



HAL
open science

Conception d'un système prédictif pour évaluer l'impact d'un dispositif numérique pour l'orientation des lycéens

Marie Gribouval

► To cite this version:

Marie Gribouval. Conception d'un système prédictif pour évaluer l'impact d'un dispositif numérique pour l'orientation des lycéens. Informatique [cs]. Université de Lorraine, 2023. Français. NNT : 2023LORR0319 . tel-04583736

HAL Id: tel-04583736

<https://hal.univ-lorraine.fr/tel-04583736>

Submitted on 22 May 2024

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.



**UNIVERSITÉ
DE LORRAINE**

**BIBLIOTHÈQUES
UNIVERSITAIRES**

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Mention : Informatique

par **Marie GRIBOUVAL**

Sous la direction de Davy MONTICOLO et Éric BONJOUR

**Conception d'un système prédictif pour évaluer
l'impact d'un dispositif numérique pour l'orientation
des lycéens**

Soutenue le 22 décembre 2023

Membres du jury :

Directeur de thèse :	M. Davy MONTICOLO	Professeur des Universités, Université de Lorraine, Nancy
Co-Directeur de thèse :	M. Éric BONJOUR	Professeur des Universités, Université de Lorraine, Nancy
Rapporteurs :	Mme Virginie GOEPP	Professeure des Universités, INSA Strasbourg
	Mme Nada MATTA	Professeure des Universités, Université de Technologie de Troyes
Examineurs :	M. Jean-Philippe KOTOWICZ	Maître de conférences, INSA Rouen Normandie
Président de jury :	M. Dominique PETITJEAN	Professeur des Universités, Université de Lorraine, Nancy

*A Solal, Romane, Lou, et Félix,
que vous puissiez devenir les adultes que vous souhaitez être.*

REMERCIEMENTS

Je tiens tout d'abord à remercier vivement Davy, pour la bienveillance avec laquelle il a accompagné ces travaux, ainsi que pour la confiance inaltérable qu'il a manifestée à mon égard durant ces trois années. Merci de m'avoir proposé ce sujet de thèse alors que nous n'avions échangé que deux fois. Un grand merci également de m'avoir impliqué dans vos cours et dans vos projets.

Je souhaite exprimer ma gratitude à Éric pour avoir accepté de co-encadrer la troisième année de ma thèse. Vous m'avez aidé à donner une nouvelle dynamique à ce travail. J'ai apprécié vos retours constructifs, votre écoute et votre bienveillance, rendant toujours nos échanges agréables.

Je tiens à remercier les membres du jury de cette thèse pour l'intérêt porté à mon travail. Je remercie particulièrement Mme Virginie Goepf et Mme Nada Matta pour avoir accepté de rapporter ma thèse et pour leurs rapports très complets. Je remercie également M. Jean-Philippe Kotowicz et M. Dominique Petitjean pour avoir accepté d'être examinateur.

Cette thèse n'aurait pas pu se dérouler sans les échanges avec les différentes parties prenantes des établissements scolaires, je tiens à remercier particulièrement Mme Sylvie Rousseau, M. Romain Mathieu, M. Alexandre Fregona, Mme Alix Jacquot, ainsi que le personnel et les lycéens du lycée Diderot de Langres.

Je remercie les membres du LITIS pour leurs enseignements de qualité pendant mes années de master, et pour leur accueil toujours chaleureux lors de mes deux premières années de thèse.

Ces trois années ont été remplies d'échanges et de moments agréables à l'ERPI, pour cela je remercie tous les membres du laboratoire et plus particulièrement, Raphaël, avec lequel j'ai eu le plaisir d'animer mes premiers Living Lab tout en traversant la région en musique ; Hakim, pour ses visites impromptues dans notre bureau qui égayaient la journée ; Juliana, pour tous les midis passés ensemble ; Valentine, pour le travail partagé autour du projet AILES ; Paco, pour sa bonne humeur contagieuse ; Neal, Maxim et Marine, avec qui j'ai été ravi de partager le bureau et nos cheminements de thèse.

Merci aux enseignants qui ont marqué ma scolarité et m'ont fait un peu trop aimer l'école/l'université au point de retarder le moment de quitter les études en effectuant un doctorat, je pense particulièrement à M. Perrier, M. Beucherie, et M. El Machkouri.

Un jour, nous sommes au lycée et puis, en l'espace d'un instant, nous mettons un point final à notre travail de thèse. Heureusement, il reste des souvenirs et des amis. Merci pour votre présence de près comme de loin qui compte beaucoup pour moi. Merci Edouard, Johanna, Julien et Pauline pour votre soutien qui embaume le fromage, les gâteaux, le chlore et les

chats. Merci Lucie pour ta présence constante. Merci David et Hélène pour nos marches qui m'ont à chaque fois remplie de bien-être et d'énergie. Et merci à Kévin et Coralie qui m'ont par leur exemple montré le chemin de la thèse.

Je me dois de finir ces remerciements par ma famille, qui me fournissent un soutien sans limites,

Ma belle-famille, Fabienne et Yves pour leur soutien discret mais toujours infaillible depuis le début de mes études supérieures. Emmanuel et Sandrine, pour leur vigilance accrue de mon moral lors de cette thèse et leurs mots/chansons de soutien ces deux dernières années.

Guillaume, mon super beau-frère, pour ses gentilles moqueries qui m'accompagnent depuis maintenant de nombreuses années et qui me font rire en tout temps.

Mimi, pour son affection, sa présence à mes côtés dans les pires et les meilleurs moments, et pour m'avoir toujours encouragé. Tu es une merveilleuse grande sœur, avec tout ce que cela englobe, tu es mon modèle, mon phare au milieu des tempêtes.

Merci à mes parents, à mon père qui ne dit rien ou si peu, mais dont je devine l'affection qu'il me porte ; à ma mère pour son soutien téléphonique quotidien, son amour et sa gentillesse infinis. Merci à tous les deux pour tous vos ravitaillements, qui bien plus qu'une aide logistique, m'ont été d'un grand réconfort.

Merci enfin à Tristan. Par tout temps, et en toutes saisons, nous en avons parcouru des kilomètres et des aventures. Tu es toujours là, lorsque c'est facile comme quand ça ne l'est pas. Merci de me suivre dans tous mes défis et de te les faire devenir tiens.

RESUME

Ce projet de thèse contribue au projet national AILES (Accompagnement à l'intégration des Lycéens dans l'Enseignement Supérieur) et s'intéresse plus particulièrement aux dispositifs numériques destinés à l'aide à l'orientation des lycéens. Le terme « dispositif numérique » désigne dans cette thèse l'ensemble des sites web, applications sur mobile et/ou sur ordinateur.

Une grande diversité d'informations est présente dans les dispositifs numériques dédiés à l'orientation. Certains d'entre eux proposent des outils d'échanges, tandis que d'autres disposent de témoignages, par exemple. Nous observons que certains dispositifs ont pour objectif d'informer les lycéens, tandis que d'autres sont dédiés à répondre à leurs questions ou à donner des idées de métiers. En considérant cette diversité, il est difficile pour les professionnels de l'orientation d'identifier quel(s) dispositif(s) est (sont) le(s) plus approprié(s) pour un lycéen, en fonction de son profil.

Plusieurs limites apparaissent face à ce problème, (1) le manque de représentation uniforme et formelle des dispositifs numériques d'aide à l'orientation, (2) l'absence de données à grande échelle sur l'impact de ce type de dispositifs, et (3) la possibilité limitée de mesurer un impact auprès de lycéens.

Partant de ces limites, cette thèse propose une ontologie et une approche pour évaluer et prédire l'impact sur les lycéens des dispositifs numériques d'aide à l'orientation.

Pour cela, nous avons travaillé sur trois propositions :

1) La description formelle des dispositifs numériques d'aide à l'orientation. Nous considérons ici la description des dispositifs comme une caractérisation de leur support, structure et contenu. Cette caractérisation permet d'avoir une description uniforme et formelle des dispositifs afin de pouvoir les comparer. Nous avons choisi d'utiliser une ontologie pour décrire ces dispositifs. Les ontologies permettent de décrire les dispositifs par des concepts et des relations possibles entre ces concepts, de créer des instanciations pour chaque dispositif et de réaliser une vérification.

2) La mesure de l'impact de ces dispositifs. Nous nous intéressons uniquement à l'impact du dispositif sur le processus d'orientation des lycéens. Plusieurs types d'impacts peuvent être considérés ainsi que plusieurs profils de lycéens. Dans cette thèse, un modèle est proposé pour mesurer l'impact des dispositifs numériques sur le processus d'orientation des lycéens.

3) La prédiction de l'impact de nouveaux dispositifs. Pour certains dispositifs, la mesure de l'impact n'est pas réalisable. De plus, le nombre de dispositifs pour lequel l'impact est connu est faible. Ainsi, cette thèse propose un modèle basé sur le raisonnement à partir de cas afin de prédire l'impact des dispositifs pour lequel il n'est pas mesurable.

Mots clefs : Ontologie, Mesure de l'impact, Prédiction, Raisonnement à partir de cas

ABSTRACT

Those research works contribute to the French national project called AILES (“Accompagnement à l'intégration des Lycéens dans l'Enseignement Supérieur”), and focus particularly on digital devices designed to help high school students with their career guidance. In this PhD thesis, the term "digital device" refers to all websites, mobile and/or computer applications.

There is a wide variety of information available about digital guidance devices. Some offer discussion tools, while others offer testimonials, for example. We can observe that some are designed to inform students, while others are dedicated to answering their questions or providing career ideas. Given this diversity, it is difficult for guidance professionals to determine which device(s) is most appropriate for a given high school student's profile.

There are several limitations to this problem: (1) the lack of a unified, formal representation of digital guidance devices, (2) the lack of large-scale data on the impact of this type of device, and (3) the limited ability to measure the impact on high school students.

Based on these limitations, this thesis proposes an ontology and an approach for evaluating and predicting the impact of digital guidance devices on high school students.

To achieve this, we worked on three proposals:

1) Formal description of digital guidance devices. Here, we consider the description of devices as a characterization of their support, structure, and content. This characterization provides a uniform and formal description of the devices in order to be compared. We have chosen to use an ontology to describe these devices. Ontologies make it possible to describe devices in terms of concepts and possible relationships between these concepts, to create instantiations for each device, and to perform verification.

2) Measuring the impact of these devices. We are only interested in the impact of the device on the guidance process. Several types of impact can be considered as well as several high school student profiles. In this thesis, we propose a model to measure the impact of digital devices on the guidance process of high school students.

3) Predicting the impact of new devices. For some devices, impact measurement is not feasible. Moreover, the number of devices for which the impact is known is small. Thus, this PhD thesis proposes a model based on case-based reasoning to predict the impact of devices for which it is not measurable.

Keywords: Ontology, Impact measurement, Prediction, Case based reasoning

TABLE DES MATIERES

Remerciements	III
Résumé	V
Abstract	VI
Table des matières	VII
Liste des figures	XII
Liste des tableaux	XIV
Remarques préliminaires.....	XVI
Chapitre 1 : Introduction générale	1
1. Contexte	2
1.1. Motivations	2
1.2. Contexte des travaux de recherche	3
2. Positionnement par rapport aux travaux de la littérature	3
3. Objectifs scientifiques et questions de recherche	5
3.1. Création d'une ontologie décrivant les dispositifs numériques d'aide à l'orientation	5
3.2. Mesure de l'impact sur le processus d'orientation des lycéens de l'utilisation des dispositifs numérique d'aide à l'orientation.....	6
3.3. Système basé sur un faible nombre de données pour prédire l'impact d'un nouveau dispositif numérique	6
4. Plan du document.....	7
Chapitre 2 : État de l'art.....	9
1. Introduction.....	10
2. Représentation des connaissances	11
2.1. Formes de représentation des connaissances.....	11
2.1.1. Représentation basée sur les réseaux (réseaux sémantiques).....	11
2.1.2. Représentation basée sur la structure (frames).....	12
2.1.3. Représentation basée sur des règles	13
2.1.4. Représentation basée sur la logique.....	13
2.2. Modèles conceptuels	14
2.3. Conclusion de la section sur la représentation des connaissances.....	15
3. Mesurer l'impact de dispositifs numériques.....	16
3.1. Définition de l'impact	16
3.2. État de l'art des indicateurs d'évaluation des dispositifs numériques.....	18
3.3. Travaux sur la recherche de corrélations entre des variables	20

3.3.1. Régression Linéaire Multiple.....	21
3.3.2. Modélisation par équation structurelle	22
3.3.3. Réseaux Bayésiens	23
3.3.4. Bilan des trois approches.....	23
3.4. Conclusion de la section sur la mesure de l'impact.....	24
4. Prédiction de l'impact.....	25
4.1. Approches de modélisation des connaissances et raisonnement	26
4.1.1. Représentation et raisonnement sur des connaissances à base logique	26
4.1.2. Raisonnement à partir de cas	26
4.1.3. Représentation et raisonnement sur des connaissances incertaines (réseaux bayésiens).....	27
4.2. Approches orientées données.....	28
4.2.1. Apprentissage supervisé	29
4.2.2. Apprentissage non supervisé	29
4.2.3. Apprentissage semi-supervisé	29
4.2.4. Apprentissage par renforcement.....	30
4.2.5. Positionnement par rapport aux approches orientées données.....	30
4.3. Synthèse des approches.....	30
4.4. Conclusion de la section sur la prédiction de l'impact.....	31
Chapitre 3 : ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation	33
1. Introduction.....	34
2. Dispositifs numériques dédiés à l'orientation scolaire des lycéens	35
2.1. Phases d'orientation vécues par les lycéens et les solutions numériques proposées.....	36
2.2. Exemples de dispositifs numériques d'aide à l'orientation à destination des lycéens	39
2.3. Ontologies relatives à l'orientation scolaire	41
2.4. Méthodologies de conceptualisation d'une ontologie	42
2.5. Synthèse et questions de recherche.....	43
3. Proposition d'une ontologie pour caractériser les dispositifs numériques d'aide à l'orientation	44
3.1. Méthodologie de conceptualisation de l'ontologie proposée	44
3.1.1. Étape 1 : Définir le domaine et la portée	45
3.1.2. Étape 2 : Identifier les ontologies existantes	45
3.1.3. Étape 3 : Lister les termes et les concepts	45
3.1.4. Étape 4 : Créer une hiérarchie de classes.....	46
3.1.5. Étapes 5 et 6 : Définir la structure interne des concepts	47
3.1.6. Étape 7 : Créer les instances	47
3.2. Ontologie des Dispositifs Numériques d'Aide à l'Orientation	47
3.2.1. Module « Nature de l'information »	49
3.2.2. Module « Interaction de l'utilisateur avec le dispositif numérique ».....	52
3.2.3. Module « Services additionnels »	53
3.2.4. Module « Utilisateur »	55
3.2.5. Module d' « Utilisabilité ».....	57

3.2.6. Module d' « Impact ».....	58
3.3. Vérification et validation de l'ontologie proposée	60
4. Conclusion	61
Chapitre 4 : Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation	63
1. Introduction.....	64
2. Description et mesure du KAB dans notre contexte	66
3. Évaluation des dispositifs numériques d'aide à l'orientation	67
3.1. Synthèse sur les indicateurs d'évaluation des dispositifs numériques	67
3.2. Définition des catégories d'indicateur dans le contexte de l'étude des dispositifs dédiés à l'orientation scolaire	69
3.2.1. Utilisateur	69
3.2.2. Ressenti global.....	70
3.2.3. Valeur perçue du dispositif	70
3.2.4. Valeur émotionnelle provoquée (hédonique).....	70
3.2.5. Valeur sociale induite	71
3.2.6. Utilité perçue du dispositif.....	71
3.2.7. Utilisation du dispositif et Valeur sensorielle provoquée	72
3.2.8. Comportement futur de l'utilisateur.....	73
3.2.9. Synthèse des indicateurs utilisés dans notre modèle	73
3.3. Modèle de relation entre l'impact et les indicateurs d'évaluation des dispositifs numériques	74
3.3.1. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques par une régression linéaire multiple	75
3.3.2. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques par une modélisation par équation structurelle	76
3.3.3. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques avec les réseaux bayésiens	77
3.3.4. Synthèse	77
4. Protocole d'estimation de l'impact.....	78
5. Conclusion	80
Chapitre 5 : Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation	81
1. Introduction.....	82
2. Présentation du fonctionnement du Raisonnement à Partir de Cas (RàPC)	83
2.1. Définition des cas	83
2.2. Cycle du RàPC	85
2.3. Récupération	85
2.3.1. Similarité locale	86
2.3.2. Similarité globale	87
2.3.3. Récupération des cas sources après les calculs de similarité	88
2.4. Réutilisation	88
2.5. Révision	89

2.6. Mémorisation.....	89
2.7. Synthèse et question de recherche.....	90
3. Notre modèle de prédiction de l'impact de dispositifs numériques.....	90
3.1. Description des cas.....	90
3.2. Propositions de prédiction basées sur le RàPC.....	95
3.2.1. Similarités locales.....	96
3.2.2. Modèle de prédiction en suivant le cycle du RàPC.....	98
3.2.3. Modèle de Raisonnement à Partir des Attributs (RàPA) inspiré du cycle du RàPC, basée sur la similarité entre chaque descripteur.....	100
3.2.4. Mémorisation des prédictions du cas cible.....	102
4. Conclusion.....	103
Chapitre 6 : Expérimentations.....	104
1. Introduction.....	105
2. Instanciation de l'ontologie pour les dispositifs Hello Charly et Onisep.....	106
2.1. Hello Charly.....	106
2.1.1. Présentation du dispositif Hello Charly.....	106
2.1.2. Instanciation de l'ontologie proposée pour le dispositif Hello Charly.....	108
2.2. Onisep.....	117
2.2.1. Présentation du dispositif Onisep.....	117
2.2.2. Instanciation du dispositif Onisep.....	119
2.3. Synthèse des instanciations.....	122
3. Estimation de l'impact.....	122
3.1. Expérimentations pour créer notre modèle.....	122
3.1.1. Protocole de mesure.....	122
3.1.2. Résultats obtenus.....	123
3.2. Expérimentation de l'estimation de l'impact.....	126
3.2.1. Mesure de l'impact du dispositif Hello Charly.....	126
3.2.2. Mesure de l'impact du dispositif Onisep.....	128
4. Expérimentation des deux modèles de prédiction proposés.....	130
4.1. Importance des caractéristiques des dispositifs numériques.....	130
4.2. Proximité entre les instances des concepts « interlocuteur », « support de l'information » et « type de test ».....	135
4.3. Prédiction de l'impact.....	137
4.3.1. Prédiction de l'impact selon le modèle de prédiction en suivant le cycle du RàPC.....	138
4.3.2. Prédiction de l'impact selon le modèle RàPA.....	140
4.3.3. Synthèse des prédictions.....	142
4.4. Étude de la sensibilité de l'impact selon le poids des critères.....	143
5. Conclusion.....	146

Chapitre 7 : Conclusion et Perspectives	147
1. Contexte et questions de recherche.....	147
2. Contribution	149
2.1. ODNAO : Une ontologie pour représenter de façon formelle les dispositifs numériques d'aide à l'orientation	149
2.2. Modèle de mesure de l'impact de dispositifs numériques d'aide à l'orientation...	150
2.3. Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation	152
3. Perspectives de recherche.....	153
3.1. Perspectives à court terme.....	153
3.1.1. ODNAO : Une ontologie pour représenter de façon formelle les dispositifs numériques d'aide à l'orientation	153
3.1.2. Modèle de mesure de l'impact de dispositifs numériques d'aide à l'orientation	154
3.1.3. Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation.....	154
3.2. Perspectives à long terme	154
 Bibliographie	 156
 Annexes	 168
Annexe A : Extrait de la première liste de termes	169
Annexe B : Questionnaire pour estimer l'impact.....	170
Annexe C : Protocole pour instancier l'ontologie ODNAO pour un dispositif numérique.....	173
Annexe D : Capture d'écran de l'instanciation du Portail de l'Onisep sur Protégé	181
Annexe E : Questionnaire AHP pour obtenir les poids des critères	182

LISTE DES FIGURES

Figure 1.1 - Organisation du mémoire de thèse -----	8
Figure 2.1 - Exemple de réseau sémantique [16]-----	12
Figure 2.2 - Représentation schématique de notre recherche de coefficients de corrélation	20
Figure 2.3 - Schéma illustrant la représentation des variables et des relations pour les modèles à équation structurelle-----	22
Figure 2.4 - Principe de fonctionnement du RàPC [82] -----	27
Figure 2.5 - Exemple de réseau bayésien avec la table de probabilité-----	28
Figure 3.1 - Processus d'orientation divisé en quatre phases-----	37
Figure 3.2 - Extrait du processus proposé pour le cycle de vie d'une ontologie [109] -----	43
Figure 3.3 - Extrait de l'Ontologie des Dispositifs Numériques d'Aide à l'Orientation-----	49
Figure 4.1 - Cadre de l'estimation de l'impact centrée sur l'expérience perçue par les lycéens -----	74
Figure 4.2 - Diagramme de cheminement PLS-SEM -----	76
Figure 4.3 - Réseau bayésien représentant les états possibles de chaque nœud-----	77
Figure 5.1 - Cycle du RàPC selon Aamodt et al. [137] -----	85
Figure 5.2 - Capture d'écran de Protégé du module « Nature de l'information » de l'ontologie proposée-----	92
Figure 5.3 - Capture d'écran de Protégé du module Services additionnels de l'ontologie proposée-----	93
Figure 5.4 - Algorithme de la fonction <i>SimilariteCategoriel()</i> -----	97
Figure 5.5 - Matrices de similarité pour les attributs de type catégoriel -----	97
Figure 5.6 - Algorithme de calcul du Paramètre de Révision proposé -----	99
Figure 5.7 - Algorithme de prédiction à partir des attributs-----	101
Figure 6.1 - Captures d'écran du site web Hello Charly-----	107
Figure 6.2 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Nature de l'information » -----	109
Figure 6.3 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Interaction de l'utilisateur avec le dispositif numérique » -----	110
Figure 6.4 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Services additionnels »-----	111
Figure 6.5 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Utilisabilité »-----	114
Figure 6.6 - Échelle des scores du SUS -----	115
Figure 6.7 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Utilisateur » -----	115

Figure 6.8 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Impact » -----	116
Figure 6.9 - Captures d'écran du portail de l'Onisep -----	119
Figure 6.10 - Distribution des données collectées -----	123
Figure 6.11 - Diagramme de cheminement PLS-SEM obtenu à partir du logiciel RStudio ---	124
Figure 6.12 - Dispersion des moyennes pour les affirmations relatives à l'utilité (1), à la valeur hédonique (2), et à la valeur sociale (3) pour le dispositif Hello Charly-----	127
Figure 6.13 - Dispersion des moyennes pour les questions relatives à l'utilité (1), à la valeur hédonique (2), et à la valeur sociale (3) pour le dispositif de l'Onisep-----	129
Figure 6.14 - Comparaison des critères « Témoignage » et « FAQ » pour le processus d'analyse hiérarchique -----	131
Figure 6.15 - Image du paperboard utilisé pour obtenir la similarité entre les différents supports d'information dans les dispositifs numériques -----	135

LISTE DES TABLEAUX

Tableau 2.1 - Exemple de représentation basée sur la structure [14] -----	13
Tableau 2.2 - Comparaison des modèles pour le traitement des données et pour la prédiction d'un impact -----	31
Tableau 3.1 - Phases d'orientation pour lesquelles les dispositifs numériques sont destinés	39
Tableau 3.2 - Temporalité et variété des informations fournies par un échantillon de dispositifs numériques d'aide à l'orientation-----	40
Tableau 3.3 - Dispositifs numériques d'orientation observés-----	46
Tableau 3.4 - Classes du module « Nature de l'information » -----	50
Tableau 3.5 - Propriétés du module « Nature de l'information » -----	51
Tableau 3.6 - Classes du module « Interaction de l'utilisateur avec le dispositif numérique » -----	52
Tableau 3.7 - Propriétés du module « Interaction de l'utilisateur avec le dispositif numérique »-----	53
Tableau 3.8 - Classes du module « Services additionnels » -----	54
Tableau 3.9 - Propriétés du module « Services additionnels »-----	55
Tableau 3.10 - Classes du module « Utilisateur »-----	55
Tableau 3.11 - Propriétés du module « Utilisateur » -----	56
Tableau 3.12 - Classes du module « Utilisabilité » -----	57
Tableau 3.13 - Propriétés du module « Utilisabilité » -----	58
Tableau 3.14 - Classes du module « Impact »-----	59
Tableau 3.15 - Propriétés du module « Impact » -----	60
Tableau 4.1 - Synthèse des catégories d'indicateur et leur utilisation possible -----	69
Tableau 5.1 - Exemple d'une structure d'un cas adapté à la prédiction-----	84
Tableau 5.2 - Attributs des cas pour le module « Nature de l'information »-----	91
Tableau 5.3 - Attributs des cas pour le module « Services Additionnels »-----	93
Tableau 5.4 - Attributs des cas pour le module « Interaction entre le dispositif et l'utilisateur » -----	93
Tableau 5.5 - Attributs des cas pour le module « Utilisabilité » -----	94
Tableau 5.6 - Attributs des cas pour le module « Utilisateur »-----	94
Tableau 6.1 - Questionnaire pour mesurer l'utilisabilité d'un dispositif -----	113
Tableau 6.2 - Valeurs des propriétés d'objet du module « Utilisabilité » pour Hello Charly	113
Tableau 6.3 - Valeur et niveau d'utilisabilité correspondants-----	113
Tableau 6.4 - Valeurs des propriétés d'objet du module « Utilisateur »-----	115
Tableau 6.5 - Valeurs des propriétés d'objet du module « Impact »-----	116

Tableau 6.6 - Instances du module « Nature de l'information » pour le dispositif du portail de l'Onisep -----	120
Tableau 6.7 - Instances du module « Interaction de l'utilisateur avec le dispositif numérique » pour le dispositif du portail de l'Onisep -----	120
Tableau 6.8 - Instances du module « Services additionnels » pour le dispositif du portail de l'Onisep -----	120
Tableau 6.9 - Valeurs des propriétés d'objet du module « Utilisabilité » pour le portail de l'Onisep -----	121
Tableau 6.10 - Valeurs des propriétés d'objet du module « Utilisateur » -----	121
Tableau 6.11 - Valeurs des propriétés d'objet du module « Impact » -----	121
Tableau 6.12 - Tableau de probabilités conditionnelles -----	125
Tableau 6.13 - Poids des différents types d'information possible -----	132
Tableau 6.14 - Poids des différents éléments disponibles dans les dispositifs numériques -----	133
Tableau 6.15 - Poids donnant l'importance des dix-sept parties des dispositifs numériques -----	134
Tableau 6.16 - Valeurs de similarité entre les différents supports d'information -----	136
Tableau 6.17 - Valeurs de similarité entre les différents interlocuteurs des outils d'échange -----	136
Tableau 6.18 - Valeurs de similarité entre les différents sujets de test -----	137
Tableau 6.19 - Description des caractéristiques des cas des dispositifs fictifs -----	138
Tableau 6.20 - Impact prédit et mesuré pour trois dispositifs numériques basé sur le cycle du RàPC -----	139
Tableau 6.21 - Impact prédit et mesuré pour trois dispositifs numériques basé sur le RàPA -----	140
Tableau 6.22 - Tableau présentant le détail de la prédiction de l'impact de Study Advisor par RàPA -----	142
Tableau 6.23 - Exemple de l'élaboration d'un vecteur de poids modifié selon l'intervalle $[-5 ; 5]$ -----	144
Tableau 6.24 - Impact moyen de dispositifs numérique en ajoutant PR en faisant varier les poids -----	145
Tableau 6.25 - Impact moyen de dispositifs numérique en utilisant PR comme un pourcentage de changement en faisant varier les poids -----	145
Tableau 6.26 - Impact moyen de dispositifs numérique en utilisant le RàPA en faisant varier les poids -----	145

REMARQUES PRELIMINAIRES

Tous les liens hypertextes présents dans ce document étaient valides au 12 octobre 2023.

Les instances, classes et propriétés ontologiques sont écrits entre guillemets doubles de la façon suivante : « instance », « propriété », « Classe ».

Les légendes sont placées en dessous des tableaux et des figures. Les tables des tableaux et figures sont disponibles ci-dessus.

INTRODUCTION GENERALE

TABLE DES MATIERES

1. Contexte	2
1.1. Motivations	2
1.2. Contexte des travaux de recherche	3
2. Positionnement par rapport aux travaux de la littérature	3
3. Objectifs scientifiques et questions de recherche	5
3.1. Création d'une ontologie décrivant les dispositifs numériques d'aide à l'orientation ...	5
3.2. Mesure de l'impact sur le processus d'orientation des lycéens de l'utilisation des dispositifs numérique d'aide à l'orientation	6
3.3. Système basé sur un faible nombre de données pour prédire l'impact d'un nouveau dispositif numérique.....	6
4. Plan du document	7

1. Contexte

Ce travail de thèse se place dans le contexte de l'orientation scolaire, et analyse plus particulièrement les dispositifs numériques dédiés à l'aide à l'orientation.

1.1. Motivations

En 2022, environ 3 millions de jeunes en France se sont inscrits à l'université pour la première fois. Avant cette inscription, ces jeunes étaient des lycéens qui ont dû réfléchir à leurs choix d'études après l'obtention du baccalauréat. Ce choix se construit au fur et à mesure de la scolarité des lycéens. Ce temps de construction implique que les lycéens sont confrontés à une variété d'informations à ce sujet. Les conseillers d'orientation, les enseignants et les proches des lycéens étaient initialement les principales sources d'information lorsqu'il s'agissait d'échanger sur l'orientation. Aujourd'hui, avec l'utilisation grandissante du numérique à l'échelle personnelle mais aussi scolaire, de nombreux sites Web et applications voient le jour dans le domaine de l'orientation.

Les lycéens ont ainsi accès à un nouveau canal de transmission de l'information par le biais de dispositifs numériques dédiés à l'orientation.

Nous utilisons dans la suite ce travail le terme « dispositifs numériques » pour désigner l'ensemble des sites web, applications sur mobile et/ou sur ordinateur.

Une grande diversité d'information est présente dans les dispositifs numériques dédiés à l'orientation. Certains d'entre eux proposent des outils d'échanges, tandis que d'autres disposent de témoignages par exemple. Nous observons que certains dispositifs ont pour objectif d'informer les lycéens, tandis que d'autres sont dédiés à répondre à leurs questions ou à donner des idées de métiers.

Aujourd'hui, il existe de nombreux dispositifs numériques. Des sites tels que celui de l'Onisep (<https://www.onisep.fr/>) et celui de l'Étudiant (<https://www.letudiant.fr/>) proposent une grande variété d'informations tandis que des applications telles que Wilbi (<https://wilbiapp.com/>) traitent de façon plus spécifique des métiers de façon concrète grâce à des vidéos réalisées par des professionnels. Une catégorie de dispositif est tournée sur les tests de personnalité comme Quel Metier (<http://www.quelmetier.fr/>). Certains dispositifs ont pour objectif de permettre aux lycéens d'échanger, de se créer un réseau comme Study Advisor (<https://www.studyadvisor.fr/>) qui vise à mettre en relation des lycéens et étudiants pour répondre notamment aux questions sur la vie étudiante.

Les rectorats, et plus particulièrement les psychologues de l'éducation nationale et les enseignants, accompagnent les lycéens dans leur démarche d'orientation, et ce pour les différents profils d'élèves (Générale, Professionnelle, Technologique) et à différentes étapes

de leur cursus au lycée : classe de Seconde, classe de Première, début de classe de Terminale, classe de Terminale avant les vœux. Une solution pour cela serait que les professionnels de l'orientation puissent conseiller aux lycéens certains supports comme les dispositifs numériques d'aide à l'orientation. Mais, face à la diversité des dispositifs numériques, il leur est difficile d'identifier quel(s) dispositif(s) est (sont) le(s) plus approprié(s) pour un lycéen, en fonction de son profil et l'étape considérée.

Nous proposons dans cette thèse un éclairage sur les dispositifs numériques d'aide à l'orientation pour aider les professionnels qui accompagnent les lycéens dans leur processus d'orientation. Cela se traduit, dans le cadre de ce travail, par la proposition d'une ontologie afin de stocker uniformément et de façon formelle les caractéristiques des dispositifs numériques, ainsi que par la proposition d'une approche pour mesurer et prédire l'impact sur les lycéens des dispositifs numériques sur leur processus d'orientation.

1.2. Contexte des travaux de recherche

Ce travail de thèse a été effectué au sein du laboratoire Équipe de Recherche sur les Processus Innovatifs (ERPI).

Nos travaux de recherche ont été effectués dans le cadre du projet AILES (Accompagnement à l'intégration des Lycéens dans l'Enseignement Supérieur). Le projet AILES est né de la réponse à un appel à projet du Programme d'Investissement d'Avenir « Dispositifs territoriaux pour l'orientation vers les études supérieures » du PIA 3 (Programme d'Investissements d'Avenir). Il bénéficie d'un financement sur une durée de 10 ans (2020-2030) et s'étend sur deux territoires, l'académie de Reims et l'académie de Nancy-Metz.

Un des objectifs de ce projet, dans lequel s'inscrit cette thèse est de donner envie aux jeunes de se lancer dans une démarche d'orientation de qualité et proactive. Les dispositifs numériques apparaissent comme une solution pour donner envie aux jeunes. Cependant, nous allons montrer qu'il n'existe pas d'étude permettant de connaître leur impact sur le processus d'orientation.

2. Positionnement par rapport aux travaux de la littérature

Plusieurs travaux de recherche décrivent des dispositifs numériques dédiés à l'orientation capables de faire des recommandations à l'utilisateur selon son profil [1][2], ou pour des profils particuliers de lycéens, par exemple pour des minorités [3]. Notre objectif n'est pas de définir l'architecture du dispositif idéal, ni de chercher celui proposant la meilleure orientation scolaire à l'utilisateur, mais de mesurer l'impact de leur utilisation sur les lycéens.

Plusieurs travaux de recherche s'intéressent à la mesure de l'impact des dispositifs numériques sur les utilisateurs. Dans le domaine de l'éducation, l'intérêt se porte sur les dispositifs dédiés à l'apprentissage d'une matière [4] ou sur les dispositifs de tests scolaires [5]. En médecine, Benson [6] s'interroge sur les façons et les raisons de la diffusion de dispositifs numériques. Les mesures de l'impact d'un dispositif numériques sur les utilisateurs sont très variées et sont dépendantes du contexte. Elles nécessitent de définir l'impact à mesurer selon le contexte ainsi que les indicateurs pour l'évaluer.

Dans cette étude, pour définir l'impact, nous nous basons sur l'objectif du projet, en l'occurrence, déterminer si l'utilisation du dispositif a contribué au dynamisme des lycéens dans la construction de leur projet d'orientation. Pour cela nous utilisons la définition de l'impact sur les utilisateurs défini par Fokkinga et al. [7], qui est séparée selon quatre aspects : l'impact en termes de comportement (actions engendrées par l'utilisation du dispositif), l'attitude (opinions, émotions), l'expérience de l'utilisateur et enfin ses connaissances.

Le calcul de l'impact d'un dispositif nécessite l'intervention d'un expert ou l'utilisation d'une méthode d'évaluation. En effet, le calcul d'un impact est un processus complexe qui requiert de multiples connaissances. Ainsi, nous observons plusieurs verrous :

- L'hétérogénéité des dispositifs numériques,
- L'hétérogénéité des profils de lycéen,
- La multiplicité des sources pouvant impacter le processus d'orientation.

Pour répondre à ces questions, nous devons collecter, de façon formelle les informations de chaque dispositif numérique à étudier. Les ontologies apparaissent comme des approches appropriées pour représenter les dispositifs numériques utilisés dans l'orientation scolaire, puisqu'elles offrent une représentation conceptuelle de la connaissance d'un domaine. Il existe des ontologies dans le domaine de l'éducation [8], [9], mais aucune ne traite à la fois de l'orientation scolaire et des dispositifs numériques.

Nous avons identifié plusieurs limites dans les travaux de la littérature dédiés à la mesure de l'impact des dispositifs numériques d'aide à l'orientation scolaire :

1. (L1) Il n'existe pas de représentation uniforme permettant de faire ressortir les spécificités de chaque dispositif telles que leur structure informationnelle, leur processus de mise à jour de l'information ou le processus de médiation avec les lycéens.
2. (L2) La mesure de l'impact d'un dispositif sur des utilisateurs repose sur des indicateurs dépendant du contexte d'utilisation du dispositif et du type d'impact recherché. Il n'existe pas de mesure de l'impact des dispositifs numériques d'aide à l'orientation sur le processus d'orientation des lycéens. Il n'existe donc pas de donnée à grande échelle de l'impact d'un dispositif ou d'un autre.

3. (L3) Cette absence de données sur l'impact ne va pas permettre l'utilisation d'algorithmes d'apprentissage qui pourraient fournir des recommandations sur les dispositifs les plus appropriés à chaque lycéen. Mesurer l'impact des dispositifs auprès des lycéens nécessite de les mobiliser spécifiquement pour notre étude, ce qui limite la possibilité de connaître l'impact pour un grand nombre de dispositifs ou un grand nombre de lycéens.

3. Objectifs scientifiques et questions de recherche

L'objectif de cette thèse peut être résumée ainsi :

Proposer une ontologie et une approche pour évaluer et prédire l'impact sur les lycéens des dispositifs numériques d'aide à l'orientation.

Cet objectif peut être décomposé en trois sous-objectifs.

Le premier consiste à analyser l'architecture des dispositifs numériques d'aide à l'orientation des lycéens afin de proposer un vocabulaire et une sémantique pour les représenter de façon uniforme. Cet objectif répond à la première limite des travaux de la littérature (L1).

Le deuxième objectif consiste à proposer des indicateurs afin de mesurer l'impact des dispositifs numériques sur le processus d'orientation des lycéens. Ce deuxième objectif répond à la limite (L2).

Enfin, le dernier objectif est de prédire l'impact de dispositifs pour lesquels il n'est pas possible de le mesurer afin d'aider le rectorat et le personnel éducatif à proposer les dispositifs les plus adaptés à une classe de lycéens. Ce dernier objectif répond à la limite (L3).

