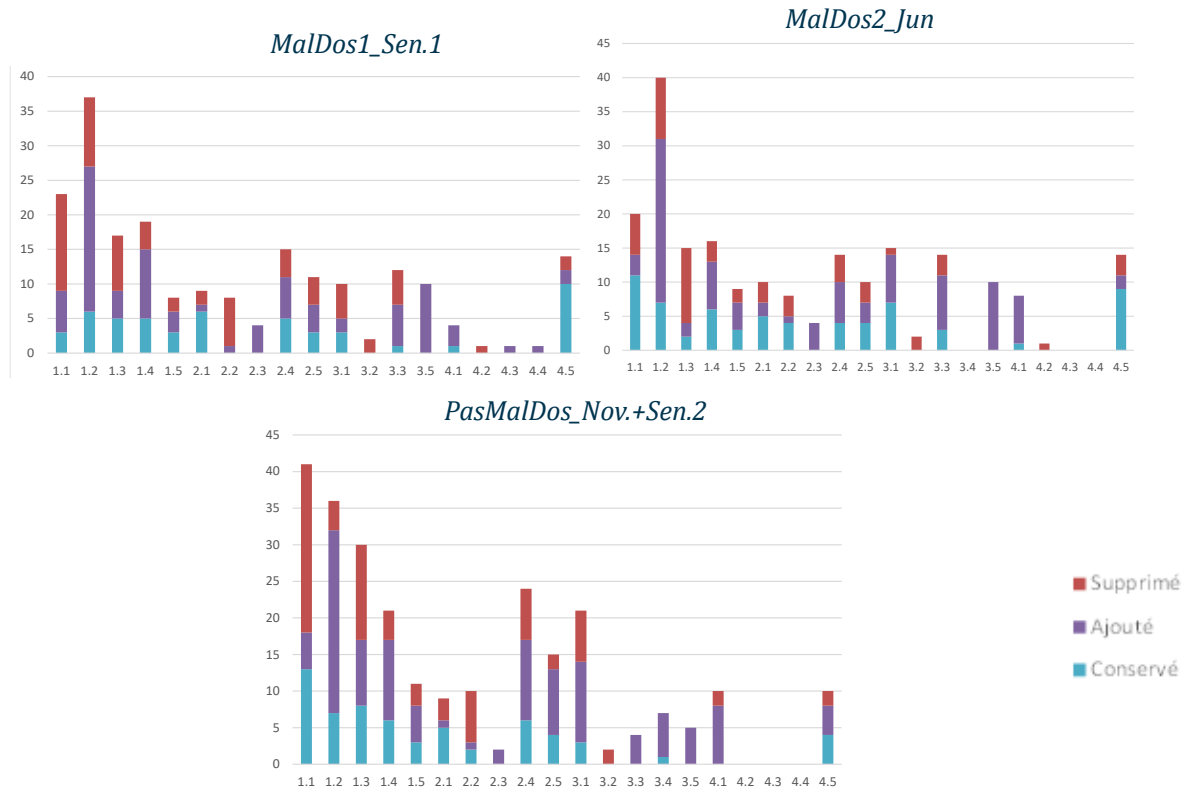


FIGURE EN ANNEXE 22 : Corrections par Critère par textes

Exemples de classifications réussies			
Verbatim	Critère	Règle	Réussite description
<i>(je ne pense pas) Il faut un mobile pendant les exercices ?</i>	1.2 – « Le principe technique du concept est-il suffisamment compréhensible par l'utilisateur ? »	<i>L'utilisateur pose une question sur les prérequis matériels nécessaires pour utiliser le concept.</i> Question_isAbout_Concept + Question_isAbout_MaterialTechnicalUseRequirements	oui
<i>Cette solution s'appuie sur une approche " objet connecté ", clairement dans l'air du temps.</i>	4.1 – « L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ? »	<i>L'utilisateur s'exprime quant à l'adéquation qu'il perçoit entre le concept et les tendances.</i> AdequacyPerception_isAbout_Concept + AdequacyPerception_isAbout_Trends	oui
<i>Il est vrai qu'elle peut apporter des informations intéressantes sur notre posture et les origines de nos douleurs.</i>	1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? »	<i>L'utilisateur exprime son intérêt pour une fonction de feedback du concept.</i> User_hasReaction_Interest + Interest_isAbout_FeedbackFunctions	oui
<i>Comme déjà évoqué dans mes précédentes réponses, je pense que ce type de produits peut répondre à des problèmes liés au manque de motivation à faire du sport, ou tout simplement à avoir moins de douleurs au dos à certaines périodes.</i>	2.4 – « Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ? »	<i>L'utilisateur exprime le fait que le concept apporte ou non une réponse à un de ses problèmes.</i> Concept_answersProblems_Problems	oui
<i>Il me paraît difficile de répondre affirmativement à cette question sans avoir essayé d'abord la solution.</i>	2.1 – « Le concept est-il simple à utiliser ? »	<i>L'utilisateur s'exprime sur la facilité/difficulté d'utilisation du concept.</i> User_hasReaction_EaseOfUsePerception + EaseOfUsePerception_isAbout_Concept	non
<i>Enfin, en termes de coût d'équipement et d'application, on n'y voit aucune information.</i>	4.5 – « Quel prix l'utilisateur consent-il à payer pour ce concept ? »	<i>L'utilisateur donne un prix pour le concept.</i> [Règle d'attribution par mot clef]	non