3.1. Création d'une ontologie décrivant les dispositifs numériques d'aide à l'orientation

Les dispositifs numériques d'aide à l'orientation des lycéens ont des architectures variées. Le premier objectif de nos travaux est de concevoir une représentation uniforme de ces dispositifs permettant d'identifier au mieux leurs spécificités et leurs ressemblances. Nous proposons de représenter ces dispositifs à l'aide d'une ontologie, c.-à-d. de construire une ontologie qui puisse être instanciée pour chaque dispositif à étudier afin de représenter ses caractéristiques. Notons que les caractéristiques intéressantes à représenter dans notre problème sont des éléments visibles par les utilisateurs et qui changent leur perception de l'utilisation, comme par exemple la présence de témoignage ou l'utilisation sur un site web ou une application mobile. Nous avons choisi d'utiliser une ontologie pour représenter les dispositifs numériques afin de pouvoir ajouter des contraintes de cardinalité et de vérifier la

cohérence des instanciations. De ce fait, notre premier objectif renvoie à la question de recherche suivante :

QR1

Comment modéliser et stocker les caractéristiques des dispositifs numériques d'aide à l'orientation tout en étant robuste à l'apparition de nouveaux dispositifs ?

Cette contribution a fait l'objet d'une publication dans la conférence internationale KEOD 2022 : « Ontology-Based Representation of Digital Devices Used in Educational Guidance » [10].

3.2. Mesure de l'impact sur le processus d'orientation des lycéens de l'utilisation des dispositifs numérique d'aide à l'orientation

Le deuxième objectif de nos travaux est de mesurer l'impact des dispositifs numérique sur le processus d'orientation des lycéens afin de proposer un indicateur utile au rectorat et au personnel éducatif pour choisir les dispositifs les plus adaptés selon la classe. Étant donné l'importance de cette information pour le personnel éducatif, il est nécessaire de proposer une mesure compréhensible et en accord avec leurs objectifs. Ainsi, à partir de l'identification des indicateurs pour mesurer l'impact, l'objectif est de lier les différents indicateurs avec les aspects de l'impact (comportement, attitude, expérience et connaissance). De ce fait, notre deuxième objectif renvoie à la problématique suivante :

QR2

Comment mesurer l'impact des dispositifs numériques d'aide à l'orientation sur les lycéens dans leur démarche d'orientation ?

Cette contribution a fait l'objet de deux publications. La première « Assessment Methods to ICT Platform Used in Educational Guidance: Application to an Immersive World » [11] a été présentée lors de la conférence internationale IICE (Ireland International Conference on Education) 2021. La seconde « Computer-assisted Career Guidance Systems User Impact: An Evaluation study » [12] a été présentée lors de la conférence internationale ICE (International Conference on Engineering, Technology, and Innovation) 2022.

3.3. Système basé sur un faible nombre de données pour prédire l'impact d'un nouveau dispositif numérique

Les décideurs dans l'Éducation Nationale sont amenés à choisir l'utilisation future des dispositifs numériques avant ou pendant leur développement. Ainsi, il n'est pas possible de

mesurer l'impact auprès des lycéens. Cette impossibilité de mesure tout en devant prendre une décision entraîne la nécessité de la mise en place de prédictions.

Les méthodes classiques de prédiction d'une valeur à partir de caractéristiques connues nécessitent un nombre de données important, alors, que dans notre contexte le nombre de dispositifs numériques pour lesquels nous pouvons mesurer l'impact est faible. Le troisième objectif de cette thèse est donc de proposer un système prédictif permettant aux décideurs de connaître la valeur de l'impact d'un dispositif alors qu'il n'est pas possible de le mesurer, en utilisant les caractéristiques (QR1) et en utilisant les dispositifs pour lesquels l'impact a été mesuré (QR2). De ce fait, notre dernier objectif correspond à la problématique suivante :

QR3

Comment concevoir un système prédictif pour obtenir l'impact de dispositifs numériques pour lesquels il n'est pas possible de le mesurer et en s'appuyant sur la connaissance d'un nombre limité de dispositifs dont la valeur d'impact est connue ?

Dans ces trois sections, nous avons présenté l'objectif de la thèse et les trois sous-objectifs qui y sont associés. Nous présentons dans la section suivante le plan du document.

4. Plan du document

Le plan de ce document de thèse est présenté dans la [Erreur ! Source du renvoi introuvable.](#) Le Chapitre « État de l'art » expose l'état de l'art existant sur nos questions de recherche. Les trois chapitres suivants présentent nos contributions pour chacune de nos trois questions de recherche (QR1), (QR2) et (QR3). Le contexte expérimental et les contributions sont quant à eux présentés dans le Chapitre « Expérimentations ». Enfin, nous concluons ce mémoire en résumant les contributions, les résultats des expérimentations et enfin une discussion autour des perspectives de recherche.

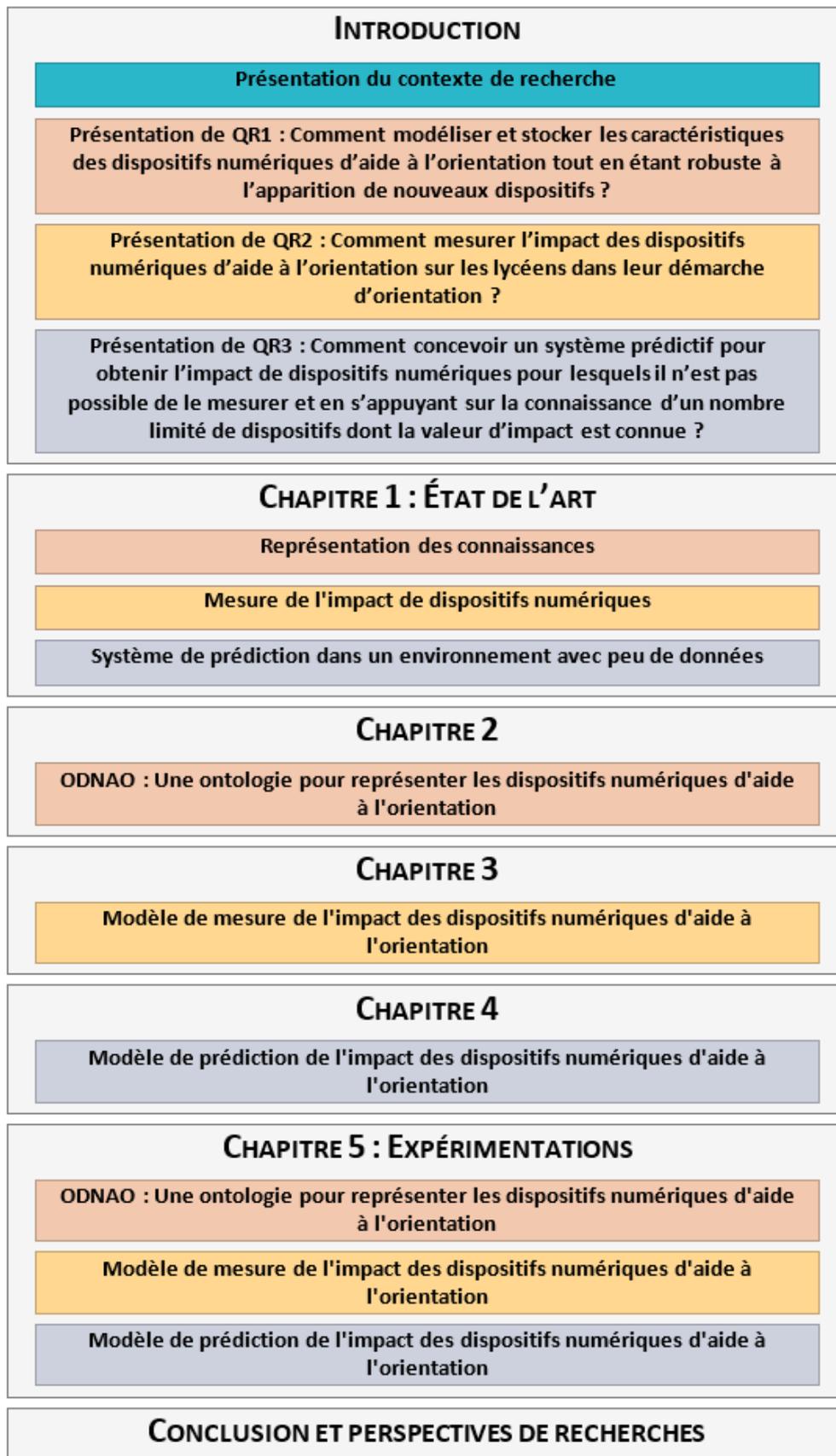


Figure 1.1 - Organisation du mémoire de thèse

ÉTAT DE L'ART

TABLE DES MATIERES

1. Introduction.....	10
2. Représentation des connaissances.....	11
2.1. Formes de représentation des connaissances.....	11
2.1.1. Représentation basée sur les réseaux (réseaux sémantiques).....	11
2.1.2. Représentation basée sur la structure (frames).....	12
2.1.3. Représentation basée sur des règles.....	13
2.1.4. Représentation basée sur la logique.....	13
2.2. Modèles conceptuels.....	14
2.3. Conclusion de la section sur la représentation des connaissances.....	15
3. Mesurer l'impact de dispositifs numériques.....	16
3.1. Définition de l'impact.....	16
3.2. État de l'art des indicateurs d'évaluation des dispositifs numériques.....	18
3.3. Travaux sur la recherche de corrélations entre des variables.....	20
3.3.1. Régression Linéaire Multiple.....	21
3.3.2. Modélisation par équation structurelle.....	22
3.3.3. Réseaux Bayésiens.....	23
3.3.4. Bilan des trois approches.....	23

3.4. Conclusion de la section sur la mesure de l'impact	24
4. Prédiction de l'impact.....	25
4.1. Approches de modélisation des connaissances et raisonnement.....	26
4.1.1. Représentation et raisonnement sur des connaissances à base logique.....	26
4.1.2. Raisonnement à partir de cas.....	26
4.1.3. Représentation et raisonnement sur des connaissances incertaines (réseaux bayésiens)	27
4.2. Approches orientées données.....	28
4.2.1. Apprentissage supervisé.....	29
4.2.2. Apprentissage non supervisé.....	29
4.2.3. Apprentissage semi-supervisé.....	29
4.2.4. Apprentissage par renforcement	30
4.2.5. Positionnement par rapport aux approches orientées données	30
4.3. Synthèse des approches	30
4.4. Conclusion de la section sur la prédiction de l'impact.....	31

1. Introduction

L'intérêt de la représentation des connaissances dépend du modèle conceptuel choisi. Le modèle choisi doit correspondre aux connaissances à représenter et à l'utilisation de ces connaissances. Différents travaux présentent ces modèles et leurs intérêts et inconvénients.

La représentation des caractéristiques des dispositifs numérique n'est pas suffisante, l'impact sur l'utilisateur des dispositifs est une connaissance importante. Afin de représenter cette connaissance, il est nécessaire de déterminer la façon dont l'obtenir, c'est-à-dire comment mesurer l'impact des dispositifs numériques.

La mesure de l'impact nécessitant l'accès aux utilisateurs, il peut être délicat pour certains dispositifs de réaliser cette mesure. Ainsi, il est intéressant d'analyser la façon dont prédire cet impact.

Dans ce chapitre, nous considérons le niveau générique des problématiques et de leurs solutions, toutefois, il est parfois nécessaire de considérer leur niveau particulier c'est-à-dire les dispositifs numériques d'aide à l'orientation et leur impact sur le processus d'orientation des lycéens. Les paragraphes traitant du niveau particulier sont mis en italiques avec un retrait et une bordure afin de les distinguer.

Ce premier chapitre présente un état de l'art des travaux de la littérature pour chacune de nos questions de recherche. Tout d'abord, dans la première section, un état de l'art est fait sur les modèles conceptuels de représentation des connaissances pour répondre à la question (QR1). La deuxième section présente les travaux existants sur la mesure de l'impact afin de répondre à la question (QR2). Finalement, la troisième section cherche à répondre à la question (QR3) en analysant les techniques de prédiction existantes.

2. Représentation des connaissances

Un grand nombre de données peuvent être extraites des dispositifs numériques d'aide à l'orientation. Une exigence importante de ce travail est de structurer les données pertinentes afin de rendre possible une comparaison entre les dispositifs numériques. Une autre exigence est de rendre accessible et réutilisables ces données pour les professionnels de l'orientation. Ainsi, la représentation des connaissances est un point central pour caractériser des dispositifs numériques d'aide à l'orientation.

La représentation des connaissances est définie comme l'utilisation de symboles formels pour représenter un ensemble de propositions [13]. Cette représentation est une approximation de la réalité puisque le nombre de propositions représenté est limité [14]. Ainsi, les connaissances sont plus ou moins mises en avant ; en réalisant une telle reproduction, un choix est fait sur ce qui est conservé ou non de la réalité.

2.1. Formes de représentation des connaissances

Nous considérons dans cette section les quatre formes les plus fréquentes de représentation des connaissances [15] qui sont les réseaux sémantiques, les frames, les règles et la logique.

2.1.1. Représentation basée sur les réseaux (réseaux sémantiques)

Les réseaux sémantiques sont adaptés pour capturer la structure taxonomique et hiérarchique d'un domaine. Les nœuds représentent des concepts ou des objets du domaine et les relations entre les concepts et objets sont représentés par des arcs [14][16].

Un exemple de réseau sémantique sur le domaine scolaire est donné dans la **Figure 2.1**. Les nœuds du réseau représentent des concepts (par exemple « le professeur »), et des instances (par exemple « John »). Les arcs entre les nœuds relient les concepts et les instances. Le réseau indique que John est un professeur délivrant des cours de calcul et qu'il travaille à l'université.

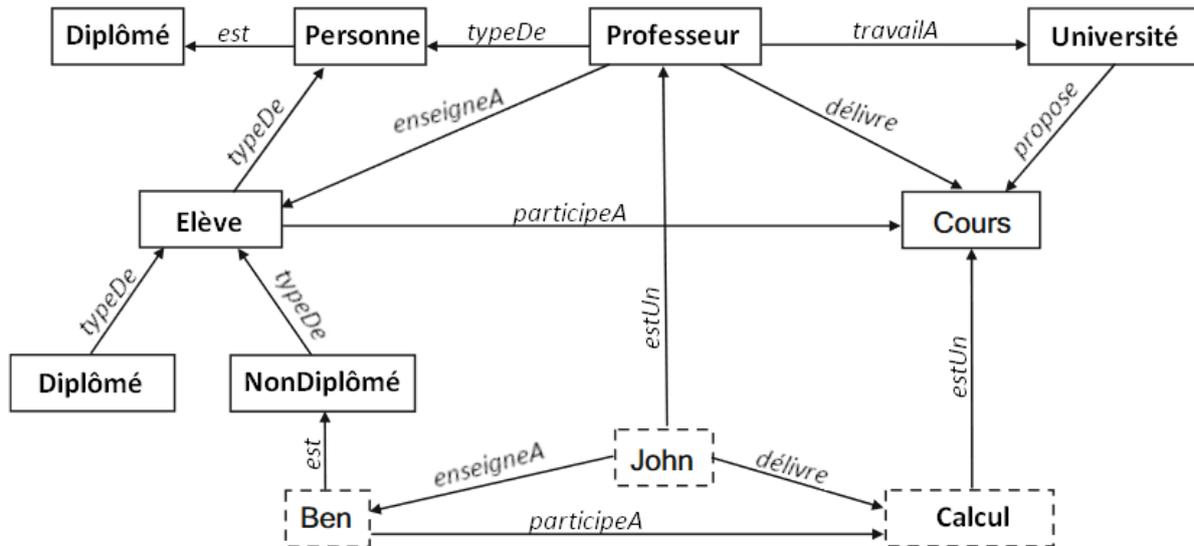


Figure 2.1 - Exemple de réseau sémantique [16]

Les réseaux sémantiques sont une représentation semi-formelle des connaissances. Les représentations semi-formelles se placent entre les représentations informelles et formelles. Les représentations informelles sont le langage naturel écrit, et les représentations formelles permettent de réaliser des tests, et d'effectuer une validation automatisée, mais ne sont compréhensibles que par des experts [17].

Le manque de formalisme est, pour les réseaux sémantiques induit par le fait que les individus (instances) et les classes sont considérés de la même façon et qu'il n'existe pas de contrainte sur les valeurs des nœuds et des arcs [15]. Ces manques induisent l'impossibilité de réaliser une validation automatique.

La modélisation semi-formelle permet de réduire le temps de conception du modèle mais peut augmenter le nombre d'erreurs introduites dans le processus de développement [18].

2.1.2. Représentation basée sur la structure (frames)

La représentation basée sur la structure peut être considérée comme une évolution des réseaux sémantiques [14]. Elle peut être considérée comme un réseau de nœuds et de relations. Les niveaux supérieurs de la structure sont fixes et représentent des propriétés qui sont toujours présentes dans le contexte donné. Les niveaux inférieurs de la structure comportent des emplacements qui doivent être remplis par des instances ou des données [19].

Ce type de représentation se place dans le paradigme du monde fermé. Cela signifie que tous les faits représentés dans le système sont vrais, et si un fait n'est pas représenté, c'est qu'il est faux [20].

Un exemple de représentation par structure est donné dans le **Tableau 2.1**.

Contrairement aux réseaux sémantiques, cette représentation permet une meilleure structure des concepts. Toutefois d'une part, elle ne permet pas de fixer des standards pour les propriétés et d'autre part, la procédure d'inférence est difficile [15].

Propriété	Valeur	Type de valeur
Type de	humain	Référence au concept parent
Anniversaire	09/04/1986	Valeur de l'instance
Age	$\text{Calcul}(\text{dateActuelle} - \text{dateAnniversaire})$	Procédure
Genre	femme	Valeur de l'instance
Nombre de jambes	2	Valeur héritée du concept parent

Tableau 2.1 - Exemple de représentation basée sur la structure [14]

2.1.3. Représentation basée sur des règles

Cette représentation permet l'utilisation de conséquences. Les règles sont de la forme « Si... Alors... ». À partir de règles peuvent découler d'autres informations [15].

Par exemple, les règles suivantes [16] :

- 1- **Si** quelqu'un est un élève diplômé **Alors** il est aussi un élève,
- 2- **Si** un professeur donne un cours à un élève **Alors** le professeur enseigne à cet élève,
- 3- **Fait** : John est un professeur qui donne un cours auquel l'élève diplômé Ben assiste.

De ces deux règles et de ce fait découle que « John enseigne à Ben ».

Les règles constituent un élément de connaissances indépendant, modulaire et relativement petit, mais elles ne permettent pas de dériver des déclarations plus complexes et générales sur le domaine qui ne sont pas de la forme d'une règle. De plus, en cas de négation, des conflits sont possibles entre les règles.

2.1.4. Représentation basée sur la logique

À partir de cette représentation, le domaine peut être décrit comme étant constitué d'objets et, il est possible de construire des formules logiques autour de ces objets formés par des prédicats, des fonctions, des variables et des connecteurs logiques [16].

Par exemple, la relation entre un professeur et une personne décrite dans la **Figure 2.1** peut être représentée de la façon suivante : $\forall x : (\text{Professeur}(x) \rightarrow \text{Personne}(x))$.

La logique est capitale pour une bonne représentation des connaissances, elle a la capacité de capturer toutes sortes d'informations et de déduire des informations complexes et pertinentes à partir de la base de connaissances [15].

2.2. Modèles conceptuels

Les modèles conceptuels permettent de représenter les connaissances. Ce sont des objets d'information qui s'appuient sur un langage de modélisation pour décrire n'importe quel aspect d'un domaine d'intérêt à des fins de compréhension et de communication [21].

Différents modèles conceptuels existent pour représenter les connaissances. Selon les modèles, les formalismes présentés ci-dessus sont variables. Nous présentons ci-dessous trois types de modèles conceptuels : les diagrammes de classe, les graphes conceptuels et les ontologies.

Les diagrammes de classes permettent de construire un modèle conceptuel d'un domaine. Les diagrammes de classes en UML (Unified Modeling Language) sont les plus fondamentaux et les plus largement utilisés [22]. Ils permettent de représenter des concepts ainsi que les relations entre les concepts. L'inconvénient des diagrammes de classe est le caractère statique de la première couche (c'est-à-dire de la structure abstraite du domaine d'application) [23] et le manque de spécification formelle basée sur la logique [24]. Par exemple, il est possible de spécifier qu'un coq est un animal qui est aussi un mâle, alors qu'il n'est pas possible de spécifier qu'une personne est un parent si et seulement si elle a au moins un enfant.

Les graphes conceptuels sont des représentations graphiques pour la logique basée sur les réseaux sémantiques [24]. Ils fournissent un formalisme pour structurer et représenter des éléments de connaissance à partir de concepts et leurs relations, d'assertions, de règles et de requêtes [25]. Leur principal intérêt est la visualisation. Ils sont basés sur la logique du premier ordre [26].

Les ontologies définissent un vocabulaire commun de terminologie scientifique [27]. Elles sont utilisées pour faciliter l'utilisation et le partage des données grâce à des concepts et relations dans un domaine d'intérêt, ainsi que des règles sur la manière de combiner ces termes et relations [28]. Le langage d'ontologie OWL (Web Ontology Web Language) permet de décrire les classes, les attributs et les relations entre les objets d'une manière que les machines peuvent comprendre [29] et de mener un raisonnement. Il est le plus couramment utilisé pour développer une ontologie [30]. Par exemple, en considérant les deux déclarations, « chaque père est un homme » et « Alex est un père », alors un raisonneur pris en charge par OWL peut en déduire qu'« Alex est un homme » [31].

2.3. Conclusion de la section sur la représentation des connaissances

Les ontologies apparaissent comme le modèle le plus complet pour représenter les connaissances.

La représentation d'un modèle conceptuel par exemple sous forme d'un diagramme de classes en UML est possible dans un premier temps pour représenter les concepts et leurs relations. Toutefois, les ontologies permettent d'ajouter des contraintes sur les concepts ou sur leurs relations et d'effectuer une vérification de la cohérence du modèle construit.

Les ontologies, dans notre cas d'étude, permettent d'offrir une solution pour représenter les connaissances sur les dispositifs numériques utilisés pour l'orientation scolaire, ce qui répond en partie à la question :

QR1

Comment modéliser et stocker les caractéristiques des dispositifs numériques d'aide à l'orientation tout en étant robuste à l'apparition de nouveaux dispositifs ?

L'ajout de contraintes possible avec les ontologies permet, par exemple, si la classe « OutilDeCommunication » a une propriété d'objet « aInterLocuteur » avec la classe « InterLocuteur », nous pouvons contraindre les instances de « Outil de communication » à avoir un unique interlocuteur et le vérifier.

Il est nécessaire dans la suite de ce travail, pour répondre à cette question de recherche, d'examiner les ontologies existantes dans le domaine des dispositifs numériques dédiés à l'orientation scolaire, c'est ce qui est réalisé dans la Section 2.3 du Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ». En effet, un atout des ontologies est la possibilité de les réutiliser intégralement ou en partie. Et, s'il n'existe pas d'ontologie répondant à cette problématique, un examen de la façon de concevoir une ontologie doit être réalisé.

Une fois les dispositifs numériques représentés de façon formelle, nous cherchons à mesurer leur impact afin de fournir aux professionnels de l'orientation une valeur de comparaison entre les dispositifs. La mesure de l'impact des dispositifs numériques est présentée dans la section suivante.

3. Mesurer l'impact de dispositifs numériques

Dans cette section, nous fournissons une définition de l'impact, une revue des différents indicateurs d'évaluation des dispositifs numériques et des approches pour déterminer les corrélations éventuelles entre l'impact et les indicateurs afin de pallier à la limite (L2).

3.1. Définition de l'impact

Une grande variété d'études existe pour mesurer l'impact de dispositifs numériques dans le domaine de l'éducation. Cela met en avant l'importance de la spécification de plusieurs éléments dans la définition du cadre dans lequel l'impact est mesuré.

Il est nécessaire de spécifier :

- (1) « Sur quoi » porte l'impact : il peut s'agir de l'impact sur le succès académique, sur les performances en mathématiques, sur les qualités d'écriture [32], sur le bien-être [33],
- (2) « Sur qui » porte l'impact : les personnes touchées par l'impact peuvent être des adolescents [33], des enseignants [34], des étudiants [35],
- (3) L'impact « de quoi » : cela peut être l'impact des technologies numériques sans distinction [33], des technologies numériques utilisées dans l'enseignement postbac [35], de jeux ou encore des ordinateurs [32],
- (4) « Comment » mesurer l'impact : l'outil utilisé pour mesurer l'impact, par exemple par un score de qualité d'écriture [36] ou à l'aide d'une échelle de Likert [35].

L'utilisation du terme impact est de plus en plus utilisée depuis les années 2000 [37], mais la littérature n'apporte pas de définition claire de ce qu'il constitue [38]. Sa définition est dépendante du contexte. Nous considérons la définition suivante de l'impact : une impression ou un effet laissé sur quelqu'un suite à une action ou un événement [39] tout en prenant en compte que cette définition nécessite des spécifications (« sur quoi », « sur qui », « de quoi » et « comment »).

Nous cherchons ici à savoir si l'utilisation d'un dispositif numérique d'aide à l'orientation permet au lycéen de démarrer ou poursuivre sa démarche d'orientation.

Le modèle « Knowledge-Attitude-Behaviour » (KAB) divise l'impact en trois processus : l'acquisition de connaissances, la création de croyances et la formation de comportement [40].

Le modèle d'impact KAB est utilisé dans de multiples domaines, par exemple en médecine sur l'impact de la formation des patients sur la prise de médicaments [40], ou dans le domaine de la prévention pour les emplois à risque [41].

Nous détaillons ci-dessous les trois éléments de ce modèle :

- « **Knowledge** ». L'élément de *connaissance* permet de témoigner de l'acquisition des connaissances. Le gain de connaissances est un élément important à évaluer. Toutefois, seul, il n'est pas suffisant. Il existe de nombreux contextes d'enseignement dans lesquels les connaissances ne représentent qu'une partie des objectifs d'apprentissage [42].

Dans le contexte de l'orientation, les connaissances ne sont qu'une partie de l'objectif qui est de rendre proactif le lycéen dans sa démarche d'orientation.

Les connaissances peuvent être mesurées à l'aide d'un test de connaissances [40], [43] ou d'un questionnaire utilisant une échelle de Likert [41].

- « **Attitude** ». L'élément d'*opinion* d'un utilisateur comprend l'ensemble des sentiments [44] à l'égard du sujet concerné par l'évaluation. Cette opinion exerce une influence sur la réponse de l'utilisateur à des situations [43]. L'opinion des personnes est mesurée à l'aide d'une échelle de Likert [41].
- « **Behaviour** ». Un *comportement* est défini comme une action observable. Il regroupe les activités et actions favorisées par ses connaissances et son opinion. Il peut être mesuré par une échelle de Likert [41] ou d'autres questions posées aux personnes concernées par l'étude [45].

Une limite de l'évaluation du modèle KAB à l'aide de questionnaires auprès des utilisateurs est la différence potentielle entre ce qui est autodéclaré par l'utilisateur et ce qui est réel [46].

En effet, le gain de connaissances et le changement peuvent être difficiles à mesurer, d'autant plus, auprès d'un public de lycéens.

L'impact des dispositifs numériques sous forme d'autres modélisations existe. Ainsi, l'impact peut être défini uniquement par l'adoption du dispositif évalué [47]. L'impact des dispositifs numériques est également souvent considéré comme étant l'efficacité des utilisateurs pour leur travail [48], [49] ou leur efficacité dans le cadre de leur apprentissage [50]. Dans le cadre du e-learning qui est le plus proche de notre contexte, l'efficacité est définie comme le "résultat de l'apprentissage", c'est-à-dire l'acquisition par les utilisateurs de nouvelles connaissances grâce à l'initiative de l'apprentissage en ligne, et la capacité de transfert (passer de l'application à la pratique) [51].

Dans notre étude, notre travail n'a pas pour but de savoir si les lycéens adoptent le dispositif utilisé. Notre intérêt porte sur l'efficacité du lycéen dans son processus d'orientation induite par l'utilisation d'un dispositif numérique. Toutefois, nous cherchons uniquement à savoir si cela a permis au lycéen de démarrer ou de poursuivre sa démarche d'orientation de façon proactive et non s'il a appris un grand nombre de nouvelles informations ou s'il arrive à appliquer ce qu'il a appris.

Cependant, nous pensons que la modélisation KAB de l'impact est intéressante dans notre contexte, car les éléments de connaissance, d'opinion et de comportement peuvent être utilisés comme un modèle multidimensionnel de la construction du projet professionnel des adolescents [52]. De plus, cette modélisation est plus adaptée aux informations que souhaitent obtenir les professionnels de l'éducation.

À partir de l'état de l'art sur la définition de l'impact réalisé dans cette section, il est nécessaire d'analyser les indicateurs d'évaluation des dispositifs numériques utilisés dans la littérature.

3.2. État de l'art des indicateurs d'évaluation des dispositifs numériques

Un indicateur est une valeur mesurée pour fournir des informations sur un phénomène ou un état à un moment donné [53]. Les indicateurs doivent être compréhensibles, disponibles et pertinents [54].

La définition des indicateurs dépend du contexte. Par exemple, pour évaluer et comparer des interfaces de test scolaire en ligne [5], une évaluation selon trois dimensions est réalisée : une dimension éducative, une dimension économique et une dimension technique. Chacune de ces dimensions contient plusieurs indicateurs. En particulier, la dimension éducative est évaluée avec des indicateurs de contenu, de présentation, d'enchaînement des questions et de commentaires suite aux réponses. Certaines évaluations s'intéressent uniquement aux utilisateurs, par exemple, pour mesurer la volonté d'acheter un produit numérique [55]. Dans ce cas, les mesures sont réalisées en interrogeant les utilisateurs en utilisant des indicateurs démographiques, des indicateurs mesurant l'innovation et la valeur qu'ils perçoivent du produit, des indicateurs pour connaître leur rapport aux innovations et des indicateurs pour déterminer leur intention d'acheter le produit.

La capacité d'adaptation de l'utilisateur à des dispositifs numériques a aussi été discutée, par Jahanmir et al. [56] et Benson [6]. Il s'agit de l'attitude initiale de l'utilisateur par rapport à la technologie. Pour Benson, la capacité d'adaptation est l'un des aspects permettant d'expliquer le taux d'adoption d'une innovation pour un groupe d'individus. Les recherches

réalisées par Jahanmir et al. montrent que les adoptants tardifs d'un dispositif se voient comme des personnes moins innovantes.

Lessiter Jane et al. proposent également un modèle pour évaluer les dispositifs numériques, et plus particulièrement des dispositifs de réalité virtuelle [57]. Leur but est de mesurer l'intérêt éventuel des utilisateurs, ainsi que l'impact culturel et commercial de ce type de dispositif. L'évaluation est réalisée en trois étapes. Des questions sont posées aux utilisateurs avant l'utilisation, juste après, puis quatorze ou vingt-et-un jours après. Des indicateurs communs reviennent par rapport à ce qui a été présenté précédemment, comme le comportement des utilisateurs face à la technologie, mais d'autres sont nouveaux comme le souhait pour l'utilisateur de renouveler l'expérience de l'utilisation de la réalité virtuelle. De nombreux indicateurs sont présents dans cette étude. Les auteurs font un découpage des indicateurs selon neuf catégories : la qualité de l'expérience du public, l'attitude du public face au contenu et à la technologie, les impacts prévus par les créateurs, les caractéristiques des utilisateurs, les possibilités d'interaction, l'impact économique, le comportement des utilisateurs, les traits personnels des utilisateurs et l'impact sur les utilisateurs ciblés.

Dans le domaine de la santé, la question de l'évaluation des dispositifs numériques est aussi posée, notamment sur les façons et les raisons de la diffusion de dispositifs numériques comme nouveaux modèles de soins. Par exemple, un coach numérique pour l'auto-management du diabète par les patients est évalué selon cinq catégories d'indicateurs [6] :

- La préparation à l'innovation : précocité de l'utilisateur à adopter de nouvelles idées par rapport aux autres,
- La confiance numérique : utilisation personnelle et de l'entourage d'appareils numériques,
- L'adoption de l'innovation : travail du personnel médical, avant, pendant et après l'usage de l'outil,
- La satisfaction de l'utilisateur : l'utilité de l'outil, sa facilité d'utilisation et la satisfaction globale,
- Le changement de comportement : capacité (par exemple physique), opportunité et motivation à utiliser le dispositif pour l'utilisateur.

Ainsi, nous venons d'identifier un ensemble de valeurs mesurables utilisées pour évaluer les dispositifs numériques dépendant de l'utilisateur, de ce qu'il perçoit, mais aussi selon les caractéristiques du dispositif visibles par l'utilisateur. Cette identification met aussi en avant la nécessité d'affiner les indicateurs selon le contexte d'évaluation et selon le type de dispositifs.

Les indicateurs d'évaluation des dispositifs numériques peuvent être regroupés selon la thématique qu'ils mesurent. Dans la littérature sur l'aide à la décision, ces regroupements d'indicateurs sont appelés des catégories (ou des dimensions).

Certaines catégories sont utilisées dans plusieurs travaux de recherche comme les caractéristiques de l'utilisateur [5], [55], [57]. D'autres catégories, comme le comportement futur des utilisateurs [57] sont moins utilisées malgré leur potentiel intérêt. Par exemple, l'analyse du comportement futur permet de capter des informations dont l'utilisateur n'a pas conscience à court terme, ou de déterminer ce dont il va se souvenir dans le temps.

Afin de déterminer les catégories pertinentes pour un contexte d'évaluation de dispositifs numérique donné, il est nécessaire d'avoir une vision globale de l'ensemble de ces catégories.

À notre connaissance, il n'existe pas de synthèse présentant les indicateurs d'évaluation des dispositifs numériques. Nous en proposons une dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ».

Dans les deux sections ci-dessus, nous avons présenté les travaux traitant de la définition de l'impact et des indicateurs utilisés pour évaluer les dispositifs numériques. Dans la section suivante, nous réalisons un état de l'art pour analyser la façon de lier ces deux éléments.

3.3. Travaux sur la recherche de corrélations entre des variables

Nous nous plaçons dans un problème de recherche de corrélation avec plusieurs variables. Dans notre contexte, nous cherchons à identifier la corrélation entre l'impact (*Impact*), et un ensemble de catégories (c_1, \dots, c_m). Cet ensemble de catégories regroupe des variables (indicateurs) pour évaluer les dispositifs numériques.

Ce problème est schématisé dans la **Figure 2.2** ; l'ensemble $\{p_1, \dots, p_m\}$ représente les coefficients de corrélation de chacune des m catégories.

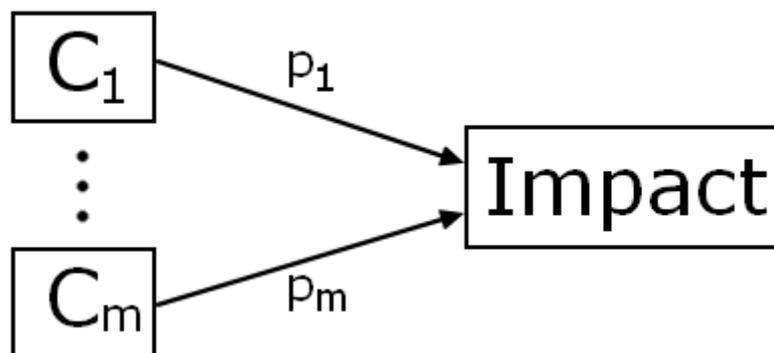


Figure 2.2 - Représentation schématique de notre recherche de coefficients de corrélation

Nous présentons dans cette section différentes méthodes d'analyse de corrélation entre une variable dépendante et un ensemble de variables indépendantes.

3.3.1. Régression Linéaire Multiple

À partir de travaux réalisés sur la corrélation dans le domaine de dispositifs numériques et l'éducation [58], nous posons l'hypothèse que la dépendance entre les variables est linéaire pour pouvoir réaliser une régression linéaire multiple et une modélisation par équation structurelle.

Il s'agit d'une méthode permettant d'estimer les coefficients de régression et d'évaluer l'importance des variables indépendantes pour estimer la variable dépendante.

Le problème peut être écrit de la façon suivante [59] en utilisant les notations associées à notre problème ci-dessus :

$$\text{(EQ1)} \quad \underbrace{\begin{pmatrix} impact_1 \\ \vdots \\ impact_i \\ \vdots \\ impact_n \end{pmatrix}}_{Impact} = \underbrace{\begin{pmatrix} 1 & c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots & \vdots \\ 1 & c_{i1} & \vdots & c_{im} \\ \vdots & \ddots & \vdots & \vdots \\ 1 & c_{n1} & \cdots & c_{nm} \end{pmatrix}}_C \underbrace{\begin{pmatrix} p_0 \\ \vdots \\ p_i \\ \vdots \\ p_m \end{pmatrix}}_P + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}}_\varepsilon$$

n observations sont obtenues à partir des questionnaires de n répondants : $(c_{i1}, \dots, c_{ip}, impact_i)$;

P : vecteur des coefficients de régression, p_0 est la valeur de l'impact lorsque toutes les valeurs des catégories sont à zéro, son vecteur estimé associé est noté \hat{P} ;

ε : vecteur de bruit (erreurs de mesure) ;

La méthode de régression linéaire consiste à estimer les coefficients du vecteur P .

Une méthode fréquente pour estimer ces coefficients est celle des moindres carrés [60], elle cherche à minimiser la distance au carré entre les variables observées et prédites. Le résultat de cette estimation est :

$$\hat{P} = (C^T C)^{-1} C^T Impact$$

La régression linéaire multiple est utilisée dans le milieu de l'éducation à plusieurs reprises pour étudier la performance des étudiants [61], [62], [58]. En cherchant par exemple les liens entre la performance scolaire générale des étudiants et leurs performances en mathématiques [63].

La régression linéaire multiple se place comme une méthode possible dans notre problème pour identifier un poids pour chacune des catégories d'évaluation. Elle permet également d'estimer la valeur de l'impact à partir de mesures faites sur les catégories d'évaluation.

Il ne s'agit pas de la seule méthode permettant cela. Nous présentons une autre méthode utilisée dans la littérature qui est la modélisation par équation structurelle.

3.3.2. Modélisation par équation structurelle

Les modèles à équation structurelle (SEM) permettent d'estimer les relations causales entre plusieurs variables observées et non observées et le poids de ces relations. Par rapport aux analyses basées sur des régressions, les SEMs permettent d'analyser un modèle dans son ensemble et pas uniquement l'influence des variables sur une autre variable [64].

Les modèles basés sur les SEMs sont représentables sous forme de graphiques. Ils utilisent deux types de variables [65] :

- Les variables manifestes/observées : il s'agit de variables observables ou mesurables, elles sont représentées par des rectangles.
- Les variables latentes : elles ne sont pas directement mesurées, mais, évaluées indirectement à l'aide de variables manifestes, elles sont représentées par des ovales.

Les relations entre les variables sont indiquées par des flèches. L'absence de flèche reliant les variables signifie qu'aucune relation directe n'a été supposée. Une flèche représente une relation directe hypothétique entre deux variables, et la variable sur laquelle pointe la flèche est la variable dépendante. Un exemple est donné dans la **Figure 2.3**.

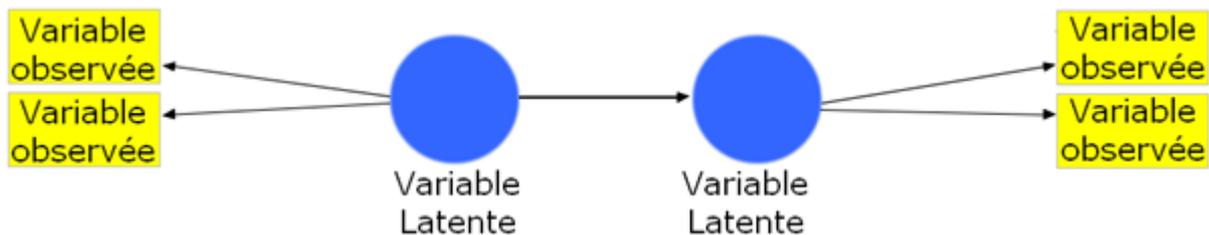


Figure 2.3 - Schéma illustrant la représentation des variables et des relations pour les modèles à équation structurelle

L'estimation des paramètres du modèle peut se faire par différentes approches. L'approche PLS (partial least squares path modeling) est privilégiée lorsque l'objet de l'étude ne dispose pas d'une base théorique bien développée, en particulier lorsqu'il existe peu de connaissances préalables sur les relations de cause à effet [66].

Dans notre problème, il n'existe pas de connaissances préalables sur les relations de cause à effet. Cette approche peut être intéressante pour notre étude puisque nous sommes dans une perspective exploratoire avec une population de petite taille.

PLS-SEM est un algorithme itératif se basant sur la variance entre les données [67]. Dans la littérature, il est utilisé, comme pour la régression linéaire multiple, pour étudier la

performance des étudiants en analysant les liens entre les émotions négatives et positives, les compétences scolaires, le stress, l'état d'esprit et la réussite académique [68]. Cette méthode est également utilisée à plusieurs reprises pour analyser la satisfaction des étudiants, leur satisfaction générale dans l'enseignement postbac [69], leur satisfaction concernant la mise en œuvre d'examens en ligne [70]. Le design du cours, la qualité de l'enseignement, les interactions entre les pairs et la qualité des technologies sont également pris en considération dans une approche PLS-SEM pour estimer la satisfaction des étudiants dans leur expérience d'apprentissage [71].

Cette approche permet d'estimer des relations causales entre les variables [72]. Dans la section suivante, nous discutons des réseaux bayésiens qui permettent de modéliser des relations causales entre les variables à partir de probabilités conditionnelles.

3.3.3. Réseaux Bayésiens

Les probabilités conditionnelles sont utilisées pour modéliser le lien entre les variables dans le cadre des réseaux bayésiens, contrairement aux deux approches présentées ci-dessus qui sont basées sur des coefficients de corrélation. Ces probabilités conditionnelles sont construites à partir d'un tableau de probabilités conditionnelles déduit des données disponibles pour le problème.

Les réseaux bayésiens sont des graphes acycliques (ils ne contiennent pas de boucle) dans lesquels chaque nœud du graphe représente une variable aléatoire et les chemins entre ces nœuds représentent les relations causales et conditionnelles entre les variables correspondantes [73]. Les probabilités conditionnelles sont données par des experts ou apprises lorsque la base de données est suffisamment importante.

Dans le domaine de l'Éducation, les réseaux bayésiens sont utilisés, par exemple, pour analyser la probabilité de pérennisation des étudiants dans les programmes de cours à distance [74], ou pour construire des évaluations à destination des étudiants [75].

3.3.4. Bilan des trois approches

Les trois approches présentées ci-dessus (régression linéaire multiple, modélisation par équation structurelle, réseau bayésien) permettent de réaliser des prédictions sur une valeur ou un état final à partir d'observations, le choix d'en utiliser une par rapport aux autres est déterminé par les données et par les liens entre les variables. Il n'est actuellement pas possible d'en retenir une plus que les autres. Ainsi, dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation », nous utilisons et comparons ces trois approches dans notre contexte.

Les approches de régression linéaire multiple, de modélisation par équation structurelle, et de réseau bayésien ont pour but, dans notre contexte d'estimer la valeur de l'impact à partir d'un ensemble d'indicateurs sans avoir besoin de mesurer l'impact directement.

3.4. Conclusion de la section sur la mesure de l'impact

Dans cette Section 3, nous avons présenté la notion d'impact, et plus particulièrement le modèle KAB sur lequel nous nous sommes basés. Ce modèle structure l'impact comme l'acquisition de connaissances, la création de croyances et la formation de comportement. Il s'agit d'un cadre pour définir l'impact, mais ce modèle n'est pas une solution pour mesurer l'impact.

Une présentation des éléments couramment mesurés lors de l'évaluation de dispositif numérique a été effectuée. Un grand nombre d'indicateurs peuvent être utilisés. Ils dépendent principalement du contexte et de l'objectif de modélisation. Toutefois, à notre connaissance, il n'existe pas de synthèse répertoriant les catégories d'indicateurs possible. Cela rend l'identification des indicateurs cohérents pour notre problème complexe. Une telle synthèse est proposée dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ».

L'impact est difficile à mesurer à cause de la différence potentielle entre ce qui est autodéclaré par les lycéens et ce qui est réel. Ainsi, nous souhaitons mettre en relation le modèle KAB avec les indicateurs d'évaluation des dispositifs numériques.

Afin de lier les indicateurs d'évaluation à l'impact, nous avons présenté trois approches de recherche de relations entre plusieurs variables : la régression linéaire multiple, la modélisation par équation structurelle et les réseaux bayésiens. Ce qui est présent dans la littérature ne permet pas de sélectionner une de ces approches. L'utilisation de l'une d'elle en particulier dépend des données et nécessite d'expérimenter ces méthodes sur nos données. Ainsi, nous conservons ces trois approches, une comparaison des approches est faite à partir des données dans le Chapitre « Expérimentations ».

La valeur d'impact n'a pas pour objectif d'être utilisée dans l'absolu. En effet, dire qu'un dispositif numérique a un impact de quatre sur sept n'est pas interprétable. La valeur d'impact est utile pour comparer plusieurs dispositifs. Le modèle KAB, quant à lui, permet d'ouvrir la discussion entre les professionnels de l'orientation.

4. Prédiction de l'impact

Dans cette section, nous nous plaçons dans une situation où l'impact n'est pas mesurable comme nous l'avons présenté précédemment, par exemple si le dispositif n'a pas encore été rendu accessible aux lycéens.

Les tâches de prédiction avancées comme celles de l'impact sur le processus d'orientation des lycéens sont des tâches complexes pour lesquelles il est difficile pour un humain de créer des algorithmes à partir de règles. L'intelligence artificielle (IA) a l'avantage d'automatiser le processus de prédiction. La prédiction se fait en utilisant des algorithmes qui traitent les données disponibles du domaine pour identifier des modèles dans ces données [2] et ainsi utiliser le modèle identifié pour prédire de nouvelles données. La prédiction, selon l'algorithme utilisé, peut être expliquée.

L'explicabilité est importante dans notre cas pour apporter des explications sur les résultats obtenus aux professionnels de l'orientation. Dans cette section, nous étudions les approches relatives à l'intelligence artificielle permettant de prédire un impact sur le processus d'orientation des lycéens tel qu'il a été défini dans le chapitre précédent.

L'Intelligence Artificielle (IA) est un vaste domaine avec de nombreuses applications comme l'acquisition, le traitement, l'interprétation [76] de données. Nous nous basons sur la structure de l'IA présentée par Bonjour et al. qui distinguent quatre grandes classes de techniques d'IA [77] : (1) les approches type recherche opérationnelle ; (2) les approches de modélisation des connaissances et raisonnement ; (3) les approches orientées données ; (4) les thèmes transverses. Plus précisément, nous nous concentrons sur les deux catégories d'approches pertinentes pour notre problème, car ces deux types d'approches permettent d'estimer des valeurs tout en étant explicables :

1. Les approches de modélisation des connaissances et raisonnement, s'intéressent à la façon dont un agent utilise ses connaissances pour prendre des décisions [78].
2. Les approches orientées données utilisent des bases de données enregistrées pour servir de référence et guider la conception d'un modèle [79].

Nous présentons brièvement ci-dessous les approches pertinentes pour notre problème dans ces deux catégories.

4.1. Approches de modélisation des connaissances et raisonnement

Généralement, ces approches se composent d'une base de connaissances disponible et exploitable, d'une base de faits caractérisant le problème et d'un programme recherchant et exploitant les données et les faits afin de raisonner et de résoudre le problème [80].

4.1.1. Représentation et raisonnement sur des connaissances à base logique

Ce type d'approche englobe les ontologies et les approches logiques. Elles s'intéressent à la façon de représenter les connaissances et de les manipuler automatiquement pour raisonner [78]. Le raisonnement se fait généralement grâce à des règles ; dans ce cas, il est nécessaire d'avoir un ensemble de règles, un ensemble de faits et un moteur d'inférence. Les règles peuvent être obtenues à l'aide d'experts.

Par exemple, dans notre contexte, ces connaissances prendraient la forme de liens entre les caractéristiques du dispositif (son type d'informations disponibles, ses outils) et son impact. Toutefois, les experts de l'orientation n'ont pas connaissance de tels liens.

4.1.2. Raisonnement à partir de cas

Le Raisonnement à Partir de Cas (RàPC) est un type de raisonnement analogique [81] qui utilise les situations similaires précédentes [82] ; il est utilisé dans différents domaines comme la médecine [83] mais aussi pour le design, les applications dans le droit, le diagnostic technique [84].

La solution à un nouveau problème est obtenue en cherchant les problèmes résolus similaires et en utilisant les solutions de ces problèmes. Un problème est représenté par un cas, le cas est quant à lui composé d'un couple (Description du problème ; Solution). Les cas sont indépendants entre eux. Cette méthode de résolution de problèmes permet d'enregistrer les problèmes résolus au fur et à mesure afin de gagner en connaissances [85]. Les problèmes résolus et stockés en mémoire sont appelés les cas sources.

Dans la **Figure 2.4**, nous présentons le fonctionnement classique et le plus fréquent du raisonnement à partir de cas en prenant l'exemple de la recherche de solution pour le Cas 4.

Dans notre étude, la description du problème pourrait correspondre aux caractéristiques connues des dispositifs numériques (données structurées avec une ontologie), et la solution d'un cas pourrait représenter la valeur de l'impact du dispositif numérique associé.

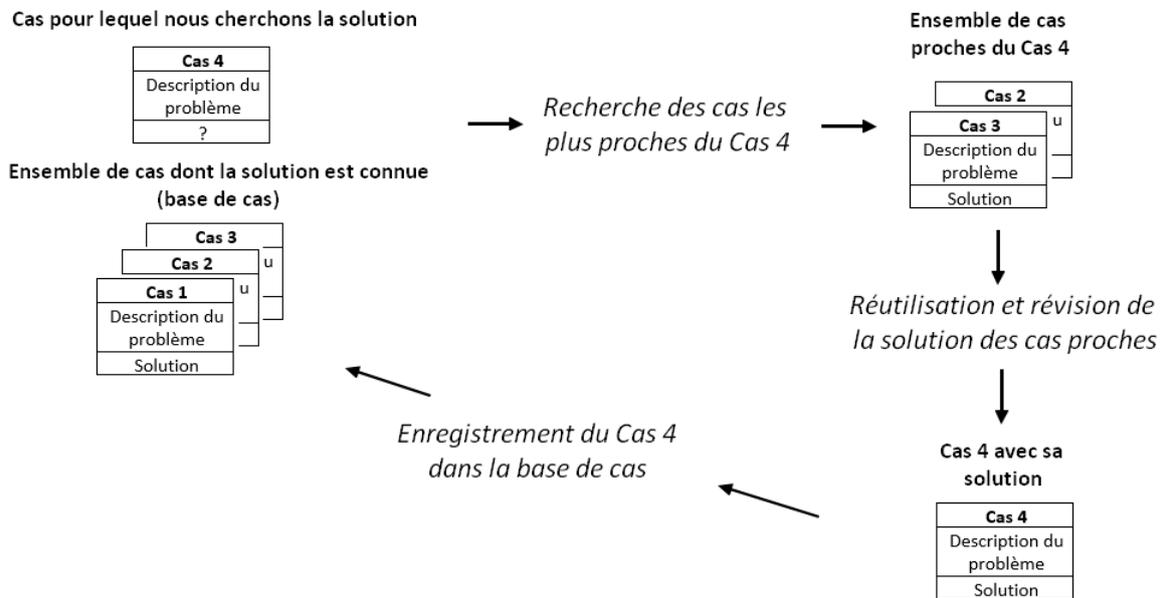


Figure 2.4 - Principe de fonctionnement du RàPC [83]

4.1.3. Représentation et raisonnement sur des connaissances incertaines (réseaux bayésiens)

Les réseaux bayésiens font partie des techniques d'intelligence artificielle. Ils sont adaptés à la collecte et à la représentation des connaissances dans les domaines incertains. C'est un outil de raisonnement dans l'incertitude. Une de ses applications est de mettre à jour des croyances [86].

Les dépendances probabilistes dans les réseaux bayésiens sont le plus souvent obtenues en utilisant des méthodes statistiques. Les réseaux bayésiens sont utiles pour apprendre les relations causales et prédire les effets futurs [87].

Les probabilités conditionnelles sont l'une des bases pour construire les réseaux bayésiens [88]. Le théorème de Bayes, pour calculer les probabilités conditionnelles est énoncé de la façon suivante [89] :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Où A et B sont des événements avec plusieurs états possibles,

$P(A)$ (respectivement $P(B)$) est la probabilité de A (respectivement B),

la probabilité de A sachant B est notée $P(A|B)$,

il est nécessaire d'avoir $P(B) > 0$.

Pour mieux comprendre le fonctionnement des réseaux bayésien, un exemple concret est présenté dans la **Figure 2.5**. Considérons des dispositifs numériques d'aide à l'orientation caractérisés par la présence de témoignages et d'informations sur la vie étudiante.

Chaque nœud représente une variable binaire qui prend la valeur vrai (V) ou faux (F). Dans cet exemple, le nœud $X1$ est la présence de témoignages ou non, le nœud $X2$ est la présence ou l'absence d'informations sur la vie étudiante. Le nœud $X3$ est la valeur de l'impact du dispositif numérique sur le processus d'orientation qui peut être « Faible » ou « Bon ». Les nœuds $X1$ et $X2$ sont les nœuds parents de $X3$. Le lien entre $X1$ et $X3$ et celui entre $X2$ et $X3$ montrent que l'état de $X1$ et celui de $X2$ modifient l'état de $X3$ selon les probabilités conditionnelles.

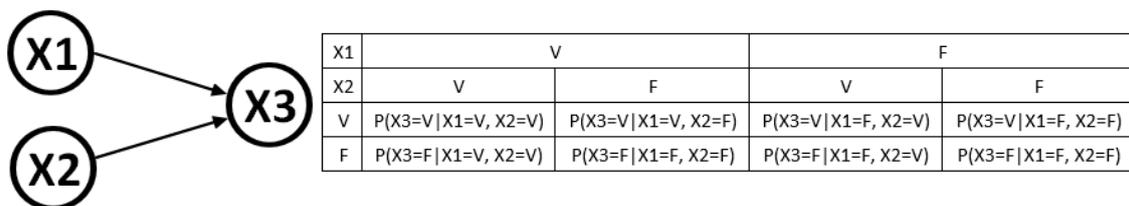


Figure 2.5 - Exemple de réseau bayésien avec la table de probabilité

Lorsqu'une information est donnée pour l'un ou les deux nœuds $X1$ et $X2$, la probabilité d'avoir un impact « Faible » ou « Bon » peut être obtenue à partir du tableau de probabilité conditionnelle.

Les valeurs du tableau de probabilités conditionnelles sont données par des experts ou apprises avec un algorithme lorsqu'une base de données est disponible, avec un nombre de données suffisamment important [87].

4.2. Approches orientées données

L'apprentissage automatique est l'ensemble des approches permettant d'apprendre à partir de données sans être explicitement programmées à l'aide de règles [90]. Il se base sur différents types d'algorithmes. Le choix de l'algorithme dépend du problème à résoudre, du nombre et du type de variable [91]. Les types d'apprentissage automatique les plus courants sont brièvement décrits ci-dessous.

Selon la disponibilité et le type de données, l'apprentissage automatique peut être supervisé, non supervisé, semi-supervisé ou par renforcement [92].

4.2.1. Apprentissage supervisé

L'apprentissage supervisé apprend à partir d'exemples. Un exemple est constitué d'une paire entrée-sortie, où l'entrée sont les données observées et la sortie la variable cible. Un apprentissage est réalisé à partir de ces exemples afin de prédire la variables cible de nouvelles données. Les algorithmes d'apprentissage supervisé induisent des modèles à partir de ces données d'apprentissage et ces modèles peuvent être utilisés pour déterminer la variable cible de données pour laquelle elle n'est pas connue [93].

Il existe de nombreuses alternatives pour réaliser cet apprentissage, par exemple, les machines à vecteurs de support, les arbres de décision [94].

L'apprentissage supervisé est lui-même divisé en deux catégories : la classification et la régression, pour considérer le type d'annotation.

Dans les problèmes de classification, la variable cible possible est un ensemble de variables discrètes, il s'agit d'une classe cible ou d'une catégorie. Par exemple, l'impact du dispositif numérique peut être associé à l'annotation « Faible » ou « Bon ».

Dans les problèmes de régression, la variable cible est une valeur réelle continue. Par exemple, si l'impact du dispositif numérique peut prendre une valeur dans l'intervalle [1 ; 7].

4.2.2. Apprentissage non supervisé

L'apprentissage non supervisé a pour objectif de séparer les données en sous-groupes, ces sous-groupes sont construits de façon à ce que les données au sein d'un sous-groupe aient une structure similaire. Les données observées sont utilisées pour grouper les données, les données ne sont pas annotées [95].

4.2.3. Apprentissage semi-supervisé

L'apprentissage semi-supervisé est une combinaison de l'apprentissage supervisé et non supervisé [96]. L'apprentissage est réalisé avec des données dont la variable cible est connue et des données pour lesquelles elle n'est pas connue. Cet apprentissage est utile lorsque l'étiquetage de toutes les données est difficile et que le nombre de données est très important.

4.2.4. Apprentissage par renforcement

L'apprentissage par renforcement est un apprentissage itératif. Les agents logiciels agissent afin de maximiser leurs récompenses. Cet apprentissage se fait par essais et erreurs [97].

4.2.5. Positionnement par rapport aux approches orientées données

Dans notre contexte, dans le cadre des approches orientées données, nous avons un couple données observées (ensemble des caractéristiques des dispositifs numériques) et variable cible (valeur de l'impact). Ainsi, dans notre problème, l'apprentissage supervisé est le plus adapté, et, lorsqu'il est possible, il s'agit de l'approche la plus recommandée. Pour notre problème, la variable cible est la valeur d'impact défini dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation » comme une valeur réelle dans l'intervalle [1 ; 7], nous nous plaçons dans un problème de régression.

4.3. Synthèse des approches

Les approches de traitement des données et de prédiction présentée ci-dessus sont comparées dans le **Tableau 2.2**. Cette comparaison présente les avantages et les inconvénients de chacune des approches étudiées selon notre objectif de recherche (compte tenu de notre problématique). Pour cela, nous comparons le type de données disponibles (données sans informations d'experts, l'acquisition de données dans le temps, et la possibilité d'effectuer des prédictions à partir d'un faible nombre de données).

À chaque ligne du **Tableau 2.2** correspond la prise en compte d'un problème spécifique. Si l'approche considérée nous permet de prendre en compte ce problème, un signe + est mis dans la case correspondante.

Les approches orientées données sont puissantes dans de multiples domaines (en médecine, en industrie, ...) où un grand volume de données est disponible. Au sein de ces approches, une grande variété d'algorithmes existe (les forêts aléatoires, les séparateurs à vaste marge, ...). Ils dépendent du type de données et du besoin d'explicabilité. Ces approches ont besoin pour donner de bons résultats, de disposer d'un grand nombre de données.

Dans notre étude, les réseaux bayésiens qui sont basées sur des calculs statistiques posent un problème de disponibilité des données, similaire à celui des approches orientées objet.

L'approche sur des connaissances à base logique a l'avantage de ne pas nécessiter un grand nombre de données. Cette approche a besoin de connaissances d'experts, c'est-à-dire d'informations de personnes spécialisées dans le domaine considéré. Un exemple concret de

règle possible pour notre problème serait que la présence combinée d'un témoignage et d'un outil de test sur les métiers dans un dispositif augmenterait la valeur d'impact globale. Toutefois, nous ne disposons pas d'informations de ce type. En effet, les dispositifs numériques d'aide à l'orientation ont été peu étudiés à ce jour. De plus, il n'existe pas de dispositif « parfait » selon les lycéens et enseignants, ce qui rend l'impact d'un dispositif très dépendant de l'utilisateur.

Concernant le raisonnement à partir de cas, une base de cas avec peu de cas entraîne le risque que les nouveaux cas cibles arrivant aient peu de similarité avec les cas sources, ce qui nécessite une étape de révision importante de la solution.

L'avantage du raisonnement à partir de cas est la possibilité d'utiliser des données sans règles énoncées par des experts, l'ajout incrémental de données lorsque l'impact de nouveau dispositif est connu, et la possibilité de prédiction à partir d'un petit nombre de données.

	Approches de modélisation des connaissances et raisonnement			Approches orientées données
	Représentation et raisonnement sur des connaissances à base logique	Raisonnement à partir de cas	Réseaux bayésiens	
Utilisation de données sans règles données par un expert		+	+	+
Ajout incrémental de données	+	+	+	+
Prédiction à partir d'un faible nombre de données	+	+		

Tableau 2.2 - Comparaison des modèles pour le traitement des données et pour la prédiction d'un impact

4.4. Conclusion de la section sur la prédiction de l'impact

Nous avons présenté différentes approches permettant d'estimer un impact. Le RàPC est une approche permettant de travailler avec un nombre de données peu important. Les informations d'experts ne sont pas un impératif au fonctionnement du RàPC lors du cycle de prédiction. De plus, le RàPC offre la possibilité de gagner en connaissances en ajoutant des données au cours du temps. Les experts sont néanmoins nécessaires pour la création de cas et ils peuvent intervenir lors des étapes de révision et d'enregistrement d'un cas dans la base de cas.

Ainsi, le raisonnement à partir de cas pour estimer l'impact sur le processus d'orientation des lycéens semble être l'approche appropriée à notre problème. Nous présentons en détails le fonctionnement du RàPC dans le Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation » afin d'estimer de quelle façon le RàPC répond à la question (QR3).

- 3 -

**ODNAO : UNE
ONTOLOGIE POUR
REPRESENTER LES
DISPOSITIFS NUMERIQUES
D'AIDE A L'ORIENTATION**

TABLE DES MATIERES

1. Introduction.....	34
2. Dispositifs numériques dédiés à l'orientation scolaire des lycéens	35

2.1. Phases d'orientation vécues par les lycéens et les solutions numériques proposées..	36
2.2. Exemples de dispositifs numériques d'aide à l'orientation à destination des lycéens.	39
2.3. Ontologies relatives à l'orientation scolaire	41
2.4. Concevoir une ontologie.....	42
2.5. Synthèse et questions de recherche	43
3. Proposition d'une ontologie pour caractériser les dispositifs numériques d'aide à l'orientation.....	44
3.1. Méthodologie de conception de l'ontologie proposée.....	44
3.1.1. Étape 1 : Définir le domaine et la portée.....	45
3.1.2. Étape 2 : Identifier les ontologies existantes	45
3.1.3. Étape 3 : Lister les termes et les concepts.....	45
3.1.4. Étape 4 : Créer une hiérarchie de classes	46
3.1.5. Étapes 5 et 6 : Définir la structure interne des concepts	47
3.1.6. Étape 7 : Créer les instances.....	47
3.2. Ontologie des Dispositifs Numériques d'Aide à l'Orientation.....	47
3.2.1. Module Nature de l'information	49
3.2.2. Module interaction de l'utilisateur avec le dispositif numérique	52
3.2.3. Module Services additionnels.....	53
3.2.4. Module Utilisateur	55
3.2.5. Module d'Utilisabilité	57
3.2.6. Module d'Impact	58
3.3. Vérification et validation de l'ontologie proposée.....	60
4. Conclusion	61

1. Introduction

Dans le cadre du processus d'orientation scolaire, différents dispositifs dédiés soutiennent les lycéens dans ce cheminement. Nous nous concentrons ici sur leur étude et leur analyse. Le terme de dispositif numérique est considéré ici comme l'ensemble des sites web, des applications mobiles et des applications ordinateurs. Ces dispositifs n'ont pas tous les mêmes composants. Certains d'entre eux proposent un outil d'échanges (c'est-à-dire permettant aux lycéens de communiquer avec quelqu'un) tandis que d'autres disposent de tests d'orientation par exemple. Ces dispositifs diffèrent également en termes d'objectifs. Nous observons que certains d'entre eux visent à informer sur la procédure d'orientation, d'autres sont dédiés à répondre à des questions ou à donner des idées de métiers. L'objectif

des professionnels de l'orientation du projet AILES est de savoir quels sont les dispositifs les plus pertinents à déployer, et lorsqu'un dispositif est en cours de développement, de prédire s'il sera pertinent ou non.

Chaque dispositif numérique propose des informations organisées différemment. Les architectures hétérogènes de ces dispositifs impliquent le besoin d'une structure conceptuelle pour collecter leurs données. La pertinence d'un dispositif dépend également de l'utilisateur. Les entretiens que nous avons menés avec des professionnels de l'orientation nous ont amenés à supposer que la pertinence n'est pas la même selon l'année précédant le baccalauréat, le type de baccalauréat, et le moment de l'utilisation dans l'année. En effet, dans le système éducatif français, il existe trois types de baccalauréat : le général, le technologique et le professionnel. Chacun de ces diplômes se déroule sur trois ans : la Seconde, la Première et la Terminale [98].

Dans ce chapitre, notre objectif est de répondre à la question suivante :

QR1 :

Comment modéliser et stocker les caractéristiques des dispositifs numériques d'aide à l'orientation tout en étant robuste à l'apparition de nouveaux dispositifs ?

Il est donc nécessaire de proposer une caractérisation formelle des dispositifs numériques dédiés à l'orientation. Cette caractérisation formelle doit contenir des informations permettant de décrire et de distinguer les dispositifs numériques. Dans ce contexte, ce chapitre présente une approche pour cartographier les données sur les dispositifs numériques afin de les caractériser.

Le reste de ce chapitre est organisé de la façon suivante. La première section décrit les dispositifs numériques dédiés à l'orientation des lycéens. Dans la deuxième section, nous posons une question de recherche et soulignons la principale contribution de ce chapitre. La troisième section présente notre approche.

2. Dispositifs numériques dédiés à l'orientation scolaire des lycéens

Cette section présente une revue de la littérature concernant les systèmes numériques dédiés à l'orientation scolaire des lycéens. Tout d'abord, les phases de l'orientation sont présentées. Ensuite, des exemples de dispositifs numériques d'orientation pour les lycéens sont donnés. Enfin, trois ontologies conçues pour ce domaine sont expliquées.

2.1. Phases d'orientation vécues par les lycéens et les solutions numériques proposées

L'évolution des nouvelles technologies a modifié les méthodes d'enseignement, d'apprentissage et d'orientation scolaire. Elles ont notamment permis de personnaliser l'apprentissage et de le rendre accessible partout et à tout moment [2]. Ces nouvelles technologies sont liées à la vie des lycéens et des étudiants [99].

Elles sont utilisées par les enseignants comme support d'apprentissage dans leur matière [100] mais aussi par les conseillers d'orientation pour soutenir les projets de vie des jeunes [101]. Dans le cadre du parcours scolaire des lycéens, le processus d'orientation est également affecté par cette évolution.

Le processus d'orientation est souvent représenté en quatre phases [2], [102] illustrées dans la **Figure 3.1**. Ces quatre phases se déroulent dans les lycées pour amener progressivement les élèves à un choix unique.

La première phase est consacrée à une **exploration large**. Au cours de cette phase, les élèves sont encouragés à rechercher de multiples sources d'information et à explorer toutes les possibilités. L'exploration des possibilités peut consister en une exploration des études après le baccalauréat ou en une exploration des options au lycée, par exemple. Cela dépend du niveau d'étude de l'élève.

La deuxième phase a pour but d'**explorer des domaines spécifiques** qui pourraient mieux convenir à l'étudiant. Une analyse d'un ou plusieurs domaines rencontrés lors de l'exploration large est réalisé.

La troisième phase est la **crystallisation des choix** à partir de l'analyse menée au cours de la phase précédente. Au cours de cette phase, les choix sont affinés et certaines idées présentes à la deuxième phase sont éliminées. Ainsi il ne reste qu'un ensemble de choix, généralement ordonné selon les préférences de l'élève.

La quatrième phase est celle de l'**exécution**, au cours de laquelle le choix est mis en œuvre.

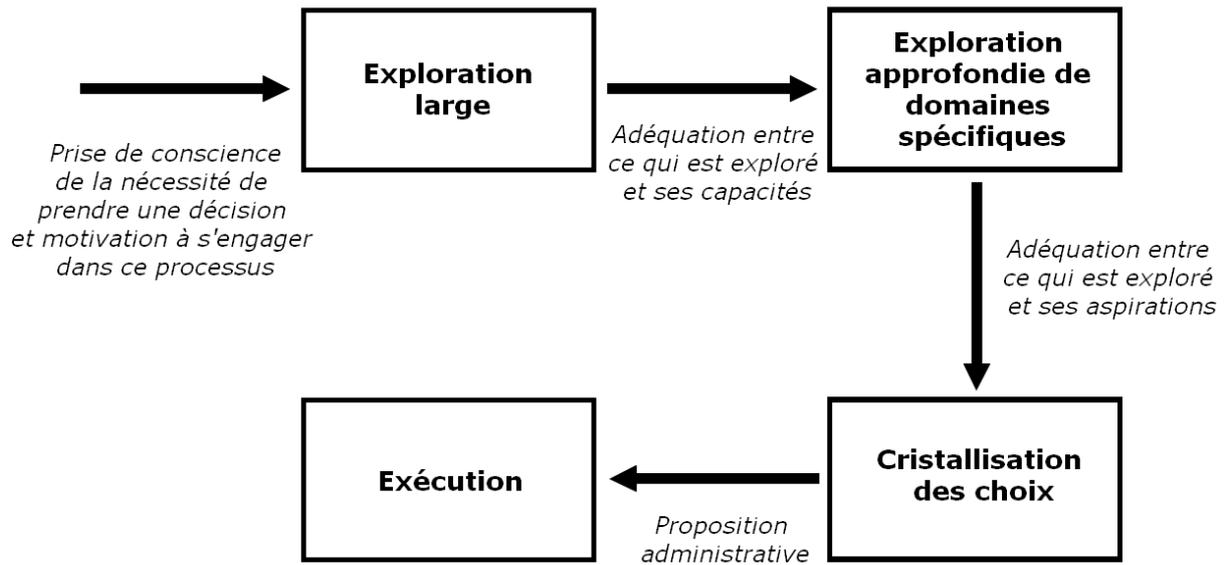


Figure 3.1 - Processus d'orientation divisé en quatre phases

Dans la littérature, des propositions de systèmes utilisant des dispositifs numériques ont été faites pour soutenir les étudiants dans leur processus d'orientation. Ce soutien est possible au cours des différentes phases de l'orientation scolaire.

Dans la phase d'**exploration large**, El Mrabet et al. proposent d'envoyer des données aux appareils électroniques du lycéen (téléphone, tablette...) en fonction de ses déplacements et des établissements géographiquement proches (par exemple une usine, une entreprise, ...) [2]. Les minorités lycéennes constituent un public particulier dans cette phase d'exploration, il s'agit de populations sous-représentées dans de nombreuses professions [3]. Greenidge souligne l'importance d'aider les minorités à se projeter dans des domaines où la croyance en leur réussite est faible [3]. L'auteure propose d'utiliser une approche de réalité virtuelle pour simuler l'environnement et les activités de diverses professions (par exemple les tâches et responsabilités quotidiennes d'un astronaute). Cette contextualisation professionnelle augmente les connaissances de l'étudiant sur une profession donnée et son auto-efficacité puisque le dispositif facilite la maîtrise des compétences en supprimant certaines limites (de temps, d'espace, etc.).

La deuxième étape du processus d'orientation (**exploration de domaines spécifiques**) est consacrée à l'adéquation entre les intérêts de l'élève et ses capacités. Cette étape permet à l'élève d'identifier des domaines particuliers et de les développer, par exemple en envisageant une carrière particulière et en recherchant le parcours d'études le plus approprié pour y parvenir en fonction de son profil. El-Sofany et al. proposent un modèle de sélection de carrières basé sur la logique floue [1]. Les étudiants disposent d'une interface leur permettant de saisir leurs caractéristiques, leurs compétences et leurs expériences, et le système leur propose une sélection de carrières. Pour cette étape, El Mrabet et al. utilisent les données collectées à l'étape précédente, c'est-à-dire les informations consultées par les étudiants sur

leurs appareils électroniques [2]. Ils croisent ces données avec le dossier académique de l'élève (notes, activités, tendances, etc.). Le système recommande alors à l'élève des professions qui pourraient lui correspondre.

La **crystallisation des choix** correspond à la stabilisation des intérêts [102]. L'élève choisit des formations ou des options. Dans le système proposé par El Mrabet et al., d'autres professions sont proposées à l'apprenant - toujours dans le secteur qui lui correspond le mieux - afin qu'il puisse choisir ce qui correspond le mieux à ses aspirations [2]. Thyrion et al. proposent un dispositif pour les jeunes en sciences humaines et sociales afin de les préparer à l'université [103]. En effet, la difficulté pour les lycéens ne s'arrête pas une fois qu'ils ont choisi leurs études. Ils ont besoin d'être sensibilisés aux changements que ce choix implique. Ce dispositif permet aux apprenants de se rendre compte des attentes académiques de l'université en leur permettant de consulter des examens corrigés sous forme de jeux. Le choix est alors fait parmi l'ensemble des solutions. Ce choix est concrétisé actuellement pour les élèves de terminale par la validation sur Parcoursup (<https://www.parcoursup.fr/>).

Enfin, lorsque les lycéens se sont engagés dans leur choix, c'est la phase d'**exécution** (par exemple, ils ont commencé un cours après le baccalauréat). L'adaptation réussie au nouvel environnement peut également être prise en compte dans le processus. En effet, le fait que la réalisation de ce choix soit bien vécue ou non est important dans ce processus, car une mauvaise intégration de l'étudiant peut avoir des répercussions non seulement pour lui, mais aussi pour ses pairs, ses cours et la vie universitaire [104]. Belskaya et al. expliquent que le principal problème d'adaptation est induit par le manque de communication des étudiants [99]. Pour y remédier, il existe toute une série de dispositifs numériques destinés à aider les étudiants à s'adapter avec succès à leur nouvel environnement (par exemple : site gouvernemental des services étudiants : <https://www.messervices.etudiant.gouv.fr/> ou l'environnement numérique de l'université de Lorraine : <https://ent.univ-lorraine.fr/>). L'environnement de l'université intelligente est un exemple de ces dispositifs. Ils permettent d'accéder à toutes les ressources d'information de l'université, à des discussions individuelles en ligne avec un conseiller, etc.

Dans cette section, nous analysons la diversité des dispositifs numériques pour les étudiants et leur existence pour toutes les phases de l'orientation **Tableau 3.1**. Dans notre recherche, nous nous concentrons sur les dispositifs destinés aux trois premières phases, c'est-à-dire la phase d'exploration, la phase d'adéquation entre les intérêts de l'élève et ses capacités, et la phase de cristallisation des choix.

Phases d'orientation Dispositif numérique	Exploration large	Exploration de domaines spécifiques	Cristallisation des choix	Exécution d'un choix
Système d'orientation basé sur l'internet des objets [2]	X	X	X	
Environnement de réalité virtuelle [3]	X			
Système d'orientation scolaire et professionnelle basé sur les données [1]		X		
Blocus [103]			X	X
Mes Services Étudiants (https://www.messervices.etudiant.gouv.fr)			X	X
ENT Université de Lorraine (https://ent.univ-lorraine.fr)				X

Tableau 3.1 - Phases d'orientation pour lesquelles les dispositifs numériques sont destinés

2.2. Exemples de dispositifs numériques d'aide à l'orientation à destination des lycéens

Les dispositifs numériques existants ont des objectifs, des thèmes ou des supports variés. Il est intéressant d'en présenter quelques-uns pour illustrer cette diversité et donc la difficulté qu'il peut y avoir à les représenter de manière uniforme.

Certains dispositifs sont destinés à être disponibles un jour par an et fonctionnent de la même manière que les journées d'orientation en face-à-face. Par exemple, le dispositif numérique Virbela a été utilisé pour la journée CapSup 4.0 (dans le cadre du Printemps de l'orientation) le 16 mars 2021 [11]. La journée CapSup est un événement qui regroupe des formations de l'enseignement postbac. Elle a pour objectif de faire découvrir aux lycéens les différentes filières de l'enseignement postbac et de les conseiller en matière d'orientation. La plateforme Virbela (<https://www.virbela.com>) est un monde immersif en 3D, chaque utilisateur est représenté par un personnage. L'utilisateur peut déplacer son personnage dans le monde virtuel pour simuler un forum d'orientation.

D'autres dispositifs ont des éléments disponibles qui varient tout au long de l'année, comme Parcoursup. Parcoursup est la plateforme nationale de préinscription en première année d'enseignement postbac en France. Pour y accéder, il faut être lycéen, et selon le trimestre de l'année et donc l'étape d'orientation, le contenu varie pour accompagner le lycéen dans son processus de choix d'études postbac. Par exemple, entre le 19 janvier et le 9 mars 2023, les

lycéens de terminale ont dû s'inscrire sur la plateforme et ont ajouté leurs vœux de formation après le baccalauréat.