<i>J'utiliserais cette solution afin de me motiver à faire des activités sportives au quotidien, ou à des moments où la fatigue accentue mes douleurs musculaires.</i>	1.5 – « L'usage du concept est-il envisageable comme un système d'usage courant ? »	<i>L'utilisateur se projette dans l'utilisation du concept pour accomplir une tâche.</i>	oui/non
		User_hasReaction_UseProjection + UseProjection_isAbout_UserTasksToAcheive	
<i>Ainsi, il serait intéressant que l'application puisse quand même donner des conseils et astuces personnalisés sur nos postures même en dehors des activités sportives avec par exemple des conseils pour les personnes qui travaillent devant un écran toute la journée, ou encore pour ceux qui sont souvent en déplacement en voiture...</i>	3.1– « L'utilisateur comprend-il que le concept le concerne et sait-il à qui d'autre il peut s'adresser ? » + 1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? »	<i>L'utilisateur s'identifie ou non lui-même comme la cible du concept.</i>	oui
		User_hasReaction_ConceptTargetPerception +ConceptTargetPerception_isAbout_Concept	
		<i>L'utilisateur exprime son intérêt pour une fonction de feedback du concept.</i>	
		User_hasReaction_Interest + Interest_isAbout_FeedbackFunctions	

TABLEAU EN ANNEXE 24 : Exemples de classifications réussies

Exemples de classifications erronées				
Verbatim	Classification algorithme	Règle	Classification analyste	Commentaire
<i>Éventuellement, à quelles solutions existantes ce concept vous fait-il penser ?</i>	1.3 – « L'utilisateur peut-il raccrocher les techniques mobilisées par le concept à des techniques existantes ? »	<i>L'utilisateur indique qu'un produit ou service concurrent est similaire au concept.</i> CompetitiveProductsServices_isSimilarTo_Concept	/	Question de l'animatrice
<i>Elle est aussi en phase avec la tendance qui est de faire plus attention à soi et de faire plus d'activités sportives (notamment avec le covid et les confinements).</i>	/	/	4.1 – « L'utilisateur trouve-t-il plausible l'arrivée du concept sur le marché tel qu'il le perçoit ? »	<i>en phase avec pas dans le lexique</i>
<i>Je suis intéressée par les deux premières</i>	/	/	1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? »	Réponse à une question (anaphore non résolue)
<i>Un signal sonore quand la position n'est pas bonne ?</i>	/	/	3.5 – « Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ? »	Manque de contexte
<i>Encore faudrait il pouvoir le porter pendant l'exercice physique.</i>	2.2 – « Le concept compléterait-il sans les concurrencer les habitudes existantes ? »	<i>L'utilisateur s'exprime sur la perception d'une nouvelle possibilité portée par le concept dans la réalisation de ses tâches.</i> User_hasReaction_NewPossibilityPerception + NewPossibilityPerception_isAbout_UserTasksToAcheive	2.4 – « Le concept résout-il des problèmes gênants rencontrés par l'utilisateur ? »	Contresens dû à la non prise en compte de <i>encore faudrait il.</i>

<p><i>une fonction permettant de mesurer son rythme cardiaque pendant la pratique sportive serait bien je pense</i></p>	<p>1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? »</p>	<p><i>L'utilisateur exprime une opinion positive ou négative générale à propos d'une fonction du concept.</i></p> <p>User_hasReaction_Opinion + Opinion_isAbout_Functions</p>	<p>3.5 – « Le concept laisse-t-il la possibilité d'imaginer des extensions d'usage ? »</p>	<p>Expression du souhait pas détectée</p>
<p><i>Pour la troisième solution, hors de question de me faire flicker par ma mutuelle. Et finalement, est-ce que ça ne sera pas ceux qui sont le plus cassé, qui finiront par payer le plus ?</i></p>	<p>/</p>	<p>/</p>	<p>1.4 – « L'utilisateur peut-il trouver et choisir les fonctions qui l'intéressent dans le concept ? » + 3.3 – « Le concept est-il compatible avec les valeurs culturelles de l'utilisateur ? »</p>	<p>Réduction sur la mutuelle non identifiée comme fonction + <i>se faire flicker</i> pas dans le lexique comme conséquence indirecte négative du Concept</p>

TABLEAU EN ANNEXE 25 : Exemples de classifications erronées