La majorité des dispositifs sont disponibles en permanence pour que les étudiants puissent y accéder dès qu'ils en ressentent le besoin. Au sein de cet ensemble de dispositifs accessibles en permanence, tous n'offrent pas les mêmes services. Certains se sont attachés à fournir un large éventail d'informations et proposent de nombreuses pages sur les métiers, la vie étudiante et les différents diplômes, comme c'est le cas du site de l'Onisep. Certains dispositifs sont spécialisés dans un type d'information. C'est le cas de l'application mobile Wilbi qui traite uniquement des métiers. Les professionnels eux-mêmes filment de petites parties de leur journée pour présenter leur métier. Wilbi dispose d'informations formelles (compétences, formation, apprentissage, rémunération, etc.) et informelles (vision des professionnels sur le domaine, avantages et inconvénients, spécificités, etc). Elle est conçue pour s'apparenter aux applications mobiles que les lycéens ont l'habitude d'utiliser.

Certains dispositifs offrent peu de contenu informatif, mais visent à aider les lycéens à mieux se connaître à travers une série de tests. Par exemple, l'application Hello Charly (<https://hello-charly.com>) est un robot qui pose plusieurs questions au cours d'une conversation et propose ensuite des idées d'emplois ou de domaines d'activité correspondant au profil de l'élève.

Dispositif Numérique	Journée CapSup sur Virbela	Parcoursup	Onisep	Wilbi	Hello Charly
Caractéristique					
Dispositif disponible temporairement	X	X			
Dispositif disponible en permanence			X	X	X
Informations variées			X		X
Informations ciblées dans un seul domaine	X	X		X	
Informations ciblées pour l'utilisateur				X	X

Tableau 3.2 - Temporalité et variété des informations fournies par un échantillon de dispositifs numériques d'aide à l'orientation

Cette section donne des exemples concrets des dispositifs que nous sommes susceptibles d'étudier et montre leur diversité. Ces informations sont résumées dans le **Tableau 3.2**. Nous envisageons maintenant de les caractériser formellement. Les ontologies permettent de structurer un monde de connaissances comme cela a été présenté dans le Chapitre « État de l'art ». Nous présentons dans la section suivante comment elles sont utilisées dans le domaine de l'orientation et de l'éducation.

2.3. Ontologies relatives à l'orientation scolaire

Dans cette section, nous recherchons les approches pour capitaliser et stocker les données à propos des dispositifs numériques. La caractérisation de chaque dispositif doit être basée sur le même modèle afin de mettre en évidence les similitudes et les différences entre chaque dispositif. De plus, la caractérisation doit être facilement réalisable par les professionnels de l'orientation.

La représentation des dispositifs numériques à partir d'une ontologie permet de les représenter de manière uniforme en tenant compte de leur diversité. Cette représentation uniforme nous permet de les comparer entre eux.

Nous avons commencé par faire une recherche large des ontologies existantes dans le domaine, y compris celles qui traitent de l'orientation scolaire.

Delestre et al. ont proposé un standard de description pour l'éducation en France, dont une partie est représentée sous la forme d'une ontologie [8]. Cette ontologie est une description des ressources éducatives, ces ressources sont notamment liées à un « public cible », et ont des « activités d'apprentissage ». Bourdeau et al. proposent également une ontologie dans le domaine de l'éducation, leur ontologie est centrée sur l'apprenant [9]. Elle permet une description de l'apprenant, comme son comportement, ses connaissances et son niveau. L'apprentissage est un domaine dans lequel plusieurs chercheurs ont construit des ontologies, par exemple l'ontologie OMNIBUS [105]. Cette ontologie vise à aider le personnel éducatif à appliquer les connaissances des théories éducatives et à établir une méthode pour construire des systèmes éducatifs en tenant compte des théories. Comme ces ontologies ne prennent pas en compte les dispositifs numériques, nous avons identifié des ontologies traitant des dispositifs numériques dans le domaine du processus de formation.

L'apprentissage en ligne est un domaine proche des dispositifs numériques d'orientation puisqu'il s'agit des mêmes utilisateurs (des lycéens) et un apprentissage au sujet de l'orientation peut avoir lieu lors de son utilisation. Les ontologies sur ce sujet se concentrent sur l'apprentissage par le jeu. L'une d'elles informe sur les aspects possibles pour personnaliser la mécanique et le dynamisme de gamification afin de concevoir des interventions de gamification personnalisées [106]. Par exemple, si l'apprenant préfère

apprendre à l'aide d'exemples, de textes ou de vidéos. Une autre ontologie permet d'organiser les scénarios de jeu liés à des situations d'apprentissage dans lesquels les apprenants interagissent les uns avec les autres et relèvent des défis [107]. Ces ontologies sont élaborées pour aider les enseignants à construire leurs activités d'apprentissage par le jeu.

Il existe des ontologies dans le domaine de l'éducation, mais celles traitant de dispositifs numériques sont spécifiques à la construction de scénarios de jeux liés à l'apprentissage. Les ontologies traitant de l'orientation se concentrent sur l'offre de formations et de diplômes. Les ontologies existantes ne traitent pas à la fois de l'orientation scolaire au niveau de son processus et des dispositifs numériques. Ainsi, il est nécessaire de construire une ontologie permettant de caractériser les dispositifs numériques dédiés à l'orientation des lycéens.

2.4. Méthodologies de conceptualisation d'une ontologie

Selon Gruber, une ontologie est « une description explicite d'une conceptualisation » [108], et selon Noy et al., une ontologie représente « les propriétés de chaque concept décrivant diverses caractéristiques et attributs du concept, et les restrictions sur les données » [27]. Les ontologies servent de cadre unificateur pour différents points de vue et capturent les concepts utilisés et les relations entre ces concepts [109] à l'aide d'un vocabulaire commun [27]. Les ontologies comportent plusieurs types de composantes :

- Les classes : elles représentent les concepts pertinents dans le domaine d'intérêt et regroupent des objets partageant des caractéristiques. Elles peuvent contenir des instances et avoir des sous-classes.
- Les relations : elles sont utilisées pour relier les concepts entre eux.
- Les instances : il s'agit d'objets nommés et concrètement identifiables.

Il existe des méthodologies pour prendre en charge l'ensemble du cycle de vie d'une ontologie, depuis la spécification des exigences que l'ontologie doit remplir jusqu'à sa maintenance, par exemple, le cadre « Linked Open Terms » (LOT) [110] présenté dans la **Figure 3.2**. Dans ce travail, notre objectif étant d'avoir une représentation unifiée des appareils numériques, nous nous sommes concentrés sur l'implémentation de l'ontologie. Pour cela, nous avons analysé les méthodologies de conceptualisation de l'ontologie.

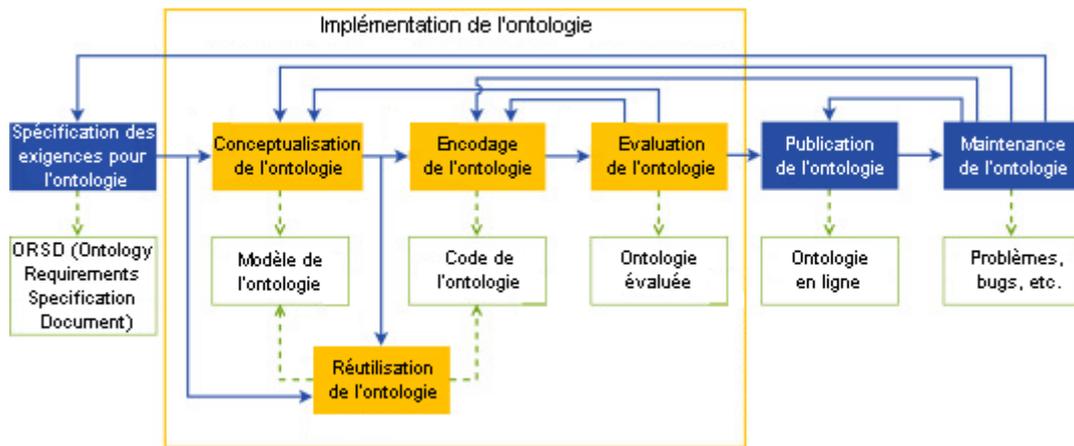


Figure 3.2 - Extrait du processus proposé pour le cycle de vie d'une ontologie [110]

L'étape de conceptualisation consiste à créer un modèle qui représente le domaine de l'ontologie. Le niveau de détail de l'ontologie dépend de son utilisation. Les questions de compétences sont la base de la conceptualisation, elles permettent de spécifier les exigences de l'ontologie [16].

Il existe plusieurs méthodologies de conceptualisation, dont certaines sont présentées par Cristani et al. [111]. Nous avons choisi d'utiliser une méthodologie évolutive, car elle est adaptée à un environnement dynamique ou difficile à comprendre. En effet, l'environnement des dispositifs numériques est dynamique et peut changer en fonction des dispositifs créés. La méthodologie choisie est « Ontology Development 101 » [27] car elle fournit une définition détaillée de chaque étape, elle est particulièrement utile pour la conceptualisation des ontologies et est citée dans divers documents de recherche [110]. Cette méthodologie comprend sept étapes.

Par la suite, dans la Section 3.1, nous présentons succinctement chaque étape et nous expliquons son application dans notre domaine d'étude, ce qui permet d'aboutir à l'élaboration d'une ontologie pour caractériser les dispositifs numériques d'aide à l'orientation.

2.5. Synthèse et questions de recherche

La diversité des dispositifs numériques dédiés à l'orientation scolaire est mise en avant dans cette section. Il existe une diversité de dispositifs en termes de disponibilité des dispositifs, de leurs sujets d'information et de leurs outils supplémentaires permettant, par exemple, d'obtenir des informations personnalisées.

La représentation de la variété des caractéristiques des dispositifs numériques et de leurs utilisateurs selon leur phase dans le processus d'orientation fait défaut actuellement.

Les ontologies semblent être une représentation des connaissances adaptée pour répondre à la question de recherche :

QR1

Comment modéliser et stocker les caractéristiques des dispositifs numériques d'aide à l'orientation tout en étant robuste à l'apparition de nouveaux dispositifs ?

Toutefois, il n'existe pas d'ontologie existante permettant de répondre à cette question de recherche. Pour y répondre, nous proposons d'élaborer une ontologie dédiée aux dispositifs numériques d'aide à l'orientation.

Elle devra fournir une caractérisation formelle des dispositifs numériques d'orientation scolaire. Cette caractérisation doit contenir des informations permettant de décrire et de distinguer les dispositifs numériques, ainsi que de cartographier et comparer les données des différents dispositifs numériques.

3. Proposition d'une ontologie pour caractériser les dispositifs numériques d'aide à l'orientation

Dans cette section, nous proposons une approche pour capitaliser et stocker les données sur les dispositifs numériques. La caractérisation de chaque dispositif doit être basée sur le même modèle afin de mettre en évidence les similitudes et les différences de chacun. De plus, la caractérisation doit être facilement réalisable par les professionnels de l'orientation.

Dans un premier temps, une représentation d'un modèle conceptuel sous forme d'UML permet de faciliter la validation des concepts et des relations par les experts de l'orientation. Plusieurs ontologies ont été modélisées avec succès à l'aide d'UML [112]. Cependant, l'utilisation des ontologies permet d'ajouter des contraintes de cardinalité et de vérifier la cohérence de l'ontologie. Nous utilisons Protégé (<https://protege.stanford.edu/>) pour concevoir l'ontologie. Les termes utilisés dans ce chapitre sont ceux issus du langage OWL et utilisés dans Protégé [113]. Les classes sont interprétées comme des ensembles d'individus. OWL possède deux types de propriété, les propriétés d'objet reliant une instance à une autre instance, et les propriétés de donnée reliant une instance à une valeur de données.

3.1. Méthodologie de conceptualisation de l'ontologie proposée

Dans cette section, nous expliquons chaque étape de la méthodologie « Ontology Development 101 » [27] en l'appliquant à la conception de l'ontologie proposée dans cette thèse.

3.1.1. Étape 1 : Définir le domaine et la portée

La première étape de la construction de l'ontologie consiste à déterminer le domaine et le champ d'application.

Dans notre domaine d'étude, l'ontologie doit permettre aux professionnels de l'orientation de caractériser les dispositifs numériques dédiés à l'orientation scolaire et de comprendre leur impact sur les lycéens pour l'orientation scolaire. L'ontologie est construite de manière à pouvoir répondre aux questions de compétences :

- Quels sont les différents types de dispositifs numériques dédiés à l'orientation ?
- Quelles sont les différentes fonctionnalités des dispositifs numériques d'aide à l'orientation ?
- Quelle est la nature de l'information proposée par ces dispositifs ?
- Comment l'utilisateur interagit-il avec le dispositif ?
- Quels services supplémentaires le dispositif fournit-il ?
- Qui est l'utilisateur du dispositif ?
- Quel est l'impact de l'utilisation du dispositif sur l'utilisateur ?
- Quelle est la facilité d'utilisation du dispositif ?

3.1.2. Étape 2 : Identifier les ontologies existantes

La deuxième étape consiste à dresser la liste des ontologies existantes qui peuvent être utilisées pour notre problème.

Dans notre domaine d'étude, les ontologies présentées dans la Section 2.3 permettent de décrire l'apprenant ou l'utilisateur du dispositif dans son contexte, mais prennent moins en compte les outils dont dispose l'utilisateur ou le type d'interface avec lequel il interagit. Dans « Web content accessibility guidelines », les auteurs, bien qu'il ne s'agisse pas d'une taxonomie ou d'une ontologie, énumèrent les éléments importants, présents dans un dispositif numérique [114]. Il s'agit d'un guide visant à rendre le contenu du web plus accessible aux personnes souffrant de handicaps tels que la cécité ou la surdité.

Ces références permettent de dresser une liste des termes importants de l'ontologie et de créer une hiérarchie entre eux aux étapes 3 et 4.

3.1.3. Étape 3 : Lister les termes et les concepts

Pour établir une liste exhaustive de termes (ou la plus exhaustive possible), nous avons utilisé trois sources d'information :

(1) Un benchmark des dispositifs d'aide à l'orientation numériques a été réalisé. Les fonctionnalités de plusieurs dispositifs numériques d'orientation sont observées, ces

dispositifs sont listés dans le **Tableau 3.3**. Nous avons choisi ces dispositifs, car ils sont parmi les plus utilisés et représentent une variété de dispositifs (ils n'ont pas tous les mêmes objectifs, ni les mêmes supports).

(2) Lors des focus groupes organisés dans le cadre du projet AILES, nous avons travaillé avec des lycéens, des enseignants et des conseillers d'orientation sur les différents dispositifs numériques d'aide à l'orientation et sur ce qu'ils attendent de ces systèmes.

(3) Les ontologies et les articles présentés à l'étape 2 ont été utilisés. Les ontologies existantes sont utilisées de façon minime car leurs objectifs sont différents des nôtres.

Nom du dispositif numérique	Lien du dispositif
Onisep	https://www.onisep.fr
Pixis	Plus accessible
Quel métier	https://www.quelmetier.fr
Bloom'r	https://www.bloomr.org
Génération 15-25	https://www.generation1525.fr
Recto Versoi	https://www.recto-versoi.com
Jobirl	https://www.jobirl.com
Study Advisor	https://www.studyadvisor.fr
Wilbi	https://wilbi-app.com

Tableau 3.3 - Dispositifs numériques d'orientation observés

Lors de la première itération de cette étape, soixante-treize termes sont ressortis, tous n'ont pas été conservés et d'autres ont été ajoutés. Un extrait de cette première liste est présenté dans l'Annexe A.

Dans la liste des termes, nous trouvons par exemple « test » et « outil de communication ». Ces termes sont présents en tant que concepts dans l'ontologie. Les « tests », en fonction du sujet, peuvent suggérer des métiers selon les domaines de prédilection de l'utilisateur, et les outils de communication permettent aux étudiants de poser des questions. L'observation des dispositifs existants lors de la création de l'ontologie évite les problèmes d'incomplétude et permet d'obtenir des modèles auto-explicatifs et facilement compréhensibles [115].

3.1.4. Étape 4 : Créer une hiérarchie de classes

La hiérarchisation des concepts permet de représenter le concept d'enfant et celui de parent. Par définition, si une classe possède une classe parent (relation de type « est un »), les membres de cette première classe sont également membre de la classe parent et héritent des conditions d'appartenance des classes en amont de leur hiérarchie [116].

Dans notre domaine d'étude, nous utilisons la hiérarchisation, par exemple, pour le concept « InformationContenu » qui a neuf concepts enfants :

- « InformationAProposDuDidpositif », « InformationEmploiStage »,
- « InformationOrientation », « InformationLycee »,
- « InformationMetier », « InformationEtablissementPostBac »,
- « InformationDiplomePostBac », « InformationPsychologiePersonalite »,
- « InformationVieEtudiante ».

3.1.5. Étapes 5 et 6 : Définir la structure interne des concepts

L'étape 5 consiste à identifier, parmi la liste des termes non utilisés en tant que concepts, ceux qui sont des propriétés d'une classe représentant un de ces concepts.

Par exemple, le concept « Outil » a comme attribut « requiereAuthentification », et le concept « OutildEchange » a comme attribut « aInterlocuteur », il s'agit d'une relation avec le concept « Interlocuteur ». Pour chaque propriété d'objet ou de donnée, l'étape 6 définit son type, dans le cas de l'attribut « requiereAuthentification », la valeur prise lors de l'instanciation de la classe associée est type booléen.

3.1.6. Étape 7 : Créer les instances

Enfin, les instances des concepts sont créées à l'étape 7.

Par exemple, l'une des instances du concept « Interlocuteur » est « etudiant ».

Ces 7 étapes conduisent à une première version de l'ontologie. Cette version doit être évaluée et affinée.

Ce processus en plusieurs étapes avec l'évaluation est itératif afin de répondre correctement aux questions de compétence et d'obtenir l'ontologie présentée dans la section suivante.

3.2. Ontologie des Dispositifs Numériques d'Aide à l'Orientation

Cette section décrit en détail l'ontologie que nous proposons et que nous avons appelée ODNAO (Ontologie des dispositifs numériques d'aide à l'orientation), il est possible de d'accéder à cette ontologie¹. L'objectif de ce modèle sémantique est de représenter les concepts et les relations dans le domaine de l'orientation scolaire afin de cartographier les données sur les dispositifs numériques pour les caractériser et les comparer entre eux. Pour créer cette ontologie, nous nous sommes basés sur une analyse des dispositifs numériques

¹ <https://github.com/MarieGribouval/OntologyDigitalDeviceForGuidance> contient le code de l'ontologie utilisée pour ces travaux.

existants et sur les travaux menés avec des professionnels de l'orientation dans le cadre du projet AILES.

Nous avons choisi de développer une ontologie modulaire. Les ontologies modulaires sont des ontologies potentiellement grandes décomposées en modules plus petits et interconnectés [117].

Ce type d'ontologie présente l'avantage (1) de faciliter la réutilisation des connaissances dans diverses applications ; (2) d'être plus facile à construire, à maintenir et à remplacer ; (3) de permettre une ingénierie distribuée des modules d'ontologie à travers différents sites et domaines d'expertise ; et (4) de permettre une gestion et une navigation entre les modules. Dans notre cas, cette construction permet d'améliorer la compréhension de l'ontologie pour les experts en orientation.

La **Figure 3.3** donne un aperçu des classes et des attributs, représentés par l'ontologie proposée. Cette ontologie est composée de six modules : (1) le module « **Nature de l'information** » permet une représentation cohérente des sujets couverts par un dispositif numérique, par exemple, des informations sur la vie étudiante fournies par des images et du texte ; (2) le module « **Interaction de l'utilisateur avec le dispositif** » vise à décrire le type d'interface et de support utilisé pour accéder au dispositif numérique ; (3) le module « **Services additionnels** » fournit des concepts et des relations qui permettent de représenter des outils qui interagissent avec l'utilisateur et ne présentent pas d'informations directement ; (4) le module « **Utilisateur** » prend en compte les connaissances relatives à la période de l'année et à la classe de l'élève qui utilise le dispositif numérique ; (5) le module « **Facilité d'utilisation (ou Utilisabilité)** » vise à représenter la facilité d'utilisation d'un dispositif numérique ; (6) le module « **Impact** » donne une représentation formelle de l'impact en termes de comportement, d'attitudes et de connaissances sur le processus d'orientation des élèves.

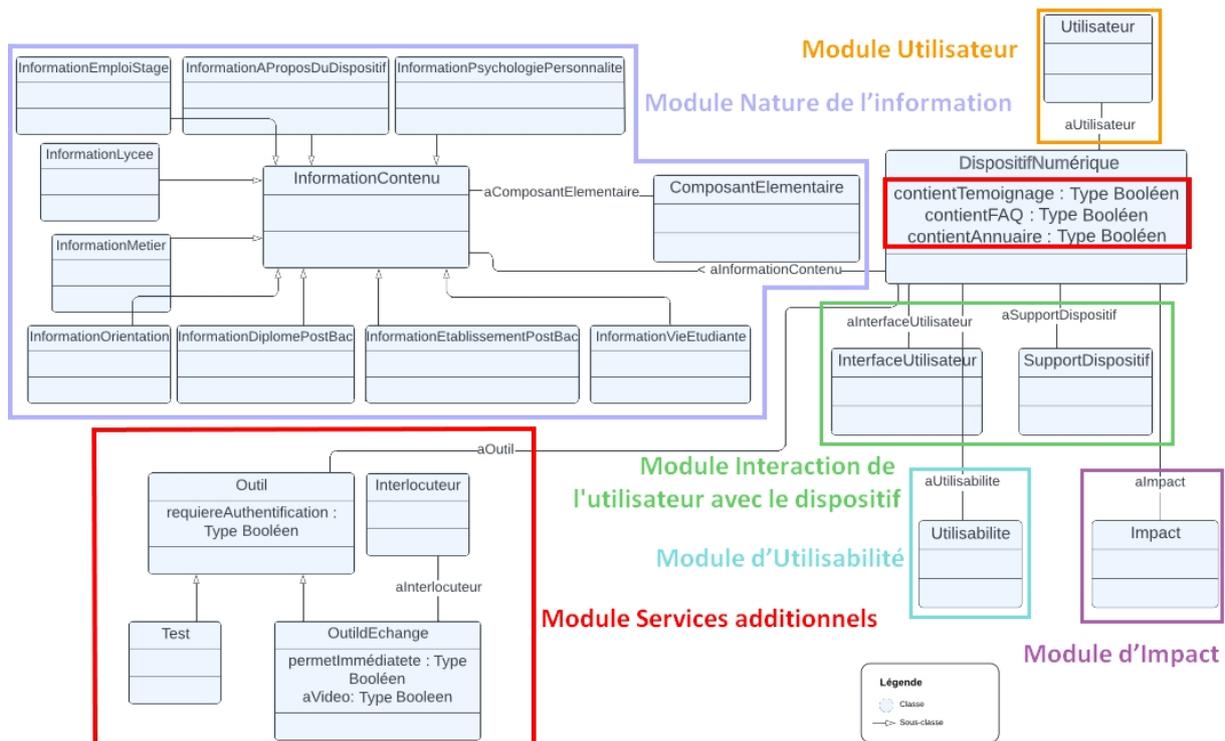


Figure 3.3 - Extrait de l'Ontologie des Dispositifs Numériques d'Aide à l'Orientation

Ces modules et leurs classes sont décrits en détail dans les sous-sections suivantes.

3.2.1. Module « Nature de l'information »

Le module « **Nature de l'information** » décrit les sujets couverts par le dispositif numérique et la manière dont l'information est présentée. Ce module contient des concepts pour les différents sujets couverts par les dispositifs d'orientation numériques et contient également des instances pour les différents types de support d'information.

Les concepts et leurs propriétés sont présentés dans le **Tableau 3.4** et le **Tableau 3.5**.

Ce module a une propriété d'objet unique car son intérêt est d'informer sur les sujets traités par le dispositif numérique et les composants (audio, image, texte et vidéo) utilisés pour les traiter.

Classe	Description	Exemple
InformationContenu	Le sujet de tout ou partie du dispositif numérique	
Information AProposDuDispositif	Description de l'objectif du dispositif et de ses membres	https://www.recto-versoi.com/qui-sommes-nous
InformationEmploiStage	Offres d'emploi ou de stage, ou conseils à ce	https://www.jobirl.com/stages-alternance/stage-etudiant

	sujet (par exemple, rédaction d'une lettre de motivation)	
InformationOrientation	Informations sur la voie choisie dans le cadre des études et sur le processus de choix	https://www.onisep.fr/orientation/l-enseignement-superieur/le-schema-des-etudes-apres-le-bac
InformationLycee	Informations au sujet du lycée (de la seconde à la terminale).	https://www.onisep.fr/orientation/le-lycee/le-lycee-general-et-technologique
InformationMetier	Informations à propos de métiers	https://www.jobirl.com/mon-orientation/metiers/environnement-eau-nature-proprete/consultant-en-developpement-durable
InformationEtablissementPostBac	Informations sur les établissements après le baccalauréat	https://www.ensgsi.univ-lorraine.fr/
InformationDiplomePostBac	Informations relatives aux formations après le baccalauréat	https://formations.univ-lorraine.fr/fr/licence/1571-licence-mention-mathematiques.html
InformationPsychologiePersonnalite	Informations relatives au fonctionnement interne de la personne (par exemple, comment savoir quelles sont ses aspirations)	https://www.cidj.com/vie-quotidienne/psycho-sante/etudiants-vous-allez-mal-c-est-normal
InformationVieEtudiante	Informations sur la vie étudiante (par exemple sur le logement étudiant)	https://www.letudiant.fr/lifestyle/logement_2/logement-crous-la-procedure-d-attribution-est-modifiee.html
ComposantElementaire	Représente tous les sons, images, textes et vidéos	Texte et image (information sur l'orientation sur ce site web : https://www.onisep.fr/orientation/l-enseignement-superieur/le-schema-des-etudes-apres-le-bac)

Tableau 3.4 - Classes du module « Nature de l'information »

Nom de la propriété	Domaine	Range ²	Description
aComposantElementaire	InformationContenu	La classe « ComposantElementaire »	Assigne une information à un composant élémentaire
aInformationContenu	DispositifNumerique	La classe « InformationContenu »	Assigne un dispositif numérique à une/des sujet(s) d'information

Tableau 3.5 - Propriétés du module « **Nature de l'information** »

La structure présentée pour les différents modules est un choix. Plusieurs possibilités de représentation existent, par exemple dans ce module il est possible de :

- Créer des sous-classes à la classe « ComposantElementaire », ces sous-classes seraient « Audio », « Texte », « Image », « Video ». Le problème avec cette représentation est le grand nombre d'instance à créer que cela provoque. Par exemple, si tous les sujets sont traités dans un dispositif, et que tous ont du texte, cela nécessiterait de créer neuf instances pour la classe « Texte ».
- Créer des propriétés de donné pour la classe « InformationContenu », ces propriétés seraient « aAudio », « aTexte », « aImage », « aVideo » et les valeurs possibles de ces propriétés seraient « Oui » ou « Non ».
- Créer des instances à la classe « ComposantElementaire », ces instances sont « audio », « texte », « image », « video ». Les instances de « InformationContenu » sont liées aux instances « ComposantElementaire » selon si l'information est présentée en utilisant le composant élémentaire.

Les deux dernières propositions sont possibles et apportent la même information. La troisième solution est privilégiée à la deuxième, car il semble plus pratique, lors de l'instanciation, s'il existe plusieurs instances pour la classe « ComposantElementaire » de créer une relation uniquement si un lien existe plutôt que de devoir remplir toutes les propriétés de données. De plus, la représentation à l'aide d'une classe et d'instances pour les composants élémentaires permet de les mettre en avant afin de mieux voir les composants existants et s'il y a un manque. Si, dans les dispositifs créés dans le futur, un autre composant

² Il est difficile de traduire ce terme en français, les termes « plage de valeur » ou « valeur admissible » seraient adaptés pour les propriétés de données mais ne le seraient pas pour les propriétés d'objet. Le terme « range » est conservé dans la suite de ce document.

élémentaire apparait, il suffira d'ajouter une instance à la classe « ComposantElementaire ».

3.2.2. Module « Interaction de l'utilisateur avec le dispositif numérique »

Ce module indique comment l'utilisateur interagit avec le dispositif, c'est-à-dire quel support peut être utilisé pour y accéder et comment le dispositif est construit. Les concepts et leurs propriétés sont présentés dans le **Tableau 3.6** et le **Tableau 3.7**.

Classe	Description	Exemple
SupportDispositif	Le dispositif numérique peut être une application mobile, une application d'ordinateur ou un site web ; cette classe décrit également le support physique utilisé	Instance ayant comme valeur de propriétés « ApplicationMobile » et « Mobile » (pour l'application mobile Wilbi)
InterfaceUtilisateur	Il décrit s'il s'agit d'une interface par page ou d'un monde immersif et la temporalité de l'interface	Instance ayant comme valeur de propriétés « Interface Par Page » et « Permanent » (par exemple pour le dispositif Onisep)

Tableau 3.6 - Classes du module « Interaction de l'utilisateur avec le dispositif numérique »

Nom de la propriété	Domaine	Range	Description
disponibilite	InterfaceUtilisateur	Une liste, les valeurs permises sont : {« Temporaire », « Permanent »}	Définie la disponibilité de l'interface utilisateur
typeInterface	InterfaceUtilisateur	Une liste, les valeurs permises sont : {« InterfaceParPage », « MondeImmersif »}	Définie le type d'interface du dispositif
typeSupport	SupportDispositif	Une liste, les valeurs permises sont : {« SiteWeb », « ApplicationMobile », « ApplicationOrdinateur »}	Définie le type de support du dispositif

supportPhysique	SupportDispositif	Une liste, les valeurs permises sont : {« Ordinateur », « Mobile »}	Définie le support physique utilisé
-----------------	-------------------	---	-------------------------------------

Tableau 3.7 - Propriétés du module « Interaction de l'utilisateur avec le dispositif numérique »

Pour ce module, le choix est fait d'utiliser des propriétés de donnée contrairement au module précédent, car ce module ne peut être composé que de deux classes. Indiquer les valeurs de leurs propriétés de donnée (« disponibilite », « typeInterface », « typeSupport » et « supportPhysique ») est donc moins chronophage que pour un module avec plus de classes.

Si, dans les dispositifs créés dans le futur, d'autres types de support ou d'autres interfaces apparaissent, les éléments possibles dans les listes devront être modifiés.

3.2.3. Module « Services additionnels »

Pour certains dispositifs, l'utilisateur peut accéder à des fonctionnalités personnalisées. Ce module décrit les types d'informations personnalisées auxquelles il est possible d'accéder. Dans l'ODNAO, ces fonctionnalités personnalisées sont appelées outils. L'ontologie considère actuellement deux types d'outils : les « tests » qui sont une série de questions auxquelles une personne doit répondre pour obtenir des informations personnalisées sur un sujet spécifique tel que l'emploi ; et les « outils de communication », qui comprennent tous les outils qui permettent à l'utilisateur d'échanger des messages instantanément ou non, sous forme de texte, d'audio, de vidéos avec des interlocuteurs. Les concepts et leurs propriétés sont présentés dans le **Tableau 3.8** et le **Tableau 3.9**.

Classe	Description	Exemple
Interlocuteur	Entité ou individu avec lequel l'utilisateur peut échanger (c'est-à-dire transmettre quelque chose et recevoir quelque chose - la chose transmise pouvant être de l'audio, de la vidéo, du texte ou des images)	« etudiant » (il s'agit de l'interlocuteur pour le dispositif Study Advisor)
SujetDeTest	Le résultat du test peut fournir des informations sur les centres d'intérêt de l'utilisateur, les emplois qui y correspondent ou sa personnalité	« metier » (il s'agit du sujet de ce test : https://www.studyrama.com/tests/etes-vous-fait-pour-devenir-enseignant-53)

Outil	Ensemble d'outils que l'utilisateur peut avoir à sa disposition dans un dispositif numérique	
OutildEchange	Un outil d'échange (c'est-à-dire transmettre quelque chose et recevoir quelque chose - la chose transmise pouvant être de l'audio, de la vidéo, du texte ou des images).	https://www.monorientationligne.fr/qr/tchat.php?popup
Test	Un formulaire que l'utilisateur peut remplir et qui renvoie un résultat	https://www.studyrama.com/tests/etes-vous-fait-pour-devenir-enseignant-53

Tableau 3.8 - Classes du module « **Services additionnels** »

Nom de la propriété	Domaine	Range	Description
permetImmédiate	OutildEchange	Une liste, les valeurs permises sont : {Oui, Non}	Indique si l'outil d'échange permet l'immédiateté ou non.
contientFaq	DispositifNumerique	Une liste, les valeurs permises sont : {Oui, Non}	Indique si le dispositif numérique contient les réponses à un ensemble de questions posées fréquemment ou non
contientAnnuaire	DispositifNumerique	Une liste, les valeurs permises sont : {Oui, Non}	Indique si le dispositif numérique contient un annuaire ou non
contientTemoignage	DispositifNumerique	Une liste, les valeurs permises sont : {Oui, Non}	Indique si le dispositif numérique possède ou non des témoignages
aInterlocuteur	OutildEchange	La classe « Interlocuteur »	Assigne un interlocuteur à un outil d'échange
aVideo	OutildEchange	Une liste, les valeurs permises sont : {Oui, Non}	Indique si un outil d'échange possède ou non une vidéo

aSujetDeTest	Test	Une liste, les valeurs permises sont : {« Interet », « Metier », « Personnalite »}	Indique le sujet abordé par le test.
requiereAuthentication	Outil	Une liste, les valeurs permises sont : {Oui, Non}	Indique si outil nécessite ou non une authentification de la part de l'utilisateur

Tableau 3.9 - Propriétés du module « **Services additionnels** »

Les instances des trois premiers modules de cette ontologie peuvent être créées par n'importe quel professionnel de l'orientation en observant le dispositif numérique et en suivant le protocole créé pour instancier l'ontologie.

3.2.4. Module « Utilisateur »

Le module « **Utilisateur** » donne des informations sur l'utilisateur. En effet, nous partons du principe que l'impact d'un dispositif numérique sur un lycéen n'est pas le même selon l'année précédant le baccalauréat, le type de baccalauréat et la période de l'année. Chacune de ces années est séparée en 4 phases. Par exemple, en seconde, ce processus permet de déterminer le type de bac choisi et les éventuelles options, alors qu'en terminale, le processus permet de déterminer le choix d'étude après le baccalauréat. Pour l'année de terminale, le processus est supporté par le dispositif Parcoursup.

L'ontologie permet de prendre en compte ces trois aspects (année, type de bac, étape du processus). Le concept d'utilisateur et ses propriétés sont présentés dans le **Tableau 3.10** et le **Tableau 3.11**.

Classe	Description	Exemple
Utilisateur	Groupe de personnes qui utilisent l'appareil	Lycéen en « Seconde » en baccalauréat « General » et dans l'étape « Cristallisation »

Tableau 3.10 - Classes du module « **Utilisateur** »

Nom de la propriété	Domaine	Range	Description
annee	Utilisateur	Une liste, les valeurs permises sont : {« Seconde », « Première », « Terminale »}	Indique la classe de l'utilisateur
aUtilisateur	DispositifNumerique	La classe « Utilisateur »	Indique que le dispositif numérique a un utilisateur
etapeDuProcessus	Utilisateur	Une liste, les valeurs permises sont : {« ExplorationLarge », « ExplorationApprofondie », « Cristallisation », « Execution »}	Indique le moment de l'année du processus d'orientation dans lequel est l'utilisateur
impactCollecte	Utilisateur	La classe « Impact »	Indique l'impact que peut collecter un utilisateur sur son processus d'orientation en utilisant un dispositif
typeDeBaccalaureat	Utilisateur	Une liste, les valeurs permises sont : {« Général », « Technologique », « Professionnel »}	Indique la filière du baccalauréat de l'utilisateur
utilisabiliteResentie	Utilisateur	La classe « Utilisabilite »	Indique l'utilisabilité que peut ressentir un utilisateur en utilisant un dispositif

Tableau 3.11 - Propriétés du module « **Utilisateur** »

3.2.5. Module d' « Utilisabilité »

Le **module d'utilisabilité** vise à représenter la facilité d'utilisation des dispositifs numériques [118]. L'utilisabilité est définie selon cinq axes [119] : (1) la navigabilité, qui représente la facilité de déplacement dans le dispositif ; (2) le contenu, en termes de quantité d'informations et de mise à jour ; (3) la représentation graphique, qui est évaluée en fonction de la lisibilité des éléments du dispositif ; (4) l'ergonomie, c'est-à-dire l'efficacité de l'appareil lors de son utilisation ; et (5) l'image de marque.

Le choix est fait d'analyser l'utilisabilité et non l'acceptabilité ou l'acceptation d'un dispositif.

L'acceptabilité est définie comme la représentation mentale plus ou moins positive que se fait l'utilisateur du dispositif avant de l'utiliser [120]. Ce n'est pas ce que nous cherchons à mesurer ici, d'autant plus que certains dispositifs ne sont pas connus par les lycéens avant qu'un professionnel de l'éducation les amène à l'utiliser. Ainsi, une représentation mentale existe pour certains dispositifs connus mais pas pour tous.

L'acceptation d'un dispositif est une évaluation pragmatique après l'utilisation [120]. L'utilisabilité est l'un des axes considéré lors de l'évaluation de l'acceptation, mais l'acceptation considère également l'intention de réutiliser le dispositif et le plaisir provoqué par le dispositif. Ce n'est pas l'objectif de ce module qui cible uniquement l'utilisabilité. Le plaisir provoqué est considéré dans l'impact, et la volonté de réutiliser un dispositif n'a pas de sens pour les dispositifs temporaires que nous considérons dans notre problème.

Le concept d'utilisabilité et les propriétés de ce module sont présentés dans le **Tableau 3.12** et le **Tableau 3.13**. Les valeurs sont dans l'intervalle [0 ; 3], une valeur de « zéro » signifie que l'attribut évalué est pauvre, une valeur de « trois » signifie que l'attribut évalué est excellent. L'échelle utilisée est l'une de celle évoquée par Martínez-Sala et al., il s'agit de celle proposant le plus de nuance [119].

Classe	Description	Exemple
Utilisabilite	Facilité d'usage du dispositif numérique	« utilisabiliteHelloCharly »

Tableau 3.12 - Classes du module « Utilisabilité »

Nom de la propriété	Domaine	Range	Description
aUtilisabilite	DispositifNumerique	La classe « Utilisabilite »	Indique l'utilisabilité qui est liée à un dispositif numérique

valeurDeNavigabilite	Utilisabilite	Valeur numérique comprise dans l'intervalle [0 ; 3]	Indique la facilité de mouvement dans le dispositif [121], [119], [122]
valeurDeContenu	Utilisabilite	Valeur numérique comprise dans l'intervalle [0 ; 3]	Indique la qualité d'informations et de mises à jour [121], [119], [122]
valeurRepresentationGraphique	Utilisabilite	Valeur numérique comprise dans l'intervalle [0 ; 3]	Indique la lisibilité des éléments du dispositif [121], [119], [122]
valeurDErgonomie	Utilisabilite	Valeur numérique comprise dans l'intervalle [0 ; 3]	Indique l'efficacité du dispositif pendant son utilisation [121], [119], [122]
valeurImageDeMarque	Utilisabilite	Valeur numérique comprise dans l'intervalle [0 ; 3]	Indique l'utilisation de l'image de marque [121], [119]
valeurUtilisabilite	Utilisabilite	Valeur numérique comprise dans l'intervalle [0 ; 3]	Indique l'utilisabilité globale

Tableau 3.13 - Propriétés du module « **Utilisabilité** »

3.2.6. Module d' « Impact »

Dans cette étude, nous nous concentrons sur l'impact du dispositif numérique sur le processus d'orientation scolaire. Nous proposons de définir l'impact sur le processus d'orientation à partir du modèle KAB. Ainsi, comme présenté dans le Chapitre « État de l'art »,

cet impact est défini comme l'impact sur les connaissances, l'impact sur l'opinion et l'impact sur le comportement de l'utilisateur [40]. L'évaluation de l'impact sur l'utilisateur est réalisée à l'aide de l'évaluation de trois catégories :

(1) La catégorie « **utilité perçue du dispositif** » se réfère à des attributs fonctionnels et instrumentaux pragmatiques qui visent à atteindre un objectif. Un dispositif dont les attributs sont essentiellement utilitaires remplit immédiatement des objectifs tangibles sans susciter d'expérience émotionnelle forte ni offrir d'activité attrayante. Les attributs utilitaires peuvent être mesurés par l'efficacité et l'utilité (la réalisation de l'objectif) et l'efficience (le temps nécessaire pour atteindre l'objectif tangible).

(2) La catégorie « **valeur hédonique** » désigne les attributs qui visent à assurer l'épanouissement de l'utilisateur (par exemple, l'amusement, l'excitation et le plaisir). Par exemple, un système dont les attributs sont principalement hédoniques répond à un besoin intangible et suscite une expérience hautement émotionnelle.

(3) La catégorie « **valeur sociale** » présente des attributs contribuant à des objectifs significatifs et sociaux. Ces attributs fournissent et représentent des activités engageantes et sociales.

Le concept d'impact et ses propriétés sont présentés dans le **Tableau 3.14** et le **Tableau 3.15**. La justification du choix de ces trois catégories et la définition de l'impact sont traitées dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ». Les valeurs des propriétés de ce module sont comprises dans l'intervalle [1 ; 7]. Cet intervalle correspond à l'échelle de Likert utilisée pour mesurer l'impact. Une valeur de un indique que les lycéens ne perçoivent pas d'apport du dispositif sur la catégorie mesurée, une valeur de sept indique que les lycéens perçoivent un fort potentiel du dispositif pour la catégorie mesurée. Le choix est fait d'avoir une valeur « ni d'accord, ni pas d'accord » dans l'échelle pour laisser une option pour les lycéens ne sachant pas quoi répondre. Et, sept semble être un nombre de réponse permettant une bonne précision [123], [124].

Classe	Description	Exemple
Impact	Changement dans les connaissances, les opinions et le comportement de l'élève envers son processus d'orientation après l'utilisation du dispositif	impactHelloCharly

Tableau 3.14 - Classes du module « Impact »

Nom de la propriété	Domaine	Range	Description
alImpact	DispositifNumerique	La classe « Impact »	Indique qu'un dispositif numérique a un impact
valeurGlobale	Impact	Valeur numérique comprise dans l'intervalle [1 ; 7]	Indique la moyenne pondérée des valeurs hédonique, sociale et d'utilité
valeurHedonique	Impact	Valeur numérique comprise dans l'intervalle [1 ; 7]	Indique le plaisir pris par l'utilisateur en utilisant le dispositif
valeurUtilitaire	Impact	Valeur numérique comprise dans l'intervalle [1 ; 7]	Indique la force de la réalisation des objectifs de l'utilisateur concernant son processus d'orientation
valeurSociale	Impact	Valeur numérique comprise dans l'intervalle [1 ; 7]	Indique la force des activités sociale induites par l'utilisation du dispositif numérique

Tableau 3.15 - Propriétés du module « Impact »

Nous avons décrit en détail l'ontologie proposée pour caractériser les dispositifs numériques d'aide à l'orientation. Dans la section suivante, nous décrivons l'évaluation de cette ontologie.

3.3. Vérification et validation de l'ontologie proposée

Il existe une grande variété d'évaluations d'ontologies. McDaniel et al. les présentent selon une phase de vérification et validation et une phase d'amélioration [125].

La vérification consiste à s'assurer que l'ontologie construite est correcte. Elle porte sur la syntaxe et la cohérence de l'ontologie. Dans notre étude, nous avons utilisé le raisonneur HerMiT [126] présent dans Protégé itérativement sur toutes les versions de l'ontologie. Il nous permet de créer des inférences et de vérifier la présence d'erreurs logiques. L'ontologie a

également été vérifiée avec OOPS! (<https://oops.linkeddata.es/>). L'outil OOPS! identifie des erreurs dans une ontologie. Il détecte les erreurs structurelles (la syntaxe et la sémantique formelle), les erreurs fonctionnelles (en considérant l'utilisation et la fonctionnalité prévues de l'ontologie proposée) et les erreurs de profilage de l'utilisabilité (contexte de communication d'une ontologie) [127]. En utilisant OOPS! sur notre ontologie, nous avons obtenu des erreurs mineures qui étaient dues au manque d'annotation pour les classes et les propriétés.

La validation consiste à s'assurer que l'ontologie construite est la bonne (pertinence et complétude), notamment en vérifiant que l'ontologie répond aux questions de compétences. Dans notre étude, l'ontologie proposée a été vérifiée par des experts en éducation. Ces derniers ont confirmé que les concepts et les relations de l'ontologie avaient un sens pour eux.

Un autre niveau de validation est atteint avec l'utilisation de l'ontologie, qui est instanciée pour les dispositifs numériques « Hello Charly » et « Onisep. Cette instanciation est présentée dans le Chapitre « Expérimentations ».

La phase d'amélioration considère la réutilisation d'une partie de l'ontologie. Ceci a été envisagé en créant une ontologie modulaire. Un module peut être utilisé indépendamment du reste de l'ontologie.

4. Conclusion

Ce chapitre a permis de présenter la conception et la structure de l'ontologie proposée pour caractériser et stocker les caractéristiques des dispositifs numérique d'aide à l'orientation.

La création de l'ontologie a été basée sur « Ontology Development 101 » et sur les dispositifs numériques d'orientation existants.

L'ontologie proposée a été vérifiée par l'utilisation de l'outil OOPS! et validée par confrontation avec des experts en éducation. Les experts en éducation ont confirmé que les concepts et les relations de l'ontologie ont un sens pour eux.

Les ontologies spécifiques au domaine de l'éducation concernent la description du processus d'apprentissage et d'enseignement, comme l'ontologie OMNIBUS [26] mais ne traitent pas des dispositifs numériques ou de l'orientation scolaire. Il n'est pas possible de comparer directement l'ontologie proposée avec les ontologies existantes.

Au fil du temps, l'architecture de l'ontologie avec ses différents modules n'évoluera pas. Toutefois, des mises à jour au sein des modules peuvent être envisagées tout en maintenant la cohérence globale de l'ontologie. Par exemple, l'ontologie peut être étendue à d'autres dispositifs numériques dédiés à l'orientation, en ajoutant des sous-concepts au concept

d'"outil". Cette extensibilité est nécessaire, par exemple, pour prendre en compte de nouveaux dispositifs innovants.

Nous avons conçu l'ontologie « Ontologie des Dispositifs Numériques d'Aide à l'Orientation » qui modélise les dispositifs numériques d'orientation pour les lycéens. Ce modèle facilite la description de la nature des informations d'un tel dispositif, de ses services additionnels, de ses utilisateurs, de la manière dont l'utilisateur peut interagir avec lui, de sa facilité d'utilisation et de son impact sur les élèves.

L'ontologie offrira aux professionnels de l'orientation la possibilité d'avoir une vue d'ensemble d'un dispositif et d'en connaître l'impact. De cette manière, les professionnels de l'orientation auront la possibilité d'identifier les caractéristiques d'un dispositif numérique et de comparer la présence ou l'absence de ces caractéristiques entre plusieurs dispositifs. Afin d'améliorer la comparaison entre les dispositifs numérique, le chapitre suivant propose de mesurer leur impact sur le processus d'orientation des lycéens. Cet impact a pour objectif d'être un élément de comparaison entre les dispositifs.

MODELE DE MESURE DE L'IMPACT DES DISPOSITIFS NUMERIQUES D'AIDE A L'ORIENTATION

TABLE DES MATIERES

1. Introduction.....	64
2. Description et mesure du KAB dans notre contexte	66
3. Évaluation des dispositifs numériques d'aide à l'orientation	67
3.1. Synthèse sur les indicateurs d'évaluation des dispositifs numériques	67
3.2. Définition des catégories d'indicateur dans le contexte de l'étude des dispositifs dédiés à l'orientation scolaire	69
3.2.1. Utilisateur.....	69
3.2.2. Ressenti global	70
3.2.3. Valeur perçue du dispositif.....	70
3.2.4. Valeur émotionnelle provoquée (hédonique)	70

3.2.5. Valeur sociale induite	71
3.2.6. Utilité perçue du dispositif	71
3.2.7. Utilisation du dispositif et Valeur sensorielle provoquée	72
3.2.8. Comportement futur de l'utilisateur	73
3.2.9. Synthèse des indicateurs utilisés dans notre modèle	73
3.3. Modèle de relation entre l'impact et les indicateurs d'évaluation des dispositifs numériques	74
3.3.1. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques par une régression linéaire multiple.....	75
3.3.2. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques par une modélisation par équation structurelle.....	76
3.3.3. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques avec les réseaux bayésiens	77
3.3.4. Synthèse	77
4. Protocole d'estimation de l'impact	78
5. Conclusion	80

1. Introduction

Une ontologie est proposée dans le chapitre précédent afin de modéliser et stocker les données associées aux dispositifs numériques d'aide à l'orientation. Pour cela, six modules sont utilisés : « **Nature de l'information** », « **Services additionnels** », « **Utilisateur** », « **Interaction de l'utilisateur avec le dispositif** », « **Utilisabilité** » et « **Impact** ». Ce dernier module, contrairement aux cinq premiers n'est pas obtenu par l'observation du dispositif mais par l'effet qu'il induit lors de l'usage sur les lycéens. De plus, il dépend de l'objectif des professionnels de l'éducation avec lesquels nous travaillons. Les professionnels de l'orientation souhaitent avoir des informations de l'utilisation d'un dispositif numérique sur le processus d'orientation. S'il s'agissait de l'impact sur un dispositif, par exemple « est-ce que les lycéens vont avoir envie de réutiliser le dispositif », l'impact à mesurer ne serait pas le même. Ainsi, il est nécessaire de détailler la façon dont est construit le module « **Impact** ».

Dans ce chapitre, nous apportons une réponse à la question de recherche suivante :

QR2

Comment mesurer l'impact des dispositifs numériques d'aide à l'orientation sur les lycéens dans leur démarche d'orientation ?

Pour répondre à cette question, nous examinons les mesures d'impact de dispositifs numériques existant dans la littérature et nous précisons notre position par rapport aux différents types de travaux existants. La définition et la façon de mesurer l'impact sont dépendantes du contexte. Or, à notre connaissance, il n'existe pas de travail dans la littérature actuelle concernant les dispositifs numériques dédiés à l'orientation scolaire. L'impact doit être en accord avec ce que les personnes utilisant l'impact cherchent à apprendre.

Dans notre contexte, l'impact est à destination des professionnels de l'orientation (le personnel du rectorat et les membres du projet AILES). L'objectif de cette mesure est d'aider les professionnels de l'orientation à comparer les dispositifs entre eux afin d'identifier les plus pertinents selon les élèves. La comparaison actuellement est délicate pour les professionnels de l'orientation en raison du grand nombre de dispositifs existants, de leur grande diversité et des besoins différents des élèves en fonction dans leur parcours dans les lycées.

Nous nous plaçons, dans le projet AILES dans l'objectif suivant : « Donner envie aux jeunes de se lancer dans une démarche d'orientation de qualité et proactive, en leur fournissant des informations simples, accessibles et attractives sur les formations et leurs débouchés ». Ainsi, nous n'évaluons pas les qualités d'un dispositif à bien orienter un lycéen. La vocation de cette mesure est de savoir si l'utilisation d'un dispositif donné permet au lycéen de démarrer ou de poursuivre sa démarche d'orientation de façon proactive. L'impact considéré est celui sur les utilisateurs principaux, qui sont les lycéens.

L'intérêt est d'avoir une mesure unique de l'impact pour tous les dispositifs numériques et de proposer un élément de comparaison uniforme entre tous les dispositifs numériques malgré leur diversité de structure.

Pour cela, nous proposons un protocole de mesure de l'impact utilisant un questionnaire à destination des lycéens après leur utilisation du dispositif. Ce protocole doit être utilisable sur différents dispositifs numériques d'aide à l'orientation et sur différentes classes de lycéens (de la seconde à la terminale).

En se basant sur la présentation générique du modèle KAB présenté dans la Section 3.1 dans le Chapitre « État de l'art », nous proposons le modèle KAB adapté à l'orientation scolaire dans une première section.

Lors de l'analyse des indicateurs d'évaluations des dispositifs numérique, la spécificité des indicateurs pour le contexte d'évaluation est relevée. Afin de déterminer les indicateurs

pertinents pour notre problème, nous proposons une synthèse générique des catégories des indicateurs fréquemment utilisés dans la section deux avant de les définir pour notre contexte.

Dans la troisième section de ce chapitre, nous définissons les modèles de recherche de corrélation présentés dans la Section 3.3 du Chapitre « État de l'art », dans notre contexte.

Enfin, le protocole de mesure de l'impact des dispositifs numériques sur le processus d'orientation des lycéens est expliqué dans une quatrième section.

2. Description et mesure du KAB dans notre contexte

L'impact qui nous intéresse est l'impact sur le processus d'orientation. Nous cherchons à savoir si l'utilisation d'un dispositif numérique donné modifie les connaissances, les opinions et les comportements d'un lycéens sur son processus d'orientation. Nous ne cherchons pas ici à mesurer l'impact sur le dispositif numérique (par exemple : est-ce que le dispositif va être adopté par les lycéens ?).

L'impact selon le modèle KAB est structuré selon trois axes : « **Knowledge** », « **Attitude** », « **Behaviour** ». Pour préciser l'évaluation de chaque axe, dans notre contexte, nous proposons des affirmations (notées de I1 à I6) que l'utilisateur pourra évaluer.

L'axe « **Knowledge** » témoigne de l'acquisition de connaissances dans le domaine de l'orientation. Les connaissances sont notamment liées aux sujets principaux des dispositifs numériques présentés dans la Section 3.2.1 du Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation » : les offres d'emploi/de stage, le lycée, les métiers, l'orientation, les diplômes postbac, les établissements postbac, la vie étudiante et la psychologie/personnalité.

Cet axe est évalué à l'aide de deux affirmations :

(I1) Après avoir utilisé le dispositif numérique, tes connaissances de ton projet d'orientation se sont améliorées,

(I2) Après avoir utilisé le dispositif numérique, tes connaissances sur toi-même se sont améliorées.

« **Attitude** » : Nous nous sommes concentrés sur l'évaluation du sentiment de stress, car le choix d'orientation est générateur de stress chez les lycéens, ce stress est principalement lié aux doutes qu'ont sur eux les lycéens [128].

Cet axe est évalué à l'aide de deux affirmations :

(I3) Après avoir utilisé le dispositif numérique, tu t'es senti(e) moins stressé(e) par ton projet d'orientation,

(I4) Après avoir utilisé le dispositif numérique tu t'es senti(e) plus à l'aise avec toi-même.

« **Behaviour** » : Par le comportement, nous évaluons si l'utilisation du dispositif induit chez le lycéen d'autres activités liées à son processus d'orientation, les deux activités mesurées sont le souhait de poursuivre ses recherches sur leur projet d'orientation ou le souhait d'en discuter.

Cet axe est évalué à l'aide de deux affirmations :

(I5) Après avoir utilisé le dispositif numérique, tu souhaites faire des recherches sur ton projet d'orientation pour en savoir plus,

(I6) Après avoir utilisé le dispositif numérique, tu souhaites discuter de ton projet d'orientation avec tes proches (amis, famille, professeur, ...).

L'impact selon les axes connaissance, opinion et comportement est défini dans notre contexte. Six affirmations sont utilisées pour le mesurer. Dans la section suivante, nous présentons les catégories d'indicateur dédiées à l'évaluation des dispositifs que nous utilisons.

3. Évaluation des dispositifs numériques d'aide à l'orientation

3.1. Synthèse sur les indicateurs d'évaluation des dispositifs numériques

À partir de l'état de l'art de des indicateurs réalisé dans la Section 3.2 du Chapitre « État de l'art », nous proposons une synthèse des indicateurs les plus fréquents et les répartissons selon neuf catégories. Des catégories ont déjà été utilisés par certains auteurs mais sans toutes les prendre en compte ou en utilisant des catégories spécifiques au contexte comme la volonté de payer [57].

Cette synthèse a pour objectif de permettre d'identifier les catégories pertinentes selon le contexte d'évaluation du dispositif numérique.

La catégorie « **utilisateur** ». Ces informations sont spécifiques à chaque utilisateur ou type d'utilisateurs. Elles permettent de le décrire.

La catégorie « **ressenti global** ». Elle mesure l'image globale que l'utilisateur a du dispositif.

La catégorie « **valeur perçue** ». Elle mesure la plus-value perçue par l'utilisateur du dispositif.

La catégorie « **valeur émotionnelle provoquée** ». Elle mesure les différentes émotions que provoque l'utilisation du dispositif.

La catégorie « **valeur sociale induite par le dispositif** ». Elle mesure le changement provoqué par l'utilisation du dispositif sur la vision que l'utilisateur a de lui-même et des autres.

La catégorie « **utilité perçue du dispositif** ». Elle mesure, dans un contexte d'apprentissage, ce que l'utilisateur pense avoir appris et s'il pense avoir appris des choses. Ces choses sont à définir selon le contexte.

La catégorie « **utilisation du dispositif** ». Elle mesure la facilité et la fluidité d'utilisation.

La catégorie « **valeur sensorielle provoquée** ». Elle mesure ce qu'a ressenti l'utilisateur selon les sens touchés par le dispositif (par exemple, l'ouïe ou la vue).

La catégorie « **comportement futur** ». Elle mesure le souvenir du dispositif après plusieurs jours/semaines, le souhait d'en parler à des proches et le renouvellement de l'expérience.

Il s'agit d'une catégorisation possible des indicateurs. Ces indicateurs permettent de mesurer la vision de l'utilisateur, son comportement par rapport au dispositif tout en considérant ses caractéristiques personnelles.

Le **Tableau 4.1** détaille chacune des catégories. Cette représentation sous forme de neuf catégories n'est pas invariable, des indicateurs peuvent se placer dans plusieurs catégories. Ce tableau permet de mettre en avant les éléments importants à considérer lors de l'évaluation de dispositifs numériques. Nous avons choisi de mettre en avant ces catégories car elles englobent les indicateurs présents dans la littérature sans dépendre du contexte. Ce sont les indicateurs mis ensuite dans les catégories qui dépendent du contexte.

Catégorie	Exemple d'indicateurs	Auteurs
Utilisateur	Âge, Genre, Niveau d'éducation, Revenu mensuel, État de santé, ...	Karacapilidis Nikos Ed [5] ; Zhang Fangyu et al. [55] ; Lessiter Jane et al. [57] ; Oliveira Renata P et al. [129]
Ressenti global	Le dispositif est nouveau dans sa catégorie, L'expérience de l'utilisation était bonne, ...	Zhang Fangyu et al. [55] ; Lessiter Jane et al. [57].
Valeur perçue	Le dispositif durera dans le temps, La qualité du dispositif est uniforme, ...	Zhang Fangyu et al. [55]
Valeur émotionnelle provoquée	Intensité du plaisir, Gain de confiance en soi, ...	Zhang Fangyu et al. [55] ; Lessiter Jane et al. [57] ; Dickinger et al. [130] ; Oliveira Renata P et al. [129] ; Choi Sujeong [131] ; Batra Rajeev et al. [132]

Valeur sociale induite par le dispositif	L'utilisation du dispositif m'apporte une approbation sociale, L'utilisation du dispositif améliore la façon dont je me perçois, ...	Karacapilidis Nikos Ed [5] ; Zhang Fangyu et al. [55] ; Dickinger et al. [130] ; Oliveira Renata P et al. [129] ; Choi Sujeong [131]
Utilité perçue du dispositif	L'utilisation du dispositif m'a engagé sur le plan intellectuel, Le sujet traité par le dispositif est adapté, ...	Karacapilidis Nikos Ed [5] ; Lessiter Jane et al. [57] ; Dickinger et al. [130] ; Batra Rajeev et al. [132]
Utilisation du dispositif	Le dispositif est facile à utiliser, Les informations graphiques et textuelles sont claires, ...	Karacapilidis Nikos Ed [5] ; Dickinger et al. [130]
Valeur sensorielle provoquée	L'interface est attractive visuellement, Les sons sont agréables, ...	Karacapilidis Nikos Ed [5]
Comportement futur	Souhait de renouveler l'expérience, Intention de parler du dispositif, ...	Lessiter Jane et al. [57].

Tableau 4.1 - Synthèse des catégories d'indicateur et leur utilisation possible

Dans cette section, nous avons identifié un ensemble de catégories d'indicateurs d'évaluation des dispositifs numériques. Les indicateurs sont mesurables, par exemple à l'aide de questionnaires auprès des utilisateurs. Il est nécessaire, pour utiliser ces catégories de les définir selon le contexte et d'analyser si chacune est pertinente à utiliser ou non. Cette adéquation entre les catégories et notre contexte est réalisée dans la section suivante.

3.2. Définition des catégories d'indicateur dans le contexte de l'étude des dispositifs dédiés à l'orientation scolaire

Les neuf catégories d'indicateurs présentées dans la section précédente sont analysées ici afin de déterminer la façon de les considérer pour notre contexte.

3.2.1. Utilisateur

La mesure de l'impact des dispositifs numériques se fait pour un type d'utilisateur fixé, les types d'utilisateur sont définies dans la Section 3.2.4 du Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation », dans le module « **Utilisateur** » de l'ontologie. Nous ne scindons pas plus finement les types utilisateur, en leur posant d'autres questions sur leur âge, leur niveau social, etc. Les informations d'impact données aux professionnels de l'orientation sont à l'échelle d'un niveau d'étude (seconde,

première et terminale) et du type de baccalauréat. L'impact mesuré n'a pas, dans cette thèse, de vocation à être personnalisé pour un lycéen.

Ce niveau de considération des types d'utilisateur (par exemple : seconde générale, première technologique, ...) nécessite de supposer que pour un même type d'utilisateur, les capacités techniques à utiliser un dispositif sont proches.

3.2.2. Ressenti global

La catégorie « *ressenti global* » n'est pas conservée dans notre modèle, car elle permet d'évaluer l'impact sur le dispositif et non sur le processus d'orientation.

3.2.3. Valeur perçue du dispositif

Nous ne conservons pas la catégorie « *valeur perçue du dispositif* » par les lycéens, car cette catégorie est utile pour estimer la volonté de continuer à utiliser un dispositif et ce n'est pas l'impact que nous cherchons à évaluer.

Toutefois, nous considérons la valeur perçue pour les différents modules existants dans les dispositifs numériques indépendamment d'un dispositif en particulier. Le terme module est celui utilisé dans l'ontologie que nous proposons. Nous nous intéressons, par exemple à la valeur perçue de la présence de témoignage, à la valeur perçue des outils d'échanges. Les valeurs perçues des éléments des différents modules sont utilisées pour le calcul de similarité entre les dispositifs numériques. Les valeurs perçues permettent par exemple de dire que deux dispositifs se ressemblent plus fortement s'ils ont en commun un élément d'un module avec une valeur perçue forte, plutôt que s'ils ont en commun un élément d'un module avec une valeur perçue faible.

Les valeurs perçues des éléments des modules évoquées dans ce paragraphe sont détaillées dans la Section 2.3.2 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation » et la Section 4.1 du Chapitre « Expérimentations ».

3.2.4. Valeur émotionnelle provoquée (hédonique)

L'importance de l'évaluation de la valeur émotionnelle décrite par des indicateurs hédoniques est prouvée dans de nombreuses études [41], par exemple, en favorisant un apprentissage de qualité, l'acquisition de compétences et un apprentissage plus durable [129].

Les indicateurs hédoniques mesurent l'épanouissement de l'utilisateur (amusement, excitation, et plaisir). Par exemple, si l'évaluation d'un dispositif montre qu'il a principalement une forte valeur hédonique, le dispositif répond à un besoin intangible et suscite une expérience hautement émotionnelle.

Les indicateurs hédoniques sont construits en s'appuyant sur trois études [131]–[133], et ont été mesurés à l'aide de six affirmations :

(H1) Tu as trouvé le dispositif agréable à utiliser,

(H2) Tu as trouvé que le dispositif possède un design moderne,

(H3) Tu as trouvé que le dispositif est ludique (présente des outils d'animations, des vidéos, des images, ...),

(H4) Tu as trouvé le dispositif interactif,

(H5) Tu as trouvé le dispositif facile à utiliser,

(H6) Tu as trouvé le dispositif pratique à utiliser,

3.2.5. Valeur sociale induite

Nous cherchons à mesurer dans cette catégorie si le dispositif contribue à des activités engageantes socialement. Ces activités ont un rôle important dans le processus d'orientation, et les dispositifs numériques favorisent notamment l'échange des lycéens avec leurs pairs [134]. La valeur sociale est mesurée à travers trois aspects en s'appuyant sur l'aspect social traité dans la littérature [129]–[131] : l'apprentissage social (c'est-à-dire, en savoir plus sur les expériences d'autres personnes), le partage d'expérience et la discussion sur le projet d'orientation avec d'autres utilisateurs. Pour cela, elle est mesurée à partir de six affirmations :

(S1) Tu as pu découvrir des expériences d'orientation d'autres personnes (cela peut être leurs sentiments, leurs craintes, leurs projets d'orientations, leurs feedbacks, leurs idées d'orientation),

(S2) Tu as pu partager ton expérience (cela peut être ton sentiment au sujet de ton projet, tes craintes, ton idée d'orientation, tes métiers désirés, tes doutes...),

(S3) Tu as été en mesure de parler de l'orientation avec d'autres utilisateurs,

(S4) Tu as eu envie d'utiliser le dispositif ou certaines de ses fonctionnalités grâce à des personnes que tu connais (amis, famille, ...),

(S5) Tu souhaites discuter avec tes proches (amis, familles, ...) de ce que tu as lu/vu sur le dispositif,

(S6) Tu as plus confiance en ton projet d'orientation après avoir utilisé le dispositif.

3.2.6. Utilité perçue du dispositif

Cette catégorie est destinée à évaluer l'utilité que les lycéens perçoivent du dispositif : est-ce qu'ils ont la sensation d'avoir appris des choses bénéfiques pour leur processus

d'orientation ? L'utilité est directement liée à la réalisation d'objectifs sur l'orientation. Un dispositif dont seule la valeur utilitaire est forte remplit rapidement les objectifs des lycéens sans susciter d'expérience émotionnelle.

Les indicateurs utilitaires sont mesurés à l'aide de six éléments [131], [133] : le système a répondu aux besoins de l'utilisateur rapidement, facilement, de façon exhaustive et personnalisée, le contenu est clair et à jour.

L'utilité perçue du dispositif est évaluée à l'aide de six affirmations :

- (U1) Le dispositif a répondu à tes besoins pour ton projet d'orientation facilement,
- (U2) Le dispositif a répondu à tes besoins pour ton projet d'orientation rapidement,
- (U3) Le dispositif t'a permis de trouver les réponses à toutes tes questions,
- (U4) Le dispositif est adapté et personnalisé pour tes besoins individuels,
- (U5) Le contenu et les fonctions présentées sur le dispositif étaient clairs,
- (U6) Le contenu et les fonctions présentées sur le dispositif étaient à jour.

3.2.7. Utilisation du dispositif et Valeur sensorielle provoquée

Nous faisons le choix de ne pas retenir ces catégories dans les éléments d'évaluation. Elles sont déjà considérées dans notre étude par la présence d'un module dans l'ontologie proposée dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ». Ces catégories sont regroupées dans le module « **Utilisabilité** ». Nous faisons ce choix car il s'agit d'une caractéristique primordiale des dispositifs numériques. De cette utilisabilité dépend la compréhension du contenu disponible et la possibilité d'utiliser sans problème le dispositif. Cette importance est mise en avant grâce au questionnaire donné aux lycéens pour prioriser l'importance des différents éléments d'un dispositif numérique.

L'utilisabilité est accessible par les professionnels de l'orientation dans l'ontologie, elle est utilisée dans le calcul de similarité entre les dispositifs numériques. Ainsi, deux dispositifs avec les mêmes modules seront très différents si l'un a une valeur d'utilisabilité mauvaise et l'autre bonne. De plus, nous avons fait le choix de placer l'utilisabilité dans les modules, car il est indispensable de la considérer même s'il n'est pas possible d'évaluer le dispositif auprès des lycéens. Pour mesurer l'utilisabilité sans que le dispositif ne soit déployé auprès de lycéens, deux solutions sont possibles. La première est de mettre à disposition le dispositif pour un petit nombre de lycéens, la seconde est de former des personnes à étudier l'utilisabilité dans notre contexte. Pour cela, nous nous basons sur le modèle d'évaluation de l'utilisabilité présenté par Martínez-Sala. Ce modèle est présenté lors de la mesure d'utilisabilité de Hello Charly dans la Section 2.1.2.2 du Chapitre « Expérimentations ».

3.2.8. Comportement futur de l'utilisateur

La catégorie du comportement futur de l'utilisateur n'est pas considérée dans notre modèle. Il s'agit d'un élément intéressant, par exemple de savoir ce que le lycéen retient à court et long terme du contenu qu'il a consulté. Toutefois, c'est une catégorie difficilement mesurable, car elle nécessite de suivre un groupe identique de lycéens sur plusieurs mois, et les lycéens ont accès dans le temps à différentes sources d'information sur l'orientation, il n'est donc pas possible de garantir que ce dont se souviennent les lycéens a été acquis lors de l'utilisation du dispositif étudié.

3.2.9. Synthèse des indicateurs utilisés dans notre modèle

Lors de l'état de l'art, neuf catégories d'évaluation des dispositifs numériques sont identifiées, un travail est réalisé afin d'adapter ces catégories à notre problème et de vérifier leur pertinence pour l'impact que nous mesurons.

Ainsi, les catégories « **utilisateur** », « **utilisation du dispositif** » et « **valeur sensorielle provoquée** » sont prises en compte dans la caractérisation des dispositifs numériques par l'ontologie que nous avons proposée, nous ne les réutilisons pas dans ce modèle de mesure de l'impact.

Les catégories « **valeur perçue** », « **ressenti global** » et « **comportement futur** » ne sont pas considérées. Les deux premières catégories sont destinées uniquement à évaluer les dispositifs numériques et n'ont pas de lien avec le processus d'orientation. La troisième catégorie pose des problèmes de mesure.

Les catégories « **valeur hédonique provoquée** », « **valeur sociale induite** » et « **utilité perçue du dispositif** » sont conservées pour construire notre modèle de mesure d'impact sur le processus d'orientation. Chacune des questions est évaluée à l'aide d'une échelle de Likert à sept points, de façon à être cohérent avec la mesure de l'impact décrite dans la Section 3.2.6 du Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ». Ainsi, les trois catégories prennent chacune une valeur entre un et sept, il s'agit de la moyenne des résultats obtenus pour leurs six questions.

Chacune de ces trois catégories ainsi que l'impact KAB est évaluée à l'aide de six questions. Le nombre de questions est réduit afin de conserver l'attention des lycéens lors du questionnaire. De plus, ces questions ont été présentées d'une part à des psychologues de l'éducation nationale pour s'assurer de leur intérêt concernant ce qui est mesuré dans l'impact, d'autre part, à des lycéens pour vérifier leur bonne compréhension des questions.

Les vingt-quatre affirmations présentées sont évaluées sur une échelle de Likert en sept points (l'échelle de réponse est la suivante ; 7 : tout à fait d'accord ; 6 : d'accord ; 5 : plutôt d'accord ; 4 : neutre ; 3 : légèrement en désaccord ; 2 : en désaccord ; 1 : fortement en désaccord). Le choix est fait d'utiliser des affirmations et non des questions de façon à ce que la formulation soit positive et simple afin de suivre les recommandations lors du développement des questionnaires basés sur une échelle de Likert [135].

Dans la section suivante, nous nous inspirons du modèle présenté par Fokkinga et al. pour construire un modèle d'estimation de l'impact à partir des trois catégories présentées dans cette section [7].

3.3. Modèle de relation entre l'impact et les indicateurs d'évaluation des dispositifs numériques

Dans la Section 3.1, nous avons défini l'impact que nous souhaitons mesurer sur le processus d'orientation des lycéens des dispositifs numériques. La Section 3.2 a été dédiée aux indicateurs d'évaluation des dispositifs numériques. Dans cette section, nous souhaitons, à partir de la « Valeur hédonique provoquée », de la « Valeur sociale induite » et de l'« Utilité perçue », proposer une mesure de l'impact. L'objectif de cette estimation est de limiter l'écart entre ce qui est déclaré par les lycéens et ce qui est réel.

Nous posons l'hypothèse que la « Valeur hédonique provoquée », de la « Valeur sociale induite » et de l'« Utilité perçue » sont corrélés à l'impact. Par exemple, si nous mesurons que le dispositif est facile à utiliser, cela indique que la valeur de l'impact est plus élevée que si le dispositif est difficile à utiliser.

Ainsi, nous cherchons à lier la « Valeur hédonique provoquée », de la « Valeur sociale induite » et de l'« Utilité perçue » à l'impact. Pour cela, nous cherchons à déterminer la corrélation entre ces trois éléments et l'impact. Cette recherche de corrélation est représentée dans la **Figure 4.1**. Pour cela, nous utilisons les approches d'analyse de relation entre les variables présentées dans la Section 3.3 du Chapitre « État de l'art ».

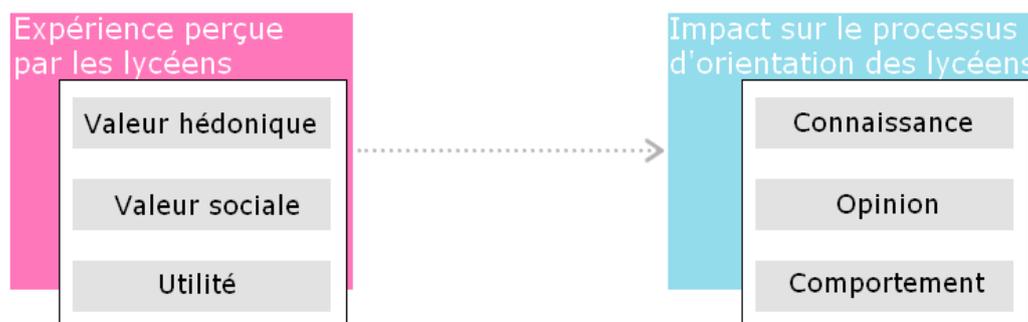


Figure 4.1 - Cadre de l'estimation de l'impact centrée sur l'expérience perçue par les lycéens

Nous décrivons ci-dessous l'estimation de l'impact avec les approches de régression linéaire multiple, de modélisation par équation structurelle et de réseaux bayésiens dans le contexte tel que nous venons de le définir. Le choix de ces trois approches est fait car elles sont utilisées dans le domaine de l'éducation. Les approches de régression linéaire sont utilisées pour étudier des liens entre la performance des élèves et d'autres caractéristiques scolaires [61], [58], [62]. La modélisation par équation structurelle cherche à analyser la satisfaction des élèves dans différents domaines [69], [71], [70]. Les réseaux bayésiens sont utilisés pour des objectifs plus variables comme la probabilité de pérennisation des étudiants dans les programmes de cours à distance ou pour construire des évaluations à destination des étudiants [74], [75].

L'utilisation de régressions linéaires multiples ou d'équations structurelles multiples se rapproche de notre problème. Dans notre cas, ce n'est pas la performance ni la satisfaction mais l'impact sur le processus d'orientation. L'utilisation des réseaux bayésiens est plus éloignée. Toutefois, il est intéressant de connaître selon la « Valeur hédonique provoquée », la « Valeur sociale induite » et l'« Utilité perçue » la probabilité d'obtenir une certaine valeur d'impact.

Ainsi, à partir du Chapitre « État de l'art » et de l'utilisation des trois approches dans le domaine de l'éducation, il n'est pas possible d'en sélectionner une en particulier. Pour cette raison, nous les considérons toutes les trois dans cette section.

3.3.1. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques par une régression linéaire multiple

En reprenant l'équation (EQ1) présentant la résolution sous forme de régression linéaire multiple, nous pouvons écrire le problème de cette façon :

$$\begin{pmatrix} impact_1 \\ \vdots \\ impact_i \\ \vdots \\ impact_n \end{pmatrix} = \begin{pmatrix} 1 & c_{11} & c_{12} & c_{13} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & c_{i1} & c_{i2} & c_{i3} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & c_{n1} & c_{n2} & c_{n3} \end{pmatrix} \begin{pmatrix} \beta \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

c_{i1} (respectivement c_{i2} , c_{i3}) est une valeur entre un et sept, représentant la valeur de la catégorie « **valeur hédonique** » (respectivement « **valeur sociale** », « **utilité** ») pour l'individu i . Le vecteur contenant les coefficients de régression et le vecteur de bruit sont estimés à partir de données connues de l'impact et des trois catégories.

Ainsi, pour obtenir l'impact d'un dispositif à partir des questions sur les 3 catégories, la formule est la suivante :

$$Impact = \beta + \varepsilon_1 + c_1 p_1 + \varepsilon_2 + c_2 p_2 + \varepsilon_3 + c_3 p_3 + \varepsilon_4$$

p_1 , p_2 et p_3 témoignent de l'importance de chacune des catégories. Une valeur élevée montre une grande influence de la catégorie sur la valeur de l'impact.

3.3.2. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques par une modélisation par équation structurelle

Pour savoir si les variables latentes (« Valeur hédonique », « Valeur social », « Utilité ») affectent de manière significative l'impact sur le processus d'orientation des lycéens, la méthode PLS-SEM est utilisée. Le diagramme PLS-SEM est illustré dans la **Figure 4.2**. L'équation du modèle PLS-SEM est présenté ci-dessous :

$$Impact = p_1 Valeur Hédonique + p_2 Valeur Sociale + p_3 Utilité + \epsilon$$

Les hypothèses sont résumées ci-dessous :

- Hypothèse 1 : La valeur hédonique influence significativement l'impact sur les lycéens,
- Hypothèse 2 : La valeur sociale influence significativement l'impact sur les lycéens,
- Hypothèse 3 : L'utilité influence significativement l'impact sur les lycéens,

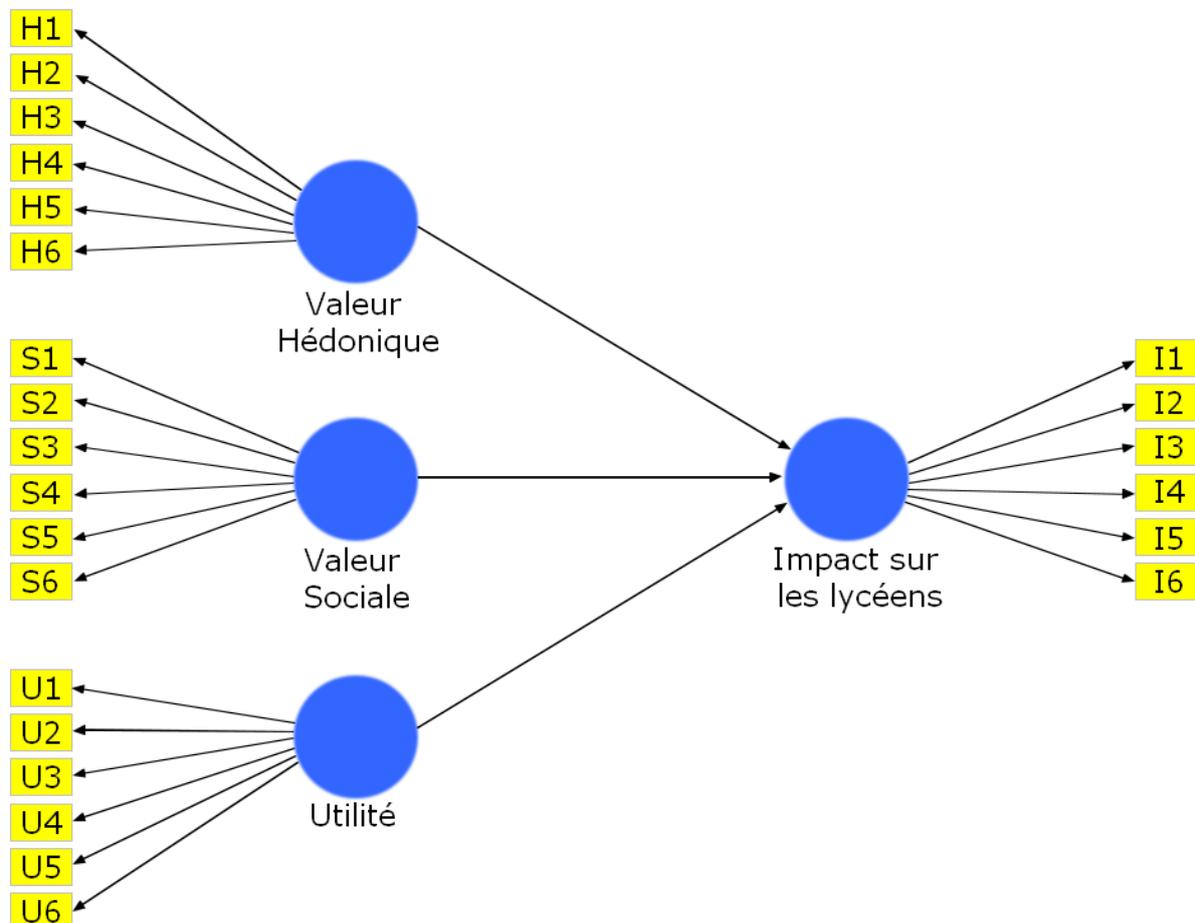


Figure 4.2 - Diagramme de cheminement PLS-SEM

3.3.3. Estimer l'impact sur le processus d'orientation des lycéens des dispositifs numériques avec les réseaux bayésiens

En conservant la même représentation de l'impact sous forme d'une valeur entre un et sept comme utilisé pour la régression linéaire et pour la PLS-SEM, le modèle a besoin d'un très grand nombre de données pour donner un résultat précis. En effet, l'utilisation des réseaux bayésiens nécessite l'existence de plusieurs données pour chaque combinaison. Ainsi, nous réduisons le nombre d'états que chaque nœud peut prendre à deux états : « Bon » si le score est supérieur à quatre, « Faible » s'il est inférieur ou égale à quatre. Ce réseau est présenté dans la **Figure 4.3**.

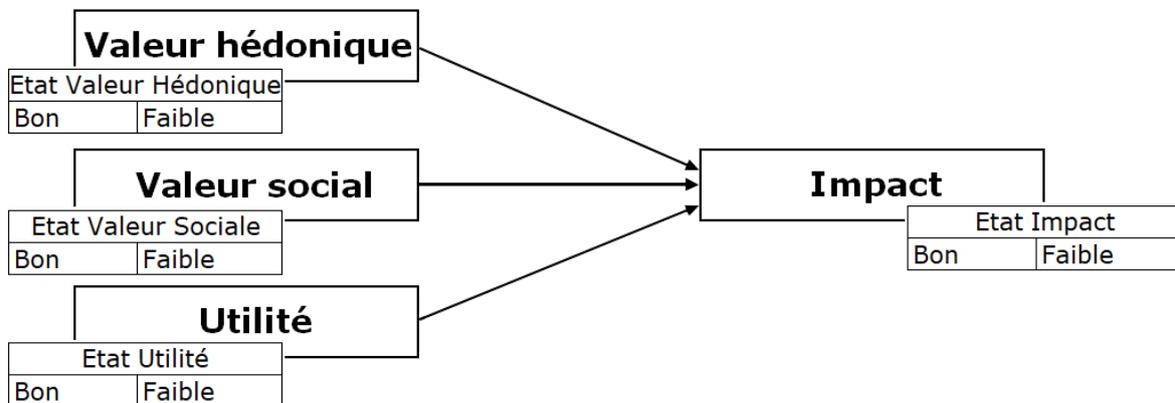


Figure 4.3 - Réseau bayésien représentant les états possibles de chaque nœud

Ce réseau nous permet de calculer la probabilité de chaque état de l'impact (« Bon » ou « Faible ») en sachant l'état de la « Valeur Hédonique », de la « Valeur sociale » et de l'«Utilité» : $p(I = Bon | Valeur Hédonique, Valeur sociale, Utilité)$.

3.3.4. Synthèse

Les modèles construits grâce aux régressions linéaires multiples, à la PLS-SEM ou aux réseaux bayésiens doivent être évalués. Une façon courante de faire est de calculer le coefficient R^2 . Ce coefficient est nécessaire pour évaluer l'aptitude globale du modèle à faire une estimation. Il mesure la corrélation au carré entre les valeurs réelles des résultats et les valeurs prédites par le modèle, il indique dans quelle mesure les variables peuvent être expliquées entre elles [69]. Plus R^2 est élevé, meilleur est le modèle. La quantité de données disponibles va influencer la qualité des résultats obtenus.

Le coefficient R^2 est utilisable dans le cas où l'estimation faite est une valeur continue. S'il s'agit d'une valeur discrète comme ce qui est fait pour les réseaux bayésiens dans notre contexte un autre coefficient doit être utilisé. Un score de précision peut être utilisé pour

évaluer les estimations de valeurs discrètes. Il s'agit du ratio entre le nombre de valeurs biens estimées et le nombre d'estimation total [136].

Les trois modèles que nous venons de présenter nécessitent pour être construits et évalués d'avoir une base de données. Cette base de données doit contenir les réponses aux questions pour les catégories « **valeur hédonique** » ((H1)-(H6)), « **valeur sociale** » ((S1)-(S6)), et « **utilité** » ((U1)-(U6)) et également les réponses aux questions relatives à l'impact ((I1)-(I6)). Les poids des modèles dépendent du type de lycéen, par exemple, la valeur hédonique a une plus grande importance sur l'impact des lycéens en seconde générale/technologique que sur l'impact des lycéens en terminale technologique.

Dans la section suivante, nous présentons le protocole d'estimation de l'impact indépendamment des trois approches d'estimation que nous venons de présenter.

4. Protocole d'estimation de l'impact

Cette section présente le protocole d'estimation de l'impact, son intérêt est de pouvoir être utilisé de la même façon quel que soit le dispositif numérique d'aide à l'orientation.

Si le contexte et l'objectif de la mesure de l'impact n'est pas le même, il est nécessaire de définir l'impact puis d'analyser chacune des catégories présentées dans le **Tableau 4.1** afin de s'assurer de la pertinence de chacune des catégories et de la façon dont les prendre en considération.

Protocole :

(1) Avant l'expérimentation, préparation du questionnaire et du scénario d'utilisation du dispositif numérique :

- Préparation d'un scénario d'utilisation pour les lycéens : le scénario construit est spécifique à chaque dispositif, il doit amener les lycéens à avoir une vision globale du dispositif. Ce scénario doit tenir compte des objectifs du dispositif pour s'assurer que le scénario permette d'utiliser les fonctionnalités principales du dispositif. Les exemples des scénarios pour Hello Charly et pour le portail de l'Onisep sont présentés dans la Section 3.2 du Chapitre « Expérimentations ».
- Préparation du questionnaire : les affirmations associées à la valeur hédonique, à la valeur sociale et à l'utilité perçues par les lycéens présentées dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation » sont utilisées pour construire le questionnaire. Le questionnaire à diffuser aux lycéens existe et est présenté en Annexe B, mais il est nécessaire pour chacune des affirmations de spécifier le nom du dispositif numérique considéré à la place du terme « dispositif »,

- Spécification des besoins auprès des professionnels en relation avec les élèves : il faut spécifier le type de lycéen considéré (seconde générale et technologique, terminale technologique, ...) ; définir sur quelle période de l'année est réalisée la mesure pour savoir quelle est l'étape du processus d'orientation correspondante ; et notifier le besoin d'un ordinateur ou d'un téléphone (selon le dispositif considéré) par élève.

(2) Pendant l'expérimentation, diffusion du scénario et du questionnaire auprès de lycéens :

- Réalisation du scénario par les lycéens, il est nécessaire de leur laisser suffisamment de temps pour effectuer le scénario et de d'intervenir uniquement pour clarifier le scénario mais pas pour aider dans l'utilisation du dispositif. Ce temps permet de s'assurer les lycéens aient utilisé le dispositif avant de répondre au questionnaire.
- Diffusion du questionnaire en étant disponible pour les lycéens pour clarifier les questions si besoin. Le questionnaire peut être donné en papier ou en version numérique. Lors des expérimentations réalisées, le questionnaire a été diffusé à partir de Framaforms (<https://framaforms.org>).

(3) Après l'expérimentation, analyse des données :

- Analyse des données en utilisant les corrélations du modèle présenté dans la Section 3.3 afin d'obtenir une valeur d'impact estimée. Les moyennes des questions pour les trois catégories sont calculées pour pouvoir utiliser le modèle d'équations linéaires multiple ou d'équations structurelles multiple. Dans le cas des réseaux bayésiens, les données sont prétraitées afin que ce ne soient plus des moyennes mais des classes, puis la probabilité la plus élevée permet de donner la classe (« Faible » ou « Bon ») de l'impact perçu du dispositif.

Ce protocole a été utilisé pour deux dispositifs, le portail de l'Onisep et Hello Charly, les spécifications et les résultats pour chacun des dispositifs sont présentés dans la Section 3.2 du Chapitre « Expérimentations ».

5. Conclusion

Ce chapitre présente une méthode d'estimation de l'impact des dispositifs numériques sur le processus d'orientation des lycéens.

Nous avons clarifié la définition de l'impact que nous souhaitons mesurer dans ce projet. Il s'agit ici de l'impact sur le processus d'orientation des lycéens en termes de connaissance, opinion et comportement. Nous avons proposé de mesurer cet impact à l'aide de six affirmations. Toutefois, il est préférable d'utiliser un autre outil de mesure pour limiter l'écart entre ce qui est répondu par le lycéen et ce qui est réel. Pour cela, nous avons synthétisé les différentes catégories d'indicateurs présents dans la littérature pour évaluer les dispositifs numériques. Les catégories d'indicateurs étant très dépendantes du contexte et de ce qui est évalué, nous avons conservé trois catégories : la « **valeur hédonique** », la « **valeur sociale** » et l'« **utilité** », ces catégories sont basées sur ce que perçoivent les lycéens. Pour chaque catégorie, six affirmations sont utilisées afin de déterminer la valeur de la catégorie.

Différentes approches sont possibles pour analyser les relations entre des variables. Les approches de régression linéaire multiple, de PLS-SEM et les réseaux bayésiens permettent d'analyser les relations entre l'impact et les trois catégories identifiées. Contrairement à d'autres approches de type boîte noire, ces trois approches permettent aux professionnels de l'éducation de comprendre la valeur de l'impact et les différents aspects perçus par les lycéens.

Il est nécessaire de construire et de valider ou non ces trois approches à partir d'une base de données dans laquelle les valeurs des trois catégories sont connues ainsi que la valeur de l'impact.

Un protocole est proposé afin que les professionnels de l'éducation puissent mettre en place l'estimation de l'impact qui est proposée dans ce chapitre.

Ce travail aidera les professionnels de l'éducation à alimenter et justifier les discussions sur les choix des dispositifs numériques proposés aux lycéens selon leur type.

MODELE DE PREDICTION DE L'IMPACT DES DISPOSITIFS NUMERIQUES D'AIDE A L'ORIENTATION

TABLE DES MATIERES

1. Introduction.....	82
2. 2. Présentation du fonctionnement du Raisonnement à Partir de Cas (RàPC)	83
2.1. Définition des cas	83
2.2. Cycle du RàPC	85
2.3. Récupération	85
2.3.1. Similarité locale	86
2.3.2. Similarité globale	87
2.3.3. Récupération des cas sources après les calculs de similarité	88

2.4. Réutilisation.....	88
2.5. Révision	89
2.6. Mémorisation	89
2.7. Synthèse et question de recherche	90
3. Notre modèle de prédiction de l'impact de dispositifs numériques.....	90
3.1. Description des cas.....	90
3.2. Propositions de prédiction basées sur le RàPC	95
3.2.1. Similarité locale	96
3.2.2. Modèle de prédiction en suivant le cycle du RàPC.....	98
3.2.2.1. Récupération et réutilisation	98
3.2.2.2. Révision	98
3.2.3. Modèle de Raisonnement à Partir des Attributs (RàPA) inspiré du cycle du RàPC, basée sur la similarité entre chaque descripteur	100
3.2.4. Mémorisation des prédictions du cas cible	102
4. Conclusion	103

1. Introduction

Un grand nombre de dispositifs étant disponibles et d'autres en cours de développement dans le domaine de l'orientation, il est profitable pour les institutions scolaires et les personnels éducatifs de savoir lesquels proposer aux lycéens. Afin d'accompagner les personnels éducatifs dans leur choix, nous avons proposé dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation » une mesure de l'impact des dispositifs numériques sur le processus d'orientation des lycéens. Cet impact est modélisé selon trois aspects, l'impact en termes de connaissance, d'opinion et de comportement.

Toutefois, si le dispositif n'est pas encore déployé, cet impact n'est pas mesurable comme nous l'avons proposé dans le chapitre précédent puisque les lycéens n'y ont pas accès, et, même pour les dispositifs déployés, il peut être délicat de mobiliser des lycéens pour réaliser les mesures.

Pour ces raisons, il est intéressant de pouvoir prédire l'impact des dispositifs numériques sur les lycéens dans leur démarche d'orientation. Les connaissances sur les dispositifs numériques sont structurées à l'aide de l'ontologie proposée dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ».

Dans ce chapitre, notre objectif est de répondre à la question de recherche suivante :

QR3

Comment concevoir un système prédictif pour obtenir l'impact de dispositifs numériques pour lesquels il n'est pas possible de le mesurer et en s'appuyant sur la connaissance d'un nombre limité de dispositifs dont la valeur d'impact est connue ?

Des techniques d'apprentissage automatique peuvent être utilisées pour prédire une classe ou une valeur à partir de données pour lesquelles la classe ou valeur est connue, par exemple, pour prédire l'engagement des étudiants pour utiliser un système d'apprentissage numérique grâce à des arbres de décision [76]. Cependant, ces systèmes requièrent beaucoup de données [137], pour obtenir de bons résultats pour la prédiction. Le nombre de données nécessaire dépend de la dimension des données, plus elle est grande, plus le nombre de données doit être grand. Ainsi, les techniques d'apprentissage automatique ne sont pas utilisables pour les problèmes avec peu de données.

La Section 4 du Chapitre « État de l'art » met en avant l'intérêt du RàPC dans notre contexte par rapport aux approches de représentation et raisonnement sur des connaissances à base logique, aux approches de représentation et raisonnement sur des connaissances incertaines et aux approches orientées données. En effet, les solutions prédites au fur et à mesure permettent d'augmenter la base de connaissances, des prédictions sont possibles même avec une faible quantité de données, et les informations des experts sont disponibles.

Afin de mettre en place une approche de RàPC répondant aux besoins de notre contexte, nous décrivons dans une première section le fonctionnement du RàPC. Dans la deuxième section, nous présentons notre modèle de prédiction de l'impact.

2. Présentation du fonctionnement du Raisonnement à Partir de Cas (RàPC)

Le RàPC est un paradigme de résolution de problème incrémental utilisant les expériences passées [138] pour résoudre de nouveaux problèmes. Il est basé sur l'hypothèse que les problèmes similaires ont des solutions similaires.

Nous abordons dans cette section la définition d'un cas et les étapes du cycle du RàPC.

2.1. Définition des cas

Nous nous basons sur la notation de Lieber pour définir les cas [139].

Un cas est la caractérisation d'un problème selon deux axes, la description du problème et sa solution. Deux types de cas existent, les cas sources, notés $(srce ; Sol(srce))$ et les cas cibles, notés $(tgt ; Sol(tgt))$. Pour les cas sources, la description du problème $(srce)$ et sa solution $Sol(srce)$ sont connues. Seule la description du problème (tgt) est connue pour les cas cibles, la solution est inconnue $(Sol(tgt))$.

Les cas sont majoritairement représentés de façon structurée, pour cela, il est nécessaire de les représenter de façon uniforme à l'aide de descripteurs et de connaître tous les descripteurs pour chaque cas [140].

La description du problème est un ensemble d'éléments que nous appelons descripteurs. Un descripteur se compose d'un attribut et d'une valeur.

Autrement dit, pour un problème donné, considérons deux cas source A et B :

Le cas A s'écrit : $((a_1 = v_1^A), \dots, (a_m = v_m^A) ; Sol(A))$,

et le cas B : $((a_1 = v_1^B), \dots, (a_m = v_m^B) ; Sol(B))$.

Le problème considéré possède m descripteurs, les attributs des descripteurs sont notés a et les valeurs des descripteurs v . Un exemple d'attribut pour un descripteur est l'âge, et une valeur possible pour ce descripteur est 37. Le **Tableau 5.1** présente un exemple de la structure d'un cas adapté à la prédiction d'une valeur.

	Attribut a	Valeur v
<i>srce</i>	Age	37
	Genre	F
	Fréquence d'entraînement par semaine	5
	Fréquence cardiaque au repos	Basse
	Catégorie lors du début de pratique	Senior
<i>Sol(srce)</i>	Temps au km	5,12

Tableau 5.1 - Exemple d'une structure d'un cas adapté à la prédiction

Dans cet exemple, le cas est représenté sous forme d'un vecteur et possède des attributs de différents types (numériques, booléens et catégoriels).

La structure du cas a un impact sur les étapes du RàPC. Dans la section suivante, nous présentons les différentes étapes du RàPC.

2.2. Cycle du RàPC

Le cycle du RàPC contient plusieurs étapes dont le nombre est variable selon les travaux. Nous présentons ici le cycle le plus couramment utilisé, il comporte 4 étapes (1) Récupération, (2) Réutilisation, (3) Révision, et, (4) Mémorisation ; que nous présentons ci-dessous et illustrées dans la **Figure 5.1**.

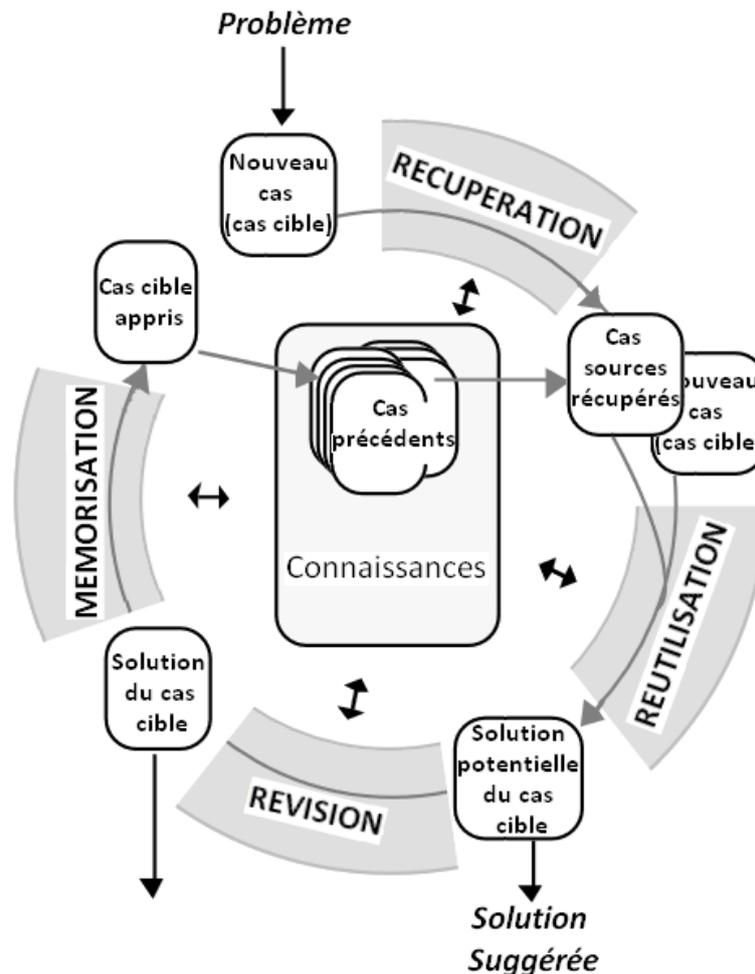


Figure 5.1 - Cycle du RàPC selon Aamodt et al. [138]

2.3. Récupération

L'étape de récupération vise à retrouver les cas sources les plus ressemblants au cas cible en considérant les descripteurs des cas. Les cas sources récupérés jouent un rôle majeur, car leur solution est utilisée pour obtenir la solution du cas cible. Ainsi, il s'agit d'une étape primordiale dans le cycle du RàPC.

Un calcul de similarité est majoritairement utilisé pour identifier les cas sources. Pour cette raison, nous nous interrogeons sur la façon dont la similarité peut être mesurée dans le processus de RàPC.

De nombreuses mesures de similarité existent. Nous évoquons différentes techniques permettant de calculer une similarité en RàPC afin d'identifier la technique appropriée à notre problème d'identification de cas similaires. Dans cette section, nous abordons différentes mesures utilisées dans la littérature.

Le calcul de similarité peut être réalisé de manière **globale** ou **locale**. S'il s'agit d'une mesure globale, une valeur de similarité est obtenue entre les cas, si elle est locale, la similarité entre chaque descripteur du cas est calculée avant de pouvoir obtenir la similarité globale.

2.3.1. Similarité locale

Une grande variété de similarités locales existe, elles sont dépendantes du type de l'attribut. Nous présentons un aperçu des mesures présentes dans la littérature selon le type d'attributs.

Dans la suite du paragraphe, considérons A et B , deux cas pour lesquels nous souhaitons mesurer la similarité, soient v_1^A (respectivement v_1^B) la valeur de l'attribut a_1 pour le cas A (respectivement B).

Attribut de type booléen (les valeurs de l'attribut à comparer peuvent prendre deux états, par exemple vrai ou faux)

$$(EQ2) \text{ sim}_{loc}(a_1^A, a_1^B) = \begin{cases} 1 & \text{si } v_1^A = v_1^B \\ 0 & \text{sinon} \end{cases}$$

Cette mesure est utilisée par exemple pour comparer en médecine la dimension de "Suspicion d'invasion" d'un cancer du sein [141] ou encore en droit le "Genre" [142]. Dans ces deux exemples, un cas est un individu.

Attribut de type numérique (les valeurs de l'attribut à comparer sont de la forme d'un nombre réel (\mathbb{R}))

$$(EQ3) \text{ sim}_{loc}(a_1^A, a_1^B) = 1 - \frac{|v_1^A - v_1^B|}{range}$$

$range(v_1)$ est l'intervalle maximal entre les valeurs de v_1 .

Cette mesure est utilisée en médecine par exemple, pour la valeur d'un antigène [141] ou encore l'âge d'individus [142].

Attribut de type chaîne de caractères (les valeurs de l'attribut à comparer peuvent prendre une suite ordonnée de caractères)

Maedche et al. [143] proposent une solution possible pour calculer la similarité entre deux chaînes de caractère :

$$sim_{loc}(a_1^A, a_1^B) = \max \left(0, \frac{\min(|v_1^A|, |v_1^B|) - ed(v_1^A, v_1^B)}{\min(|v_1^A|, |v_1^B|)} \right)$$

où, $ed(v_1^A, v_1^B)$ est la distance de Levenshtein. Il s'agit du minimum d'insertions, de délétion et de substitution à effectuer pour passer d'une chaîne de caractères à l'autre.

Attribut de type catégoriel (les valeurs de l'attribut à comparer se placent dans une hiérarchie ou dans un corpus)

Différentes mesures existent pour comparer les valeurs catégorielles [144] ; par exemple, il est possible de considérer la longueur du chemin reliant les valeurs dans une taxonomie. Comparer la valeur informationnelle (fréquence d'apparition dans un corpus) des deux valeurs est une autre solution.

Lorsque la similarité locale entre chacun des attributs est calculée, il est nécessaire d'agréger l'ensemble des similarités afin d'obtenir une valeur de similarité globale.

2.3.2. Similarité globale

Comme pour la similarité locale, différentes mesures existent.

Considérons $sim_{loc}(A, B)$, le vecteur contenant les similarités locales entre les cas A et B .

$$(EQ4) \quad sim(A, B) = \sum_i w_i * sim_{loc}(A, B)_i$$

w_i est le poids donné à l'attribut a_i .

Cette mesure, utilisée par exemple dans le domaine industriel pour la gestion des risques des projets de construction [145], a l'avantage d'utiliser des poids pour considérer l'importance de chaque attribut. Le poids est donné par des experts, soit directement [142], [146], soit calculé à partir d'un questionnaire [145], ou peut être obtenue à partir d'algorithmes [141], [147].

Dans le cas où toutes les caractéristiques sont de type numérique, il est possible d'obtenir directement la similarité globale sans passer par les similarités locales grâce par exemple à la similarité cosinus.

Similarité cosinus :

Soit v_A (respectivement v_B), le vecteur des caractéristiques du cas A (respectivement B).

$$sim_{cos}(A, B) = \frac{v_A \cdot v_B}{\|v_A\| \cdot \|v_B\|}$$

Cette mesure est par exemple utilisée dans la prédiction de consommation d'électricité [148].

2.3.3. Récupération des cas sources après les calculs de similarité

Une fois la/les mesure(s) de similarités définie(s) pour un problème donné, il faut identifier les cas à récupérer pour l'étape de réutilisation.

Pour cela, il existe plusieurs possibilités. (1) Une façon de faire est de fixer un seuil de similarité [142],[147],[148]. Ainsi, tous les cas sources ayant une similarité avec le cas cible supérieure au seuil sont récupérés. Cela garantit que les cas sources sont utilisés seulement s'ils sont assez proches du cas cible. Cela peut rendre le temps de calcul lors de l'étape de réutilisation long s'il y a un grand nombre de cas utilisés, et il faut savoir à partir de quel seuil de similarité les cas sont assez proches pour avoir des solutions similaires. (2) Une autre solution est de fixer un nombre k de cas sources à utiliser pour l'étape suivante. Dans cette situation, le fonctionnement s'approche de celui d'un algorithme de k plus proches voisins [141]. Les k cas sources les plus similaires au cas cible sont utilisés dans l'étape de réutilisation. Le choix du nombre k entraîne le risque d'inclure des cas sources qui ne sont pas proches du cas cible, mais permet de ne pas avoir un grand nombre de cas pour l'étape suivante.

Dans cette section, nous avons présenté l'étape de récupération dans le RàPC. L'étape de récupération est couramment réalisée en mesurant la similarité entre les cas. Pour cette raison, nous avons énoncé les mesures de similarité les plus utilisées dans les problèmes de RàPC. Ces mesures étant dépendantes du type de caractéristiques stockées dans les cas, nous traitons du choix de la mesure pour notre problème dans la Section 3 de ce chapitre.

Lorsque les cas sources les plus similaires au cas cible sont identifiés, l'étape de réutilisation est effectuée.

2.4. Réutilisation

Cette étape utilise les cas sources récupérés pour déduire une solution [85].

La réutilisation peut être simple, si la ou les solution(s) retrouvée(s) est reprise directement comme solution du nouveau problème. C'est le cas pour les tâches de classification avec un nombre limité de classes et un grand nombre de cas dans la base de cas. Dans un tel scénario, chaque solution est généralement contenue dans la base de cas [149].

Dans les tâches de classification, un vote peut également être fait. La solution pour l'étape de réutilisation est alors la classe majoritaire parmi les cas sources récupérés.

Si le problème est un problème de régression, une moyenne ou moyenne pondérée peut être calculée à partir des cas les plus similaires afin d'obtenir la solution de cette étape pour le cas cible [85].

Si les solutions au problème ne sont pas des classes ou des valeurs numériques, la combinaison des solutions sources récupérées peut être plus complexe.

2.5. Révision

La révision permet d'adapter la solution du cas cible pour mieux lui correspondre. Différents types d'adaptation existent [85].

- La substitution, elle consiste à remplacer une ou des parties de la solution,
- La transformation, elle permet d'ajouter, modifier ou supprimer des éléments de la solution,
- L'adaptation générative, une nouvelle solution est générée.

Cette étape est réalisée avec un expert, ou par apprentissage des connaissances de révision à partir des cas de la base de cas, ou de connaissances du domaine.

Les étapes de réutilisation et de révision ne sont pas toujours dissociées. Il est difficile de réutiliser directement les solutions sources si le cas cible est très différent des cas sources. Dans d'autres utilisations du RàPC, l'étape de réutilisation est faite seule, sans révision.

2.6. Mémorisation

Lorsque la solution du cas cible est obtenue, se pose la question de la mémorisation de ce nouveau cas en tant que cas source. Cette étape est un atout du RàPC puisque la mémorisation permet d'enrichir les connaissances de l'algorithme. En effet, une base de cas trop petite est susceptible de ne pas contenir de cas sources similaires aux cas cibles [85].

Il n'est pas pertinent d'enregistrer tous les nouveaux cas cibles dont la solution a été trouvée, car cela risque de surcharger la base de cas et rendre plus longue l'étape de récupération. Les nouveaux cas doivent être ajoutés à la base de cas uniquement s'ils apportent des connaissances, par exemple si les valeurs de leurs caractéristiques sont éloignées de celles présentes dans la base de cas.

Cette étape peut aussi permettre de maintenir la base de cas en vérifiant que les cas en mémoire sont toujours pertinents, s'il faut en supprimer ou en remplacer. Un expert peut réaliser cette étape.

Malgré son importance, cette étape est peu décrite dans la littérature lors de l'utilisation concrète du RàPC [150].

2.7. Synthèse et question de recherche

Dans cette section, nous avons présenté le fonctionnement du RàPC, la façon dont sont définis les cas et les quatre étapes du cycle. Selon le contexte et le type de données représentant les cas, les étapes de récupération, réutilisation et révision sont variables. L'utilisation du RàPC peut être problématique si le nombre de cas source est très faible, car il est alors possible que le cas cible n'ait pas de cas source similaire. Le processus de révision peut pallier ce problème en modifiant la solution prédite. Toutefois, dans la littérature, cette étape est peu utilisée et présentée car elle est dépendante du contexte.

Nous présentons, dans la section suivante l'application du RàPC au problème de prédiction de l'impact de dispositifs numériques d'aide à l'orientation des lycéens afin de répondre à la question de recherche :

QR3

Comment concevoir un système prédictif pour obtenir l'impact de dispositifs numériques pour lesquels il n'est pas possible de le mesurer et en s'appuyant sur la connaissance d'un nombre limité de dispositifs dont la valeur d'impact est connue ?

3. Notre modèle de prédiction de l'impact de dispositifs numériques

Pour décrire ce modèle, nous abordons tout d'abord la formalisation des cas puis les propositions réalisées pour prédire l'impact. Les mesure similarités de similarité sont présentées. Elles sont communes à nos différentes propositions, puis les étapes de révision. Enfin nous discutons de la mémorisation des nouveaux cas dans la base de cas.

3.1. Description des cas

Une représentation uniforme des dispositifs numériques à l'aide d'une ontologie est présentée dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ». Nous obtenons, à partir de cette représentation, une structure de cas comportant la description du problème et sa solution. La description du problème possède cinq modules, (1) « **Nature de l'information** », (2) « **Services additionnels** », (3) « **Interaction de l'utilisateur avec le dispositif** », (4) « **Utilisabilité** », (5) « **Utilisateur** ». Cette structure en cinq modules provient des cinq modules de l'ontologie proposée. La solution du problème est l'impact des dispositifs numériques sur les lycéens dans leur démarche d'orientation.

Nous détaillons les descripteurs de chacun de ces modules et la solution du problème en spécifiant l'attribut et son type de valeur. La colonne d'indexation de l'attribut est utilisée pour faciliter la compréhension des formules utilisées plus tard dans ce chapitre :

- (1) « **Nature de l'information** » : traduit les sujets abordés par le dispositif et la façon dont ils le sont (à l'aide de texte, de vidéo, ...). Ce module est lui-même scindé en neuf parties. Chacune de ces parties se compose de deux attributs, l'un est de type booléen, l'autre de type catégoriel. Les attributs et le type des valeurs qu'ils peuvent prendre sont présentés dans le **Tableau 5.2**. Le module de l'ontologie correspondant est présenté dans la **Figure 5.2**.

Parties	Indexation de l'attribut a_i	Attribut	Type de valeur
À propos	1	Sujet traité	Booléen \in {Vrai, Faux}
	2	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Diplôme postbac	3	Sujet traité	Booléen \in {Vrai, Faux}
	4	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Emploi et stage	5	Sujet traité	Booléen \in {Vrai, Faux}
	6	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Établissement postbac	7	Sujet traité	Booléen \in {Vrai, Faux}
	8	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Lycée	9	Sujet traité	Booléen \in {Vrai, Faux}
	10	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Métier	11	Sujet traité	Booléen \in {Vrai, Faux}
	12	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Orientation	13	Sujet traité	Booléen \in {Vrai, Faux}
	14	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Psychologie	15	Sujet traité	Booléen \in {Vrai, Faux}
	16	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}
Vie étudiante	17	Sujet traité	Booléen \in {Vrai, Faux}
	18	Composant utilisé	Catégoriel \in {Texte, Image, Vidéo, Audio}

Tableau 5.2 - Attributs des cas pour le module « **Nature de l'information** »

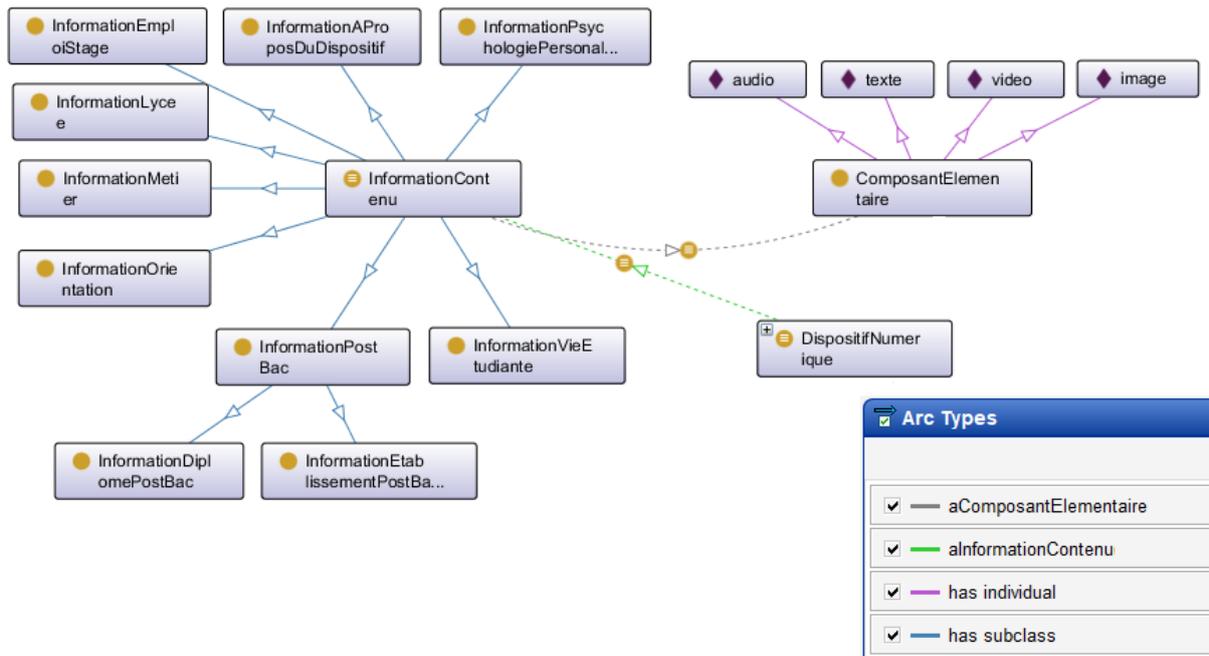


Figure 5.2 - Capture d'écran de Protégé du module « **Nature de l'information** » de l'ontologie proposée

(2) « **Services additionnels** » : contient les outils ou informations particulières auxquels l'utilisateur peut accéder. Ce module est scindé en cinq parties et est présenté dans le **Tableau 5.3**. Le module de l'ontologie correspondant est présenté dans la **Figure 5.3**, tous les attributs ne sont pas visibles sur cette figure car Protégé ne permet pas de représenter visuellement les propriétés des classes.

Parties	Indexation de l'attribut a_i	Attribut	Type de valeur
Témoignage	19	Présence	Booléen ∈ {Vrai, Faux}
	20	Nécessite authentification	Booléen ∈ {Vrai, Faux}
FAQ	21	Présence	Booléen ∈ {Vrai, Faux}
	22	Nécessite authentification	Booléen ∈ {Vrai, Faux}
Annuaire	23	Présence	Booléen ∈ {Vrai, Faux}
	24	Nécessite authentification	Booléen ∈ {Vrai, Faux}
Outil d'échange	25	Présence	Booléen ∈ {Vrai, Faux}
	26	Nécessite authentification	Booléen ∈ {Vrai, Faux}
	27	Permet instantanéité	Booléen ∈ {Vrai, Faux}
	28	Type interlocuteur	Catégoriel ∈ {Lycéen, Étudiant, Professeur de lycée, Professeur postbac, Conseiller d'orientation, Professionnel, Robot}
	29	Visioconférence	Booléen ∈ {Vrai, Faux}

Outil de test	30	Présence	Booléen ∈ {Vrai, Faux}
	31	Nécessite authentification	Booléen ∈ {Vrai, Faux}
	32	Sujet du test	Catégoriel ∈ {Métier, Psychologie et personnalité, Centre d'intérêt}

Tableau 5.3 - Attributs des cas pour le module « Services Additionnels »

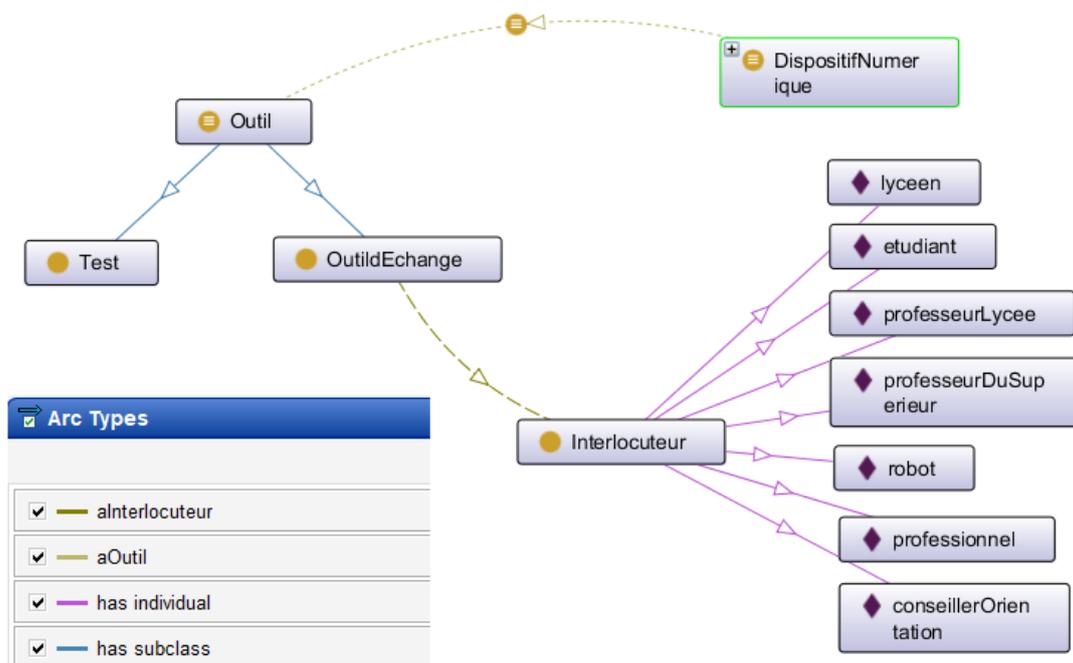


Figure 5.3 - Capture d'écran de Protégé du module Services additionnels de l'ontologie proposée

(3) « **Interaction de l'utilisateur avec le dispositif** » : indique la façon dont l'utilisateur utilise le dispositif, c'est-à-dire le support utilisé et la façon dont l'interface est construite. Les attributs de cette sous-partie sont décrits dans le **Tableau 5.4**.

Partie	Attribut	Indexation de l'attribut a_i	Type de valeur
Support numérique	Support physique	33	Catégoriel ∈ {Ordinateur, Téléphone}
	Type d'application	34	Catégoriel ∈ {Application mobile, Application ordinateur, Site web}
Interface Utilisateur	Disponibilité	35	Booléen ∈ {Temporaire, Permanent}
	Type Interface	36	Catégoriel ∈ {Interface par page, Monde immersif}

Tableau 5.4 - Attributs des cas pour le module « Interaction entre le dispositif et l'utilisateur »

(4) « **Utilisabilité** » : représente le degré selon lequel un produit peut être utilisé, ce module se compose de six attributs présentés dans le **Tableau 5.5**. Les valeurs sont

dans l'intervalle $[0 ; 3]$, une valeur de « zéro » signifie que l'attribut évalué est « pauvre », une valeur de « trois » signifie que l'attribut évalué est « excellent ».

Partie	Attribut	Indexation de l'attribut a_i	Type de valeur
Utilisabilité	Navigabilité	37	Numérique $\in [0 ; 3]$
	Contenu	38	Numérique $\in [0 ; 3]$
	Représentation graphique	39	Numérique $\in [0 ; 3]$
	Ergonomie	40	Numérique $\in [0 ; 3]$
	Image de Marque	41	Numérique $\in [0 ; 3]$
	Utilisabilité globale	42	Numérique $\in [0 ; 3]$

Tableau 5.5 - Attributs des cas pour le module « **Utilisabilité** »

(5) Utilisateur : décrit le type de lycéen qui utilise le dispositif, les trois attributs de ce module sont décrits dans le **Tableau 5.6**.

Partie	Attribut	Indexation de l'attribut a_i	Type de valeur
Caractéristiques Utilisateur	Année	43	Catégoriel $\in \{\text{Seconde, Première, Terminale}\}$
	Type de baccalauréat	44	Catégoriel $\in \{\text{Général, Technologique, Professionnel}\}$
	Étape dans le processus d'orientation	45	Catégoriel $\in \{\text{Exploration, Exploration approfondie, Choix, Adaptation}\}$

Tableau 5.6 - Attributs des cas pour le module « **Utilisateur** »

(Solution du problème) « Impact » : il s'agit de la valeur de l'impact sur le processus d'orientation des lycéens du dispositif numérique. L'impact est une valeur numérique comprise dans l'intervalle $\in [1 ; 7]$. Cet intervalle correspond à l'échelle de Likert qui est utilisé pour mesurer l'impact pour les dispositifs pour lesquels l'impact est mesurable. Ces dispositifs servent de cas source.

Nous venons, dans cette section de proposer la structure des cas grâce à l'instanciation de l'ontologie présentée dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ». Les attributs des cas sont formalisés dans l'ontologie par une classe, une propriété de classe, ou une relation. Dans notre problème, les cas sont décrits par quarante-cinq attributs, les attributs sont de type booléen, catégoriel et numérique. La solution des cas est de type numérique, elle représente la valeur de l'impact sur le processus d'orientation des lycéens.

Les cas sont stockés de façon linéaire dans notre base de cas. Il s'agit du stockage de cas le plus fréquent en RàPC, tous les cas sont au même niveau. L'autre solution est d'organiser hiérarchiquement les cas [140].

La structure des cas étant décrite, nous présentons deux propositions basées sur le RàPC permettant de prédire l'impact sur le processus d'orientation des lycéens de dispositifs numériques.

La première proposition est un modèle suivant le cycle du RàPC dans lequel nous proposons un paramètre de révision permettant d'obtenir la solution prédite du cas cible lors de l'étape de révision.

La seconde proposition est une prédiction du cas cible à partir des attributs. Cette proposition s'inspire du cycle du RàPC mais en raisonnant à l'échelle des attributs et non à celle des cas.

Ces deux propositions sont détaillées dans la section suivante.

3.2. Propositions de prédiction basées sur le RàPC

Les deux propositions sont basées sur les mêmes mesures de similarité locales. Dans cette section, nous présentons d'abord la méthode de calcul de similarité locale. Puis nous détaillons chaque contribution : un modèle de calcul en suivant le cycle du RàPC et un modèle de Raisonnement à Partir des Attributs (RàPA).

La similarité locale entre deux cas est l'ensemble des similarités entre chaque descripteur du cas.

Pour plus de fluidité, dans ce paragraphe, nous utilisons les notations suivantes :

- A est le cas source obtenu à l'étape de récupération, sa solution est notée $Sol(A)$,
- B est un cas source obtenu à l'étape de récupération, sa solution est notée $Sol(B)$,
- C est le cas cible, sa solution que nous souhaitons prédire est notée $Sol(C)$.

Exemple :

Nous présentons dans les sections suivantes un exemple des deux propositions. Nous nous plaçons dans un problème avec les cas fictifs suivants qui se composent de trois attributs :

Attribut a_i	Poids de l'attribut w_i	Cas A	Cas B	Cas C
a_1	0,2	"Vrai"	"Vrai"	"Faux"
a_3	0,5	"Faux"	"Vrai"	"Vrai"
a_5	0,3	"Vrai"	"Faux"	"Vrai"
Solution	/	$Sol(A) = 4$	$Sol(B) = 5$	$Sol(C)$

3.2.1. Similarités locales

Les cas possèdent des attributs de trois types différents, des attributs de type booléen, de type numérique et de type catégoriel (voir Section 3.1).

Considérons A et B , deux cas pour lesquels nous souhaitons calculer la similarité locale de leurs attributs. Leur similarité locale notée $Sim_{loc}(A, B)$, est un vecteur de dimension m , où m est le nombre d'attributs dans la description du problème.

Si l'attribut i est de type **booléen**, $Sim_{loc}(A, B)_i$ est calculée à partir de l'équation (EQ2).

Si l'attribut i est de type **numérique**, $Sim_{loc}(A, B)_i$ est calculée à partir de l'équation (EQ3).

Si l'attribut i est de type **catégoriel**, $Sim_{loc}(A, B)_i$, est calculée à partir de la fonction $SimilariteCategoriel()$.

Cette fonction est présentée dans la **Figure 5.4**. Elle utilise les matrices de similarités présentées dans la **Figure 5.5**. L'utilisation d'une taxonomie n'ayant pas de sens pour nos catégories, nous utilisons des matrices de similarité entre tous les éléments possibles de nos catégories. Pour construire ces matrices, nous avons travaillé avec des lycéens. Le protocole pour obtenir ces matrices est détaillé dans la Section 4.2 du Chapitre « Expérimentations ». Il s'agit de matrices symétriques.

Exemple du calcul de similarité catégoriel :

Nous cherchons à calculer la similarité locale de l'attribut « Interlocuteur ». Pour cet attribut, la valeur du cas A est « étudiant », et est « lycéen » pour le cas B . Nous utilisons la matrice $S_{Interlocuteur}$ de la **Figure 5.5**, et nous avons $S_{Interlocuteur}[\text{étudiant}, \text{lycéen}] = 0,4$.

Nous en déduisons que la similarité locale pour cet attribut est de 0,4.

Nous nous sommes placés dans un exemple où les cas ont une unique catégorie, mais il est possible qu'il y en ait plusieurs. Par exemple, pour le module « **Nature de l'information** », les informations sur la vie étudiante peuvent prendre la forme de texte et d'image pour un

dispositif numérique donné. Dans cette configuration, la valeur de son attribut est {texte, image}.

```
def SimilariteCategoriel(liste_elmtA,liste_elmtB,M_simi):
    """Entrée de la fonction :
    liste_elmtA : la liste des catégories du cas A pour l'attribut considéré,
    liste_elmtB : la liste des catégories du cas B pour l'attribut considéré,
    M_simi : La matrice de similarité correspondant à l'attribut considéré.

    Sortie de la fonction :
    somme_simi/indice : valeur comprise dans l'intervalle [0,1],
    il s'agit de la similarité moyenne entre Les cas A et B pour l'attribut considéré,
    """

    somme_simi=0
    indice=0
    liste_elmtB_traites=[]

    """Tous les éléments du cas A sont considérés, pour chacun d'entre eux, la similarité avec l'élément du cas B
    donnant la similarité maximale est prise.
    """
    for i in liste_elmtA:
        somme_simi+=max(M_simi[i][j] for j in liste_elmtB)
        liste_elmtB_traites.append(M_simi[i].index(max(M_simi[i][j] for j in liste_elmtB)))
        indice+=1

    """Les éléments du cas B non considérés dans la boucle précédente sont utilisés,
    pour chacun d'entre eux, la similarité avec l'élément du cas A
    donnant la similarité maximale est prise.
    """
    liste_elmtB_nontraites=[e for e in liste_elmtB if e not in liste_elmtB_traites]

    for i in liste_elmtB_nontraites:
        somme_simi+=max(M_simi[i][j] for j in liste_elmtA)
        indice+=1

    """Retourne la similarité moyenne"""
    return somme_simi/indice
```

Figure 5.4 - Algorithme de la fonction *SimilariteCategoriel()*

$$S_{Interlocuteur} = \begin{matrix} & \begin{matrix} lycéen & étudiant & professeur lycée & professeur supérieur & professionnel & conseiller orientation & robot \end{matrix} \\ \begin{matrix} lycéen \\ étudiant \\ professeur lycée \\ professeur supérieur \\ professionnel \\ conseiller orientation \\ robot \end{matrix} & \begin{pmatrix} 1 & & & & & & \\ 0,4 & 1 & & & & & \\ 0 & 0 & 1 & & & & \\ 0 & 0 & 0,6 & 1 & & & \\ 0 & 0,4 & 0,4 & 0,4 & 1 & & \\ 0 & 0 & 0,6 & 0 & 0 & 1 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

$$S_{SupportInformation} = \begin{matrix} & \begin{matrix} vidéo & texte & image & audio \end{matrix} \\ \begin{matrix} vidéo \\ texte \\ image \\ audio \end{matrix} & \begin{pmatrix} 1 & & & \\ 0,4 & 1 & & \\ 0,4 & 0,2 & 1 & \\ 0,6 & 0,4 & 0,4 & 1 \end{pmatrix} \end{matrix}$$

$$S_{SujetTest} = \begin{matrix} & \begin{matrix} métiers & personnalité & centres intérêt \end{matrix} \\ \begin{matrix} métiers \\ personnalité \\ centres intérêt \end{matrix} & \begin{pmatrix} 1 & & \\ 0,2 & 1 & \\ 0,6 & 0,8 & 1 \end{pmatrix} \end{matrix}$$

Figure 5.5 - Matrices de similarité pour les attributs de type catégoriel

Les mesures de similarité étant fixées, nous présentons nos deux propositions de prédiction de la solution.

3.2.2. Modèle de prédiction en suivant le cycle du RàPC

Dans cette proposition, nous décrivons le déroulement de chacune de ses étapes, qui sont celles du RàPC.

3.2.2.1. Récupération et réutilisation

Les similarités locales entre les cas sources et le cas cible sont calculées comme présenté dans la Section 3.2.1. Ainsi, nous avons m vecteurs de similarités locales, correspondants aux similarités locales entre le cas cible et les m cas sources de la base de cas.

À partir de ces similarités locales sont calculées les similarités globales en utilisant l'équation (EQ4). Ce calcul de similarité globale utilise une pondération pour chacune des dix-sept parties. Les dix-sept parties sont celles présentées dans la Section 3.1 en excluant la partie « *Caractéristiques de l'Utilisateur* ». La pondération permet de donner plus ou moins d'importances à certains attributs dans le calcul de similarité. Nous avons obtenu ces pondérations à partir de lycéens.

À l'issue des calculs de similarité globale, nous obtenons m valeurs de similarité globale. La similarité globale est une valeur comprise dans l'intervalle $[0 ; 1]$.

Le cas source ayant la valeur de similarité globale la plus élevée est récupéré. Cette étape permet d'aboutir à une solution potentielle pour le cas cible. Une solution potentielle est la solution du cas cible qui n'a pas encore été révisé, elle est issue des cas sources directement sans avoir été retravaillé. La solution potentielle de notre cas cible à la fin de cette étape est la solution du cas source le plus similaire. Cette solution potentielle est révisée dans l'étape suivante.

3.2.2.2. Révision

Les étapes précédentes permettent d'obtenir une solution potentielle de $Sol(C)$ issue directement de la base de cas. L'étape de révision a pour objectif de retravailler la solution potentielle en la diminuant ou en l'augmentant. Son intérêt est d'améliorer la solution estimée en considérant les spécificités du cas cible. Le principe de cette révision est d'augmenter la valeur après l'étape de réutilisation de $Sol(C)$ s'il dispose de plus d'éléments que le cas A , ou de diminuer cette valeur s'il a des éléments en moins.

Pour effectuer cette révision, nous posons l'hypothèse que si un attribut témoigne de la présence d'un élément, cela signifie que cet élément est fonctionnel.

La révision s'effectue en modifiant la solution du cas cible selon un paramètre de révision PR . PR est obtenu en comparant les attributs du cas cible (C) et du cas source récupéré (A). Ces deux cas sont

- (1) comparés au niveau des attributs booléens témoignant de l'absence ou de la présence d'informations et de services additionnels,
- (2) comparés au niveau l'attribut d'utilisabilité.

L'algorithme calculant le Paramètre de Révision est présenté dans la **Figure 5.6**.

<p>Entrée $srce_A$: couples (attribut, valeur) du cas source A, il s'agit du cas le plus similaire à C, tgt_C : couples (attribut, valeur) du cas cible C, w : vecteur de poids des attributs.</p>
<p>Paramètre de Révision (PR) $PR = 0$ (1) Pour chaque attribut $a_i \in \{a_1, a_3, a_5, a_7, a_9, a_{11}, a_{13}, a_{15}, a_{17}, a_{19}, a_{21}, a_{23}, a_{25}, a_{30}\}$: Si $v_i^A \neq v_i^C$ et $v_i^C = Vrai$ Alors, $PR = PR + w_i$ Si $v_i^A \neq v_i^C$ et $v_i^C = Faux$ Alors, $PR = PR - w_i$ (2) $PR = PR + \frac{(v_{42}^C - v_{42}^A)}{3} \times w_{42}$</p>
<p>Sortie PR : paramètre de révision de la prédiction</p>

Figure 5.6 - Algorithme de calcul du Paramètre de Révision proposé

Le Paramètre de Révision est une valeur comprise entre $[-1; 1]$. Si la valeur PR est négative, la valeur estimée d'impact du cas C est plus petite que celle du cas A , et inversement.

À partir de ce Paramètre de Révision, nous proposons deux révisions possibles sur la valeur solution $Sol(C)$:

- a) Le Paramètre de Révision est ajouté à la valeur solution du cas source :

$$Sol(C) = Sol(A) + PR$$

- b) Le Paramètre de Révision est utilisé comme un pourcentage de changement à effectuer sur la solution du cas source :

$$Sol(C) = Sol(A) + (Sol(A) \times PR)$$

La valeur estimée pour la proposition a) est dans l'intervalle $[-1; 8]$, et $Sol(C)$ aura au maximum un point de moins ou de plus que $Sol(A)$.

La valeur estimée pour la proposition b) est dans l'intervalle $[0; 14]$, et $Sol(C)$ aura au maximum sept points de moins ou de plus que $Sol(A)$.

Exemple :

Le cas le plus similaire à C est le cas A , nous calculons le Paramètre de Révision :

A l'initialisation $PR = 0$

Pour a_1 : la valeur de A est "Vrai", celle de C est "Faux",

alors : $PR = PR - w_1 = 0 - 0,2 = -0,2$

Pour a_3 : la valeur de A est "Faux", celle de C est "Vrai",

alors : $PR = PR + w_2 = -0,2 + 0,5 = 0,3$

Pour a_5 : la valeur de A est "Vrai", celle de C est "Vrai",

alors : PR ne change pas.

Le paramètre de révision est donc de 0,3.

- Révision a) : $Sol(C) = Sol(A) + PR = 4 + 0,3 = 4,3$,
Finalement, avec la révision a), l'impact prédit de C est 4,3.
- Révision b) : $Sol(C) = Sol(A) + (Sol(A) \times PR) = 4 + (4 \times 0,3) = 5,2$,
Finalement, avec la révision b), l'impact prédit de C est 5,2.

L'étape de réutilisation/révision étant effectuée, l'étape de mémorisation doit être réalisée, nous la présentons dans la Section 3.2.4 car elle fonctionne de la même façon pour les deux modèles de prédiction que nous proposons.

Nous venons dans cette section de présenter un modèle de prédiction basé sur le RàPC. L'utilisation de ce modèle est illustrée dans le Chapitre « Expérimentations ».

Dans ce modèle de prédiction, les similarités locales et globales sont utilisées ; dans le cas du modèle proposé dans la section suivante, seule la similarité locale est utilisée.

3.2.3. Modèle de Raisonnement à Partir des Attributs (RàPA) inspiré du cycle du RàPC, basée sur la similarité entre chaque descripteur

Le second modèle proposé est décrit dans cette section. Ses étapes s'inspirent de celles du RàPC, mais en raisonnant à l'échelle des attributs et non à celle des cas. De plus, les étapes de réutilisation et de révision ne sont pas distinctes.

Les similarités locales entre les cas sources et le cas cible sont calculées comme présenté dans la Section 3.2.1. Ainsi, nous avons m vecteurs de similarités locales, correspondant aux similarités locales entre le cas cible et les m cas sources de la base de cas.

(1) Pour chacune des dix-sept parties (comme définies dans la Section 3.1), le cas source le plus similaire selon cette partie est récupéré. Si plusieurs cas sources ont la même similarité maximale, la moyenne de leur solution est utilisée. La similarité au sein d'une partie P_i entre les cas A et B est définie de la façon suivante :

$$sim_{partie}(P_i^A, P_i^B) = \frac{1}{|P_i|} \sum_{a \in P_i} sim_{loc}(a^A, a^B)$$

(2) Un pourcentage de la solution du cas source est récupéré (ou une moyenne des pourcentages des cas sources), ce pourcentage dépend de la pondération de la partie considérée selon le vecteur w .

Enfin, les valeurs obtenues pour chaque partie sont sommées afin d'obtenir une prédiction de la solution du cas cible.

Ces deux étapes du modèle RàPA proposé sont présentées dans la **Figure 5.7**. L'opérateur *arg max* utilisé nous permet de sélectionner le cas donnant la similarité la plus élevée.

<p>Entrée $\{(srce ; Sol(srce))\}_m$: ensemble des m cas sources, tgt_C : couples (attribut, valeur) du cas cible C, w : vecteur de poids des parties.</p>
<p>Étape de réutilisation du RàPA $Sol(C) = 0$ (1) Pour chaque partie $P_i \in \{P_1, \dots, P_{17}\}$: Cherchons le cas source A tel que $A = \underset{Cas \in \{(srce ; Sol(srce))\}_m}{arg \max} (sim_{partie}(P_i^{Cas}, P_i^C))$ (2) $Sol(C) = Sol(C) + Sol(A) \times w_i$</p>
<p>Sortie $Sol(C)$: solution prédite pour le cas cible C</p>

Figure 5.7 - Algorithme de prédiction à partir des attributs

La valeur prédite pour le modèle RàPA est dans l'intervalle $[1 ; 7]$, et la valeur de $Sol(C)$ prédite peut être très différente des valeurs des cas de la base de cas. Ce modèle suppose toutefois que plus un dispositif a d'éléments, plus il est susceptible d'avoir une valeur prédite de solution élevée.

Exemple :

L'exemple est réalisé sur les trois attributs du problème fictif, mais le principe est le même que sur les parties. La similarité entre les attributs des cas sources et du cas cible est calculée :

Attribut a_i	$sim_{loc}(a_i^A, a_i^C)$	$sim_{loc}(a_i^B, a_i^C)$
a_1	0	0
a_3	0	1
a_5	1	0

À l'initialisation $Sol(C) = 0$

Pour a_1 : $sim_{loc}(a_1^A, a_1^C) = sim_{loc}(a_1^B, a_1^C) = 0$ La valeur de similarité est la même pour les cas A et B, nous utilisons la moyenne de la solution de ces deux cas,

$$\text{alors : } Sol(C) = Sol(C) + \frac{Sol(A)+Sol(B)}{2} \times w_1 = 0 + \frac{4+5}{2} \times 0,2 = 0,9$$

Pour a_3 : La valeur de similarité pour cet attribut est la plus grande avec le cas B, avec $sim_{loc}(a_3^B, a_3^C) = 1$,

$$\text{Alors : } Sol(C) = Sol(C) + Sol(B) \times w_3 = 0,9 + 5 \times 0,5 = 3,4$$

Pour a_5 : La valeur de similarité pour cet attribut est la plus grande avec le cas A, avec $sim_{loc}(a_5^A, a_5^C) = 1$,

$$\text{Alors : } Sol(C) = Sol(C) + Sol(A) \times w_5 = 3,4 + 4 \times 0,3 = 4,6$$

Finalement, l'impact prédit de C est 4,6.

L'étape de réutilisation/révision étant effectuée, l'étape de mémorisation doit être réalisée, son déroulement est présenté dans la section suivante.

3.2.4. Mémorisation des prédictions du cas cible

Lors de cette étape, le cas C est de la forme $(srce ; Sol(srce))$. Il est enregistré dans la base de cas et devient un cas source. Toutefois, il est nécessaire de vérifier manuellement que la solution de C a du sens. Pour cela, l'intervention d'un expert est indiquée.

Si le dispositif représenté par le cas C est utilisé auprès de lycéens, il est intéressant de mesurer l'impact comme présenté dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation » et de changer la solution prédite dans la base de cas si elle est différente de la solution mesurée.

4. Conclusion

Nous nous sommes intéressés dans ce chapitre aux travaux concernant la prédiction d'une valeur numérique à partir d'approches d'IA. En effet, ces approches sont très utilisées dans les problèmes de prédiction et donnent de bons résultats. Les approches utilisées dépendent du problème et des données disponibles.

De ce fait, pour notre problème, les systèmes de RàPC se placent comme étant les plus appropriés. Ils fonctionnent de façon incrémentale et s'inspirent du raisonnement humain en utilisant les problèmes passés pour résoudre les problèmes futurs. Les problèmes sont appelés des cas, et leur structure doit être définie.

Par la suite, dans ce chapitre, nous avons proposé une structure des cas basé sur l'ontologie ODNAO, les modules « **Nature de l'information** », « **Services additionnels** », « **Interaction de l'utilisateur avec le dispositif** », « **Utilisabilité** », « **Utilisateur** » décrivent le problème, tandis que le module « **Impact** » décrit la solution.

Les cas correspondant à notre problème possèdent quarante-cinq attributs, ces attributs sont répartis en dix-sept parties pour permettre une pondération des éléments par des experts. Ces parties sont réparties dans les cinq modules afin de faciliter leur utilisation.

Le cycle du RàPC se compose de plusieurs étapes que nous avons décrites en détail dans ce Chapitre. Les étapes de récupération, réutilisation et prédiction du RàPC sont dépendantes de la structure des cas et des possibilités de révision selon le contexte. Les travaux existants ne se plaçant pas dans une recherche de valeur d'impact des dispositifs numériques à destination des lycéens, la seconde contribution porte sur la conception de deux modèles de prédiction. Le premier modèle est un système de RàPC classique avec une étape de révision adaptée à nos données. Le second modèle prédit la solution du problème à partir des attributs les plus similaires et non à l'échelle des cas, il s'inspire du cycle du RàPC.

Il est nécessaire de tester et de comparer ces deux propositions sur des données, cette comparaison est réalisée dans la Section 4.3 dans le Chapitre « Expérimentations ».

EXPERIMENTATIONS

TABLE DES MATIERES

1. Introduction.....	105
2. Instanciation de l'ontologie pour les dispositifs Hello Charly et Onisep	106
2.1. Hello Charly	106
2.1.1. Présentation du dispositif Hello Charly	106
2.1.2. Instanciation de l'ontologie proposée pour le dispositif Hello Charly	108
2.2. Onisep	117
2.2.1. Présentation du dispositif Onisep.....	117
2.2.2. Instanciation du dispositif Onisep	119
2.3. Synthèse des instanciations.....	122
3. Estimation de l'impact.....	122
3.1. Expérimentations pour créer notre modèle	122
3.1.1. Protocole de mesure	122
3.1.2. Résultats obtenus.....	123
3.1.2.1. Régression linéaire multiple.....	123
3.1.2.2. Modélisation par équation structurelle	124
3.1.2.3. Réseaux bayésiens	124
3.1.2.4. Discussion sur les modèles de relation entre l'impact et les indicateurs d'évaluation des dispositifs numériques	125

3.2. Expérimentation de l'estimation de l'impact.....	126
3.2.1. Mesure de l'impact du dispositif Hello Charly	126
3.2.2. Mesure de l'impact du dispositif Onisep	128
4. Expérimentation des deux modèles de prédiction proposés.....	130
4.1. Importance des caractéristiques des dispositifs numériques	130
4.2. Proximité entre les instances des concepts	135
4.3. Prédiction de l'impact.....	137
4.3.1. Prédiction de l'impact selon le modèle de prédiction en suivant le cycle du RàPC	138
4.3.1. Prédiction de l'impact selon le modèle RàPA	140
4.4. Étude de la sensibilité de l'impact selon le poids des critères	143
5. Conclusion	146

1. Introduction

Cette section présente les expérimentations qui ont été réalisées lors de cette thèse. Elles ont été menées au sein du lycée polyvalent Denis Diderot à Langres et du laboratoire ERPI à Nancy, en collaboration avec des psychologues de l'éducation nationale et auprès de lycéens en seconde générale et technologique.

Nous nous plaçons dans le cadre du projet AILES et plus particulièrement dans l'évaluation des dispositifs numériques destinés à aider les lycéens dans leur processus d'orientation. Dans le cadre de ce projet, le dispositif numérique Hello Charly a été mis en place dans le lycée Diderot pour deux années scolaires, de 2021 à 2023. Ainsi, les professionnels de l'éducation du projet AILES s'interrogent particulièrement sur l'impact de ce dispositif. Pour cette raison, les expérimentations réalisées lors de cette thèse se sont concentrées sur ce dispositif.

Le portail numérique de l'Onisep est également considéré, car il s'agit d'un dispositif de référence porté par l'Éducation Nationale, utilisé très largement par les lycéens dans leur processus d'orientation et également par les professionnels qui les entourent.

La première section de ce chapitre présente la modélisation et le stockage des caractéristiques du dispositif Hello Charly et du portail de l'Onisep, à partir de l'ontologie proposé dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ».

Dans la deuxième section, nous présentons tout d'abord les expérimentations réalisées pour obtenir les valeurs sur les arcs du modèle d'impact proposé dans la Section 3.3 du Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ».

Puis, à partir de ce modèle, nous mesurons l'impact auprès de lycéens de seconde du dispositif Hello Charly et du portail de l'Onisep.

Finalement, dans la troisième section, dans un premier temps, nous présentons les expérimentations réalisées pour obtenir les poids des caractéristiques des dispositifs numériques pour le calcul de similarité. Puis, nous expérimentons l'utilisation d'une approche basée sur le « Raisonnement à Partir de Cas » pour prédire l'impact de dispositifs numériques à partir des deux propositions décrites dans le Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation ».

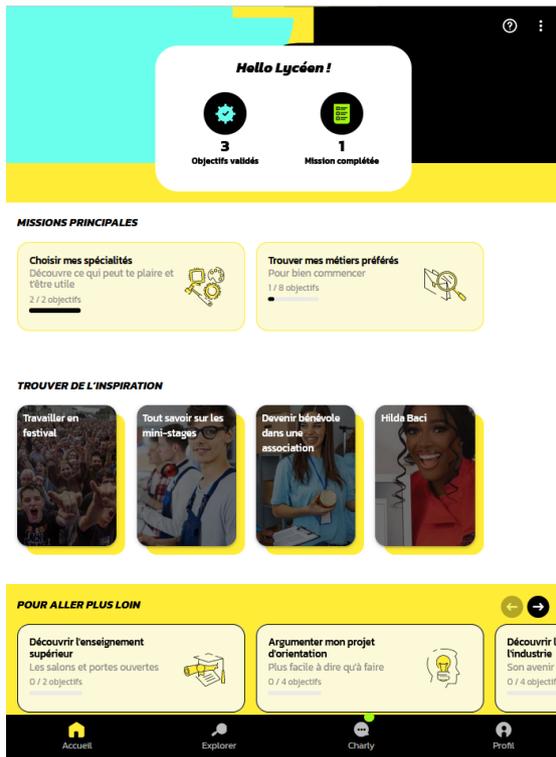
2. Instanciation de l'ontologie pour les dispositifs Hello Charly et Onisep

Cette section présente l'instanciation de l'ontologie proposée dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation » pour deux dispositifs numériques : Hello Charly et le portail de l'Onisep. Pour instancier l'ontologie pour un dispositif particulier, un protocole doit être suivi. Ce protocole, présenté en Annexe C, est à destination des professionnels de l'orientation pour les guider lors de l'instanciation et pour s'assurer que les modules de l'ontologie sont compris et que tous sont considérés.

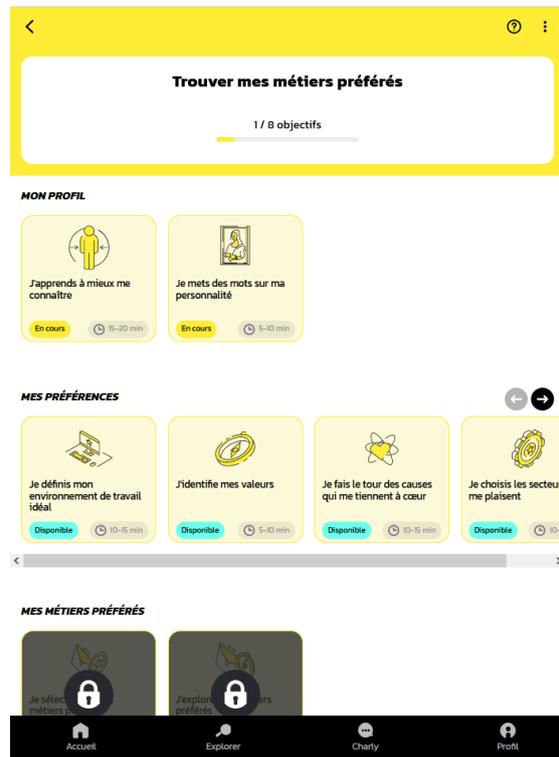
2.1. Hello Charly

2.1.1. Présentation du dispositif Hello Charly

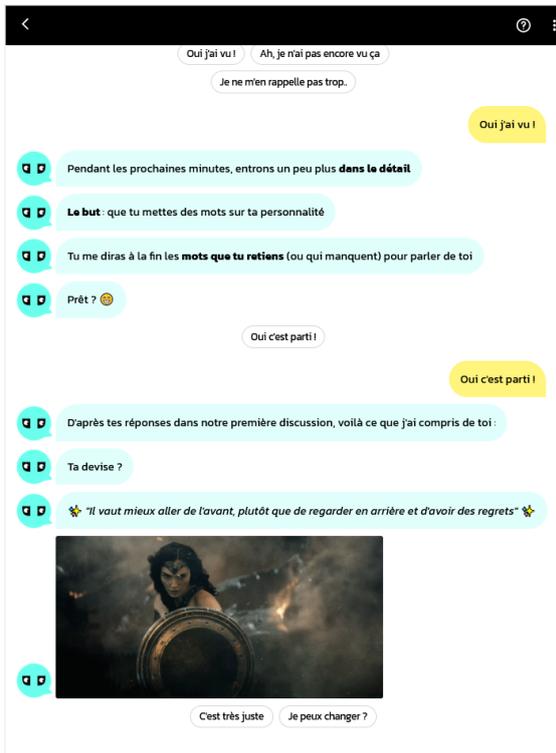
Créé en 2015, Hello Charly est un dispositif qui s'utilise sur smartphone, tablette et ordinateur. Il est conçu selon les codes sociaux des jeunes et basé sur le principe de messagerie instantanée. À travers une discussion, le robot, nommé « Charly », propose un parcours guidé et personnalisé pour atteindre des objectifs concernant l'orientation des élèves. Concrètement, « Charly » pose des questions à l'utilisateur, et en fonction des réponses, il peut émettre des recommandations personnalisées. Les lycéens répondent à « Charly » en choisissant parmi un ensemble de réponses possibles. Plusieurs sujets peuvent être abordés par « Charly ». Ce sont les lycéens qui choisissent lors de leur utilisation les sujets qui les intéressent. Par exemple, ils peuvent échanger sur le sujet du choix de spécialité au lycée ou sur celui des métiers qu'ils préfèrent. Des captures d'écran de l'interface du site web de Hello Charly sont présentées dans la **Figure 6.1**. Les lycéens de Langres accèdent à Hello Charly à partir de leur environnement numérique de travail.



Page d'accueil



Discussions possibles avec « Charly » sur les métiers préférés



Interface d'une discussion avec « Charly »



Page dédiée au profil de l'utilisateur

Figure 6.1 - Captures d'écran du site web Hello Charly

2.1.2. Instanciation de l'ontologie proposée pour le dispositif Hello Charly

L'instanciation se fait pour commencer par la création de l'instance « helloCharly », qui est l'instance du concept « DispositifNumerique ». Ce concept est utilisé pour relier les autres concepts.

2.1.2.1. Instanciation des modules « Nature de l'information », « Interaction de l'utilisateur avec le dispositif numérique » et « Services additionnels »

Le module « **Nature de l'information** » est complété en premier.

À cet effet, nous nous intéressons au concept « InformationContenu ». Il contient neuf sous-concepts. Si le sujet d'une sous-classe existe dans le dispositif numérique, une instance correspondante est créée. Les instances de ce module permettent de décrire le dispositif à partir des sujets abordés par celui-ci. Pour ce concept, six instances sont créées, une pour chacune de ces sous-classes : « InformationAProposDuDispositif », « InformationLycee », « InformationMetier », « InformationOrientation », « InformationDiplomePostBac », et « InformationPsychologiePersonnalite »

Chacune de ces six instances est ensuite reliée à une ou plusieurs instances du concept « ComposantElementaire » grâce à la relation « aComposantElementaire ».

Par exemple, l'instance « informationPsychologiePersonnaliteHelloCharly » est liée aux instances « image » et « texte ». Cette relation permet de décrire la façon dont les informations liées à la psychologie et à la personnalité sont présentes dans le dispositif Hello Charly.

La capture d'écran de l'instanciation de ce module pour Hello Charly est présentée dans la **Figure 6.2**.

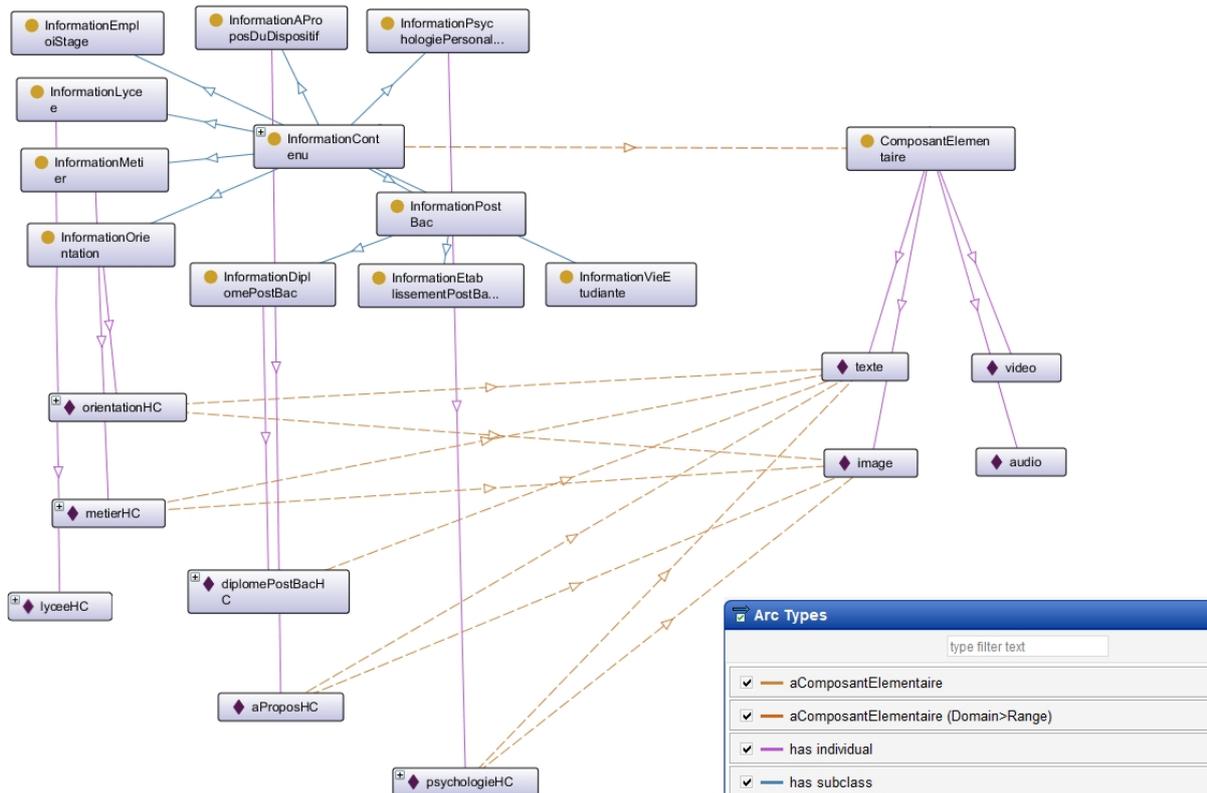


Figure 6.2 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « **Nature de l'information** »

Dans le module « **Interaction de l'utilisateur avec le dispositif numérique** », nous créons deux instances :

- L'instance « **supportDispositifHelloCharly** » pour la classe « **SupportDispositif** ». Cette instance a pour propriété « **typeSupport** », et « **supportPhysique** ». La valeur de la première propriété est « **SiteWeb** », la valeur de la seconde est « **Ordinateur** ». Nous considérons dans cette instanciation uniquement le site web de Hello Charly, l'application mobile n'est pas prise en compte. Ce choix est fait, car Hello Charly est utilisé par les élèves du lycée Diderot principalement en classe sur leur ordinateur. Il s'agit de leur ordinateur personnel, distribué par la région Grand Est lors de leur entrée en seconde.
- L'instance « **interfaceUtilisateurHelloCharly** » pour la classe « **InterfaceUtilisateur** ». Cette instance a deux propriétés qui prennent les valeurs suivantes : « **typeInterface = Interface Par Page** », « **disponibilite = Permanent** ». L'interface de Hello Charly est construite avec plusieurs pages et les lycéens ont accès toute l'année au dispositif.

La capture d'écran de l'instanciation de ce module pour Hello Charly est présentée dans la **Figure 6.3**.

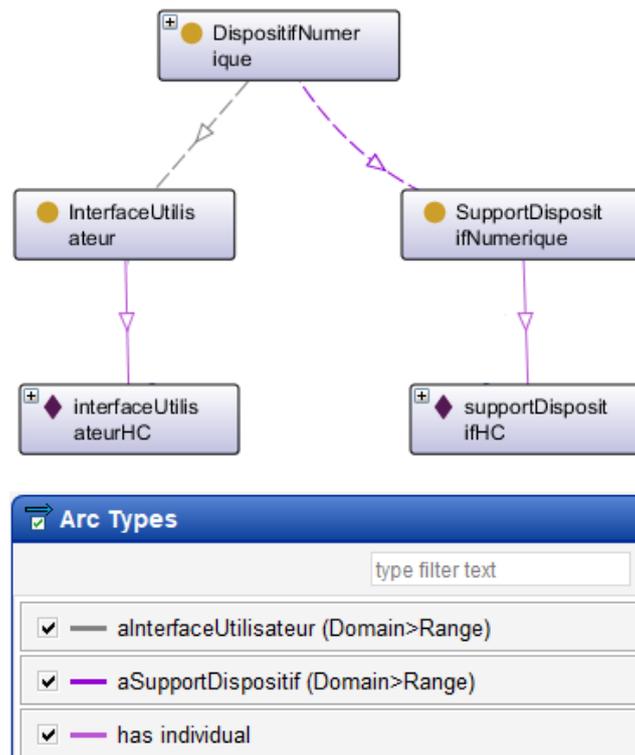


Figure 6.3 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « **Interaction de l'utilisateur avec le dispositif numérique** »

Enfin, les services additionnels fournis par Hello Charly sont indiqués avec le module « **Services additionnels** ».

Hello Charly comprend un outil de communication avec un chatbot. Ainsi, le concept « OutildEchange » est instancié avec une instance nommée « outildEchangeHelloCharly ». Cette instance est liée à « robot » par la relation « aInterlocuteur ». « robot » est une instance possible du concept « Interlocuteur ».

Les propriétés du concept « OutildEchange » permettent d'indiquer que :

- (1) L'outil permet une communication instantanée (« permetImmediate = Vrai »),
- (2) Il n'est pas possible de communiquer par vidéo (« aVideo = Faux »),
- (3) Pour pouvoir discuter avec le robot Charly, il faut s'identifier (« requiereAuthentication = Vrai »).

Ce module contient également trois propriétés dans le concept « DispositifNumerique », permettant de préciser si

- (1) Le dispositif contient ou non un ensemble de Questions Fréquemment Posées (FAQ) (« contientFaq = Faux »),
- (2) Le dispositif contient ou non un annuaire (« contientAnnuaire = Faux »),
- (3) Le dispositif contient ou non des témoignages (« contientTemoignage = Faux »).

La capture d'écran de l'instanciation de ce module pour Hello Charly est présentée dans la **Figure 6.4**.

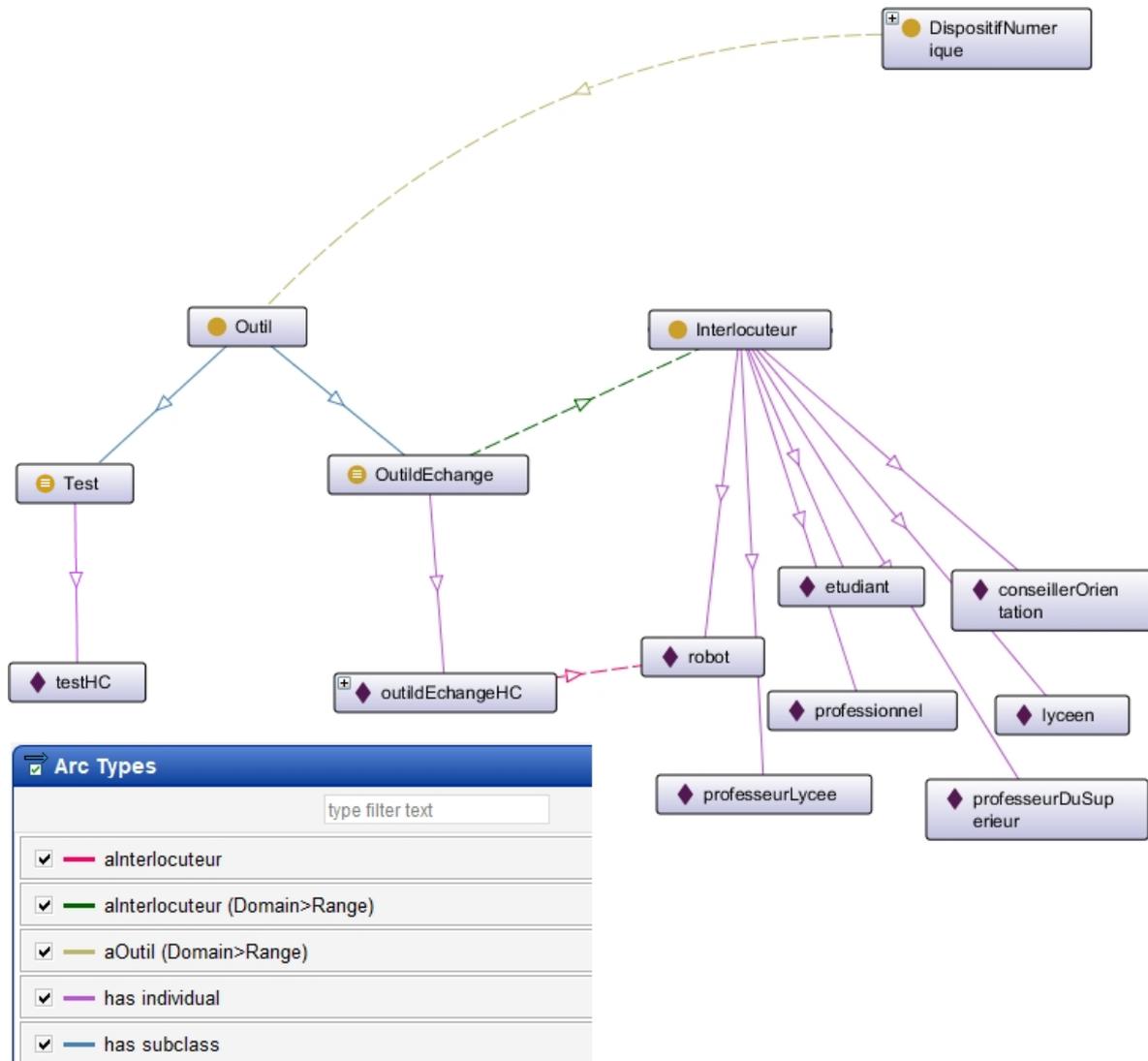


Figure 6.4 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « **Services additionnels** »

2.1.2.2. Instanciation du module « Utilisabilité »

Pour instancier ce module, nous nous appuyons sur le modèle d'évaluation de l'utilisabilité présenté dans la Section 3.2.5 du Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation ». Ce modèle a l'avantage de nécessiter uniquement un petit nombre d'experts préalablement formés.

Dans notre problème, l'utilisabilité de certains dispositifs doit être mesurée avant que les lycéens n'y aient accès, donc le nombre d'utilisateurs possibles est faible. Il est alors avantageux que le modèle puisse être évalué par un petit nombre d'experts. Les résultats sont discutés entre les experts pour s'assurer de leur fiabilité.

Ces axes sont évalués à travers vingt affirmations. Chaque expert doit attribuer une note comprise entre zéro et trois à chacune des questions présentées dans le **Tableau 6.1** : zéro correspond à « mauvais » et trois à « excellent ». Les notes individuelles sont additionnées et moyennées pour obtenir une appréciation globale pour chaque axe.

L'expert remplit la colonne score après avoir suivi un parcours utilisateur spécifié. Ce parcours utilisateur considère l'objectif du dispositif et le profil du lycéen pour qui nous instancions l'ontologie. Tous les experts ont le même parcours à suivre.

Navigation
J'ai accès à une barre de menu/navigation permanente et saillante.
Je peux faire une identification de base des contenus (titre, source, date).
Je peux effectuer une navigation non séquentielle efficace.
Des éléments de navigation à distance/systèmes de navigation complémentaires sont présents (ex : table des matières, index, plan du site).
Il est possible de rechercher/récupérer des informations par le biais de questions ou mots clef.
L'accès aux pages est facile (par exemple, pas de retour en arrière excessif/de clics et d'accès par de multiples chemins).
Je peux savoir facilement où je me trouve sur le dispositif.
Le dispositif n'ouvre pas trop de nouvelles fenêtres/de pop-up lorsque je me déplace.
Contenu
La quantité de contenu sur chaque page/blocs lisibles est appropriée.
Les sujets traités sont à jour.
Représentation graphique
Le contraste est adéquat entre les figures et le fond (correspond à l'image de marque et aux tendances actuelles sans compromettre la convivialité),
La typographie (couleur, police et taille du texte) correspond à l'image de marque et aux tendances actuelles, sans nuire à la facilité de lecture.
La taille et la résolution des images sont adaptées.
Logos et icônes sont distincts.
Il n'y a pas de fonctions irritantes (texte défilant/clignotant/animations en boucle...).
Ergonomie, efficacité
La vitesse de chargement du contenu est appropriée (je ne dois pas attendre trop longtemps après avoir réalisé une action).
La vitesse de chargement des éléments visuels est appropriée.
La conception du dispositif est optimisée pour le support utilisé (par exemple : pas besoin de faire défiler les pages de gauche à droite lorsque nous lisons).
Image de marque

Mise en œuvre des éléments fonctionnels et émotionnels de la marque, le logo de la marque, les schémas de couleurs, la typographie, le style, le ton, ainsi que le rôle des images et des graphiques.

Présence de liens vers le blog de la marque et les principaux réseaux sociaux.
--

Tableau 6.1 - Questionnaire pour mesurer l'utilisabilité d'un dispositif

L'utilisabilité est mesurée à partir des réponses de quatre personnes du laboratoire ERPI travaillant sur le projet AILES, et préalablement formées par les auteurs à partir du scénario suivant :

« Choisissez un métier/domaine d'activité et recherchez des informations sur celui-ci (retour d'expérience, conditions de travail, salaire), les études dans ce domaine, et les options les plus adaptées à choisir pour la première année. »

Cette expérimentation a été réalisée avec ces quatre experts, car l'objectif était d'illustrer l'instanciation et d'expérimenter le protocole de mesure d'utilisabilité.

Les personnes ont répondu individuellement au questionnaire décrit dans le **Tableau 6.1**, puis une discussion a eu lieu pour les différentes réponses et comprendre des différences éventuelles. Enfin, les résultats obtenus permettent de compléter les propriétés de l'instance « utilisabiliteHelloCharly ». Les valeurs de chacune de ces propriétés sont présentées dans le **Tableau 6.2**. La capture d'écran de l'instanciation de ce module, pour Hello Charly, est présentée dans la **Figure 6.5**.

Propriété d'objet	Valeur
valeurDeNavigabilite	1,9
valeurDeContenu	2,1
valeurRepresentationGraphique	2,4
valeurDErgonomie	2,4
valeurImageDeMarque	1
valeurUtilisabilite	1,9

Tableau 6.2 - Valeurs des propriétés d'objet du module « **Utilisabilité** » pour Hello Charly

La valeur de la propriété « valeurUtilisabilite » est obtenue en calculant la moyenne des cinq autres propriétés.

Intervalle de la valeur	Niveau d'utilisabilité
[0 ; 0,75[Mauvais
[0,75 ; 1,5[Pauvre
1,5	Modéré
]1,5 ; 2,25]	Bon
]2,25 ; 3]	Excellent

Tableau 6.3 - Valeur et niveau d'utilisabilité correspondants

Les valeurs entre zéro et trois de chacune des propriétés sont interprétées selon le **Tableau 6.3**.

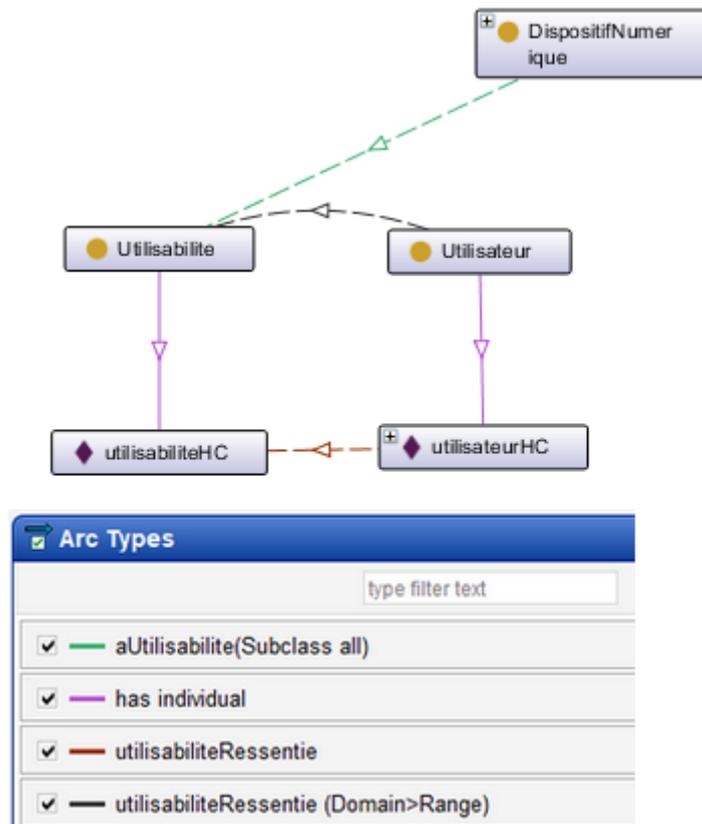


Figure 6.5 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « **Utilisabilité** »

Afin de valider cette mesure d'utilisabilité, le System Usability Scale (SUS) de Hello Charly a été mesuré auprès de trente-quatre élèves de seconde. Le SUS est un questionnaire composé de dix questions. Il permet d'évaluer l'utilisabilité et la convivialité d'un service/dispositif. Les dix questions posées utilisent l'échelle de Likert composée de 5 items allant de « Pas du tout d'accord » à « Tout à fait d'accord » [151].

Le score obtenu se situe entre un et cent. Plus le score est proche de cent, plus l'utilisateur trouve que le service a une très bonne utilisabilité.

Le test SUS a été retenu, car il présente deux atouts majeurs :

- Il s'agit d'un test standardisé couramment utilisé pour la mesure de satisfaction d'un produit ou d'un service numérique.
- Il est très court (dix questions seulement), ce qui favorise l'obtention de réponses fiables.

Le score obtenu est de 71/100, ce qui signifie que les utilisateurs trouvent l'utilisabilité bonne (**Figure 6.6**) et est en accord avec les résultats obtenus avec notre mesure d'utilisabilité puisqu'elle est également bonne.

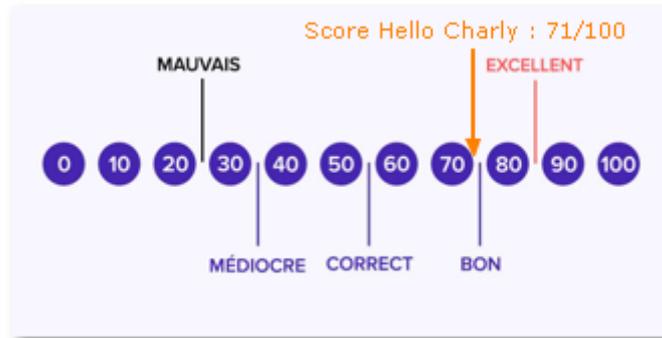


Figure 6.6 - Échelle des scores du SUS

2.1.2.3. Instanciation du module « Utilisateur »

Une instance « utilisateurHelloCharly » du concept « Utilisateur » est créée avec les propriétés présentées dans le **Tableau 6.4**.

Propriété de l'objet	Valeur
annee	Seconde
etapeDuProcessus	Cristallisation
typeDeBaccalaureat	{Général, Technologique}

Tableau 6.4 - Valeurs des propriétés d'objet du module « Utilisateur »

La capture d'écran de l'instanciation de ce module, pour Hello Charly, est présentée dans la **Figure 6.7**.

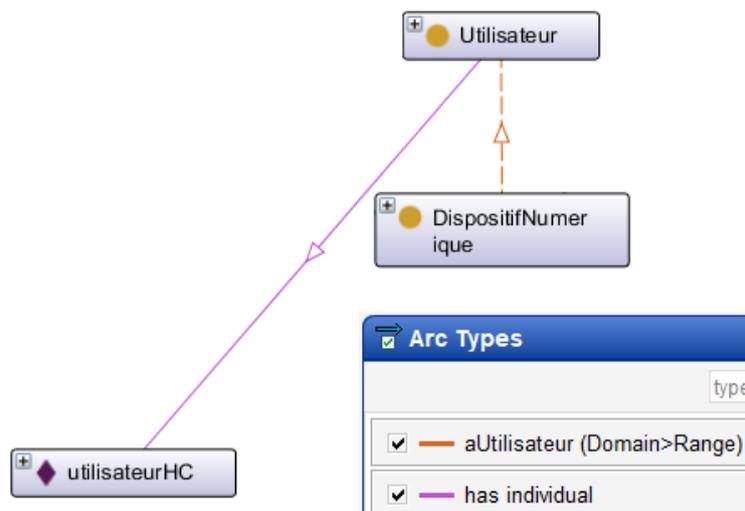


Figure 6.7 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « Utilisateur »

2.1.2.4. Instanciation du module « Impact »

Pour obtenir l'impact de Hello Charly sur le processus d'orientation des lycéens selon le modèle d'estimation de l'impact présenté dans le Chapitre « Modèle de mesure de l'impact

des dispositifs numériques d'aide à l'orientation », les valeurs hédoniques, sociales et l'utilité perçue par les lycéens lors de leur utilisation de Hello Charly sont mesurées.

Les résultats du **Tableau 6.5** sont obtenus de la façon suivante : pour une propriété d'objet, nous avons posé six questions (les questions sont présentées dans la Section 3.2 du Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation »). La réponse à chaque question est une valeur comprise entre un et sept. Pour un individu, nous avons pris la moyenne de ces six valeurs. Enfin, la moyenne des réponses de tous les individus est calculée pour obtenir une valeur unique pour chaque catégorie.

Quatre-vingt-huit lycéens ont répondu au questionnaire. Les moyennes obtenues sont présentées dans le **Tableau 6.5**, ces lycéens ont le profil décrit dans le module « **Utilisateur** ».

La capture d'écran de l'instanciation de ce module, pour Hello Charly, est présentée dans la **Figure 6.8**.

Propriété de l'objet	Valeur
valeurHedonique	4,99
valeurUtilitaire	4,56
valeurSociale	3,58
valeurGlobale	$p_1 \times 4,99 + p_2 \times 4,56 + p_3 \times 3,58$ où p_1, p_2 et p_3 sont les poids reliant la valeur globale d'impact aux valeurs des trois catégories

Tableau 6.5 - Valeurs des propriétés d'objet du module « **Impact** »

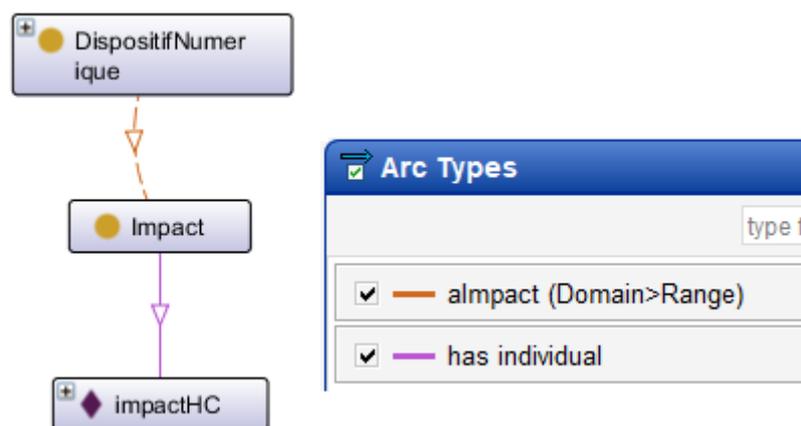


Figure 6.8 - Capture d'écran de l'interface Protégé pour l'instanciation de Hello Charly pour le module « **Impact** »

2.1.2.5. Synthèse de l'instanciation de Hello Charly

Un grand nombre de concepts de l'ontologies est instancié lors de l'instanciation de Hello Charly. Les deux outils actuellement possibles sont instanciés. Les concepts qui ne sont pas instanciés sont des sous-concepts de « Information Contenu » car la version considérée de Hello Charly ne donne pas d'informations sur des offres d'emploi ou de stage, ni sur des établissements post-baccalauréat, ni sur la vie étudiante.

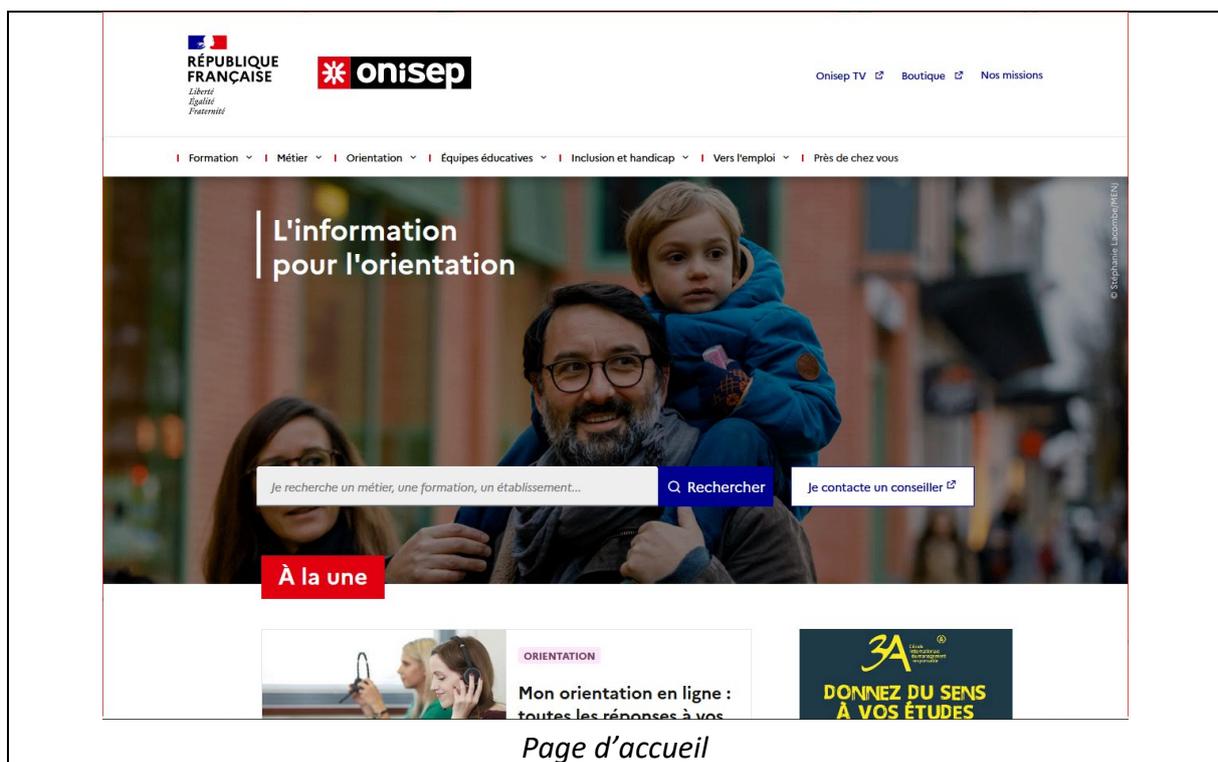
2.2. Onisep

2.2.1. Présentation du dispositif Onisep

Le portail numérique de l'Onisep relève du ministère de l'Éducation nationale et de la Jeunesse et du ministère de l'Enseignement supérieur et de la Recherche. L'Onisep, produit et diffuse toute l'information sur les formations et les métiers. Il offre également des services aux élèves, aux parents et aux équipes éducatives. Ce dispositif a pour objectif d'informer les élèves et leur famille sur les formations, les métiers et les secteurs professionnels.

Ce dispositif propose des actualités nationales et régionales sur les formations et les métiers, la localisation des lieux de formation et d'information, les structures destinées à l'information ou à la scolarisation des élèves en situation de handicap. Il présente des informations précises "Formation", "Métier", "Orientation", "Équipes éducatives", "Inclusion et handicap"...

Des captures d'écran de l'interface du portail de l'Onisep sont présentées dans la **Figure 6.9**.



2. INSTANCIATION DE L'ONTOLOGIE POUR LES DISPOSITIFS HELLO CHARLY ET ONISEP

The screenshot shows the Onisep website homepage. At the top left is the logo of the République Française (Liberté, Égalité, Fraternité) and the Onisep logo. To the right are links for 'Onisep TV', 'Boutique', and 'Nos missions', and a search bar with the text 'Rechercher'. Below the header is a navigation menu with items: 'Formation', 'Métier', 'Orientation', 'Équipes éducatives', 'Inclusion et handicap', 'Vers l'emploi', and 'Près de chez vous'. The main content area features a large image of two women talking, with the text 'L'information pour l'orientation' and 'Métier'. Below this is a search bar with the text 'Rechercher un métier' and a search button. Underneath, there are four buttons: 'Découvrir le monde professionnel', '10 métiers dans un univers qui vous attire', 'Des métiers qui recrutent', and 'Découvrir les univers des professionnels en images'. At the bottom, there is a form to search for a job, with the text 'Pour trouver un métier, précisez : Quel métier ? Un nom de métier' and a search button, followed by a link 'Voir tous les métiers'.

Page d'accueil « Métier »

The screenshot shows the Onisep website job profile page for 'Animateur nature / animatrice nature'. At the top left is the logo of the République Française and the Onisep logo. To the right are links for 'Onisep TV', 'Boutique', and 'Nos missions', and a search bar with the text 'Educateur/trice environnement'. Below the header is a navigation menu with items: 'Formation', 'Métier', 'Orientation', 'Équipes éducatives', 'Inclusion et handicap', 'Vers l'emploi', and 'Près de chez vous'. The main content area features a large red background with the text 'Animateur nature / animatrice nature'. Below this is a description of the job: 'Sorties pour découvrir la flore et la faune locales, création de sentiers, organisation d'une exposition... l'animateur ou l'animatrice nature ne manque pas d'idées pour sensibiliser petits et grands à la préservation de l'environnement. De quoi former des écocitoyens.' To the left of the main content is a sidebar with a 'SOMMAIRE' section containing links: 'Le métier', 'Compétences requises', 'Où l'exercer ?', 'Les études', 'Emploi et secteur', 'Salaire du débutant', 'Autres ressources', and 'Pour aller plus loin'. To the right of the main content is a 'Partager la page' section with social media icons. Below the main content, there is a box with the following information: 'Niveau minimum d'accès : Bac + 2', 'Salaire débutant : 1747 €', 'Statuts : Statut fonctionnaire, Saisonnier, Statut salarié', 'Synonymes : Animateur/trice environnement, Eco-interprète, Educateur/trice environnement, Educateur/trice nature, Guide animateur/trice', 'Secteurs professionnels : Environnement, Fonction publique, Tourisme', and 'Centres d'intérêt : J'aime la nature, J'aime organiser, gérer, J'aimerais travailler dehors, Je veux'.

Exemple de fiche métier

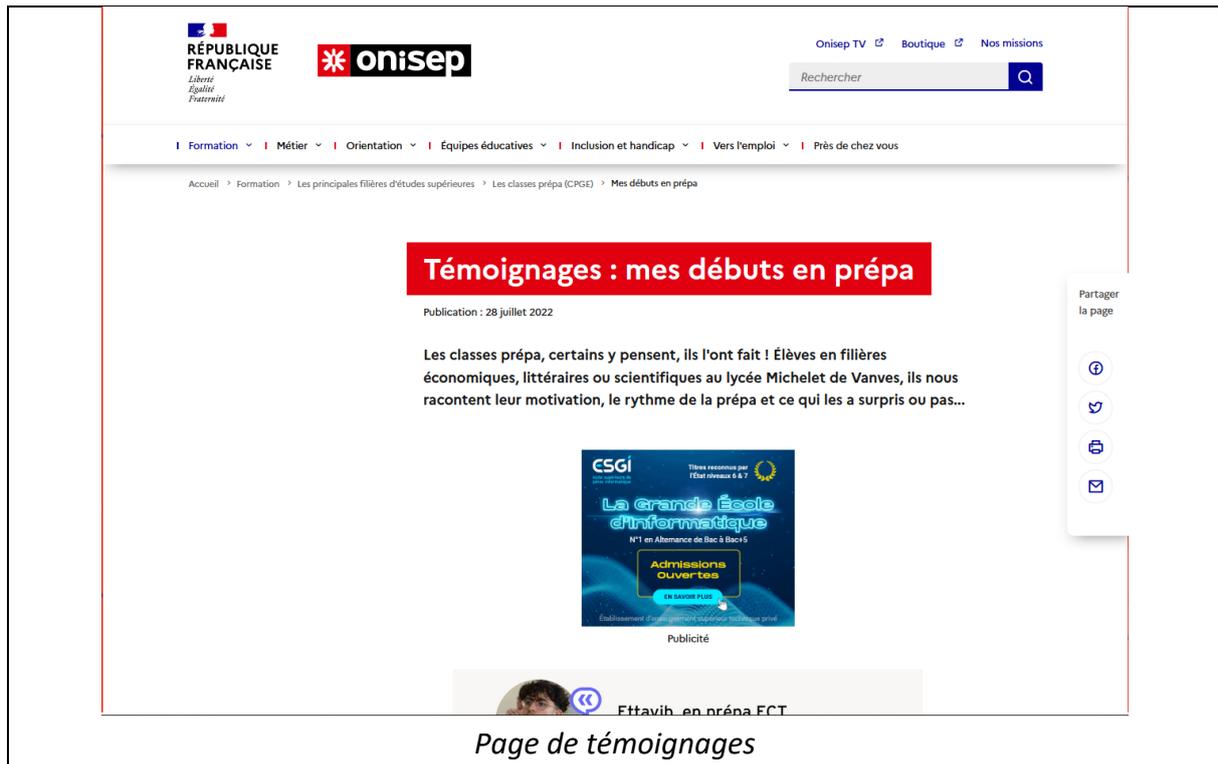


Figure 6.9 - Captures d'écran du portail de l'Onisep

2.2.2. Instanciation du dispositif Onisep

L'instanciation se fait pour commencer par la création de l'instance "onisep", qui est l'instance du concept « DispositifNumerique ». Ce concept est utilisé pour relier les autres concepts.

Les instances pour le module « **Nature de l'information** » et leurs relations sont décrites dans le **Tableau 6.6**.

Instance	Classe de l'instance	Instances mises en relation avec la propriété « a ComposantElementaire »
informationAProposDuDispositifOnisep	Information AProposDuDispositif	« image », « texte »
informationEmploiStageOnisep	InformationEmploiStage	« image », « texte »
informationOrientationOnisep	InformationOrientation	« image », « texte »
informationLyceeOnisep	InformationLycee	« image », « texte »
informationMetierOnisep	InformationMetier	« image », « texte »
informationEtablissementPostBacOnisep	Information EtablissementPostBac	« texte »

informationDiplomePostBacOnisep	Information DiplomePostBac	« image », « texte »
informationVieEtudianteOnisep	InformationVieEtudiante	« image », « texte »

Tableau 6.6 - Instances du module « **Nature de l'information** » pour le dispositif du portail de l'Onisep

Dans le module « **Interaction de l'utilisateur avec le dispositif numérique** », nous créons deux instances décrites dans le **Tableau 6.7**.

Instance	Classe de l'instance	Propriétés et valeurs pour une instance
supportDispositifOnisep	SupportDispositif	« supportPhysique = Ordinateur » « typeSupport = SiteWeb »
interfaceUtilisateurOnisep	InterfaceUtilisateur	« typeInterface = Interface Par Page », « disponibilite = Permanent »

Tableau 6.7 - Instances du module « **Interaction de l'utilisateur avec le dispositif numérique** » pour le dispositif du portail de l'Onisep

Le module « **Services additionnels** » signale la présence d'un annuaire, d'une FAQ, de tests et d'un outil d'échange, ces informations sont présentées dans le **Tableau 6.8**.

Instance	Classe de l'instance	Propriétés et valeurs pour une instance	Relation de l'instance pour le module « Services additionnels »
onisep	DispositifNumerique	« contientFaq = Vrai », « contientAnnuaire = Faux », « contientTemoignage = Vrai »	
outil dEchangeOnisep	OutildEchange	« permetImmediatete = Faux », « aVideo = Faux », « requiereAuthentication = Faux »	Relation de type « aInterlocuteur » avec l'instance « conseiller dOrientation »
testOnisep	Test	« requiereAuthentication = Faux », « aSujetDeTest » = « Metier »	

Tableau 6.8 - Instances du module « **Services additionnels** » pour le dispositif du portail de l'Onisep

L'utilisabilité du portail de l'Onisep est mesurée de la même façon que celle de Hello Charly à l'aide de quatre personnes du projet AILES. Les valeurs des propriétés de l'instance « utilisabilitéOnisep » du module « **Utilisabilité** » sont indiquées dans le **Tableau 6.9**.

Propriété d'objet	Valeur
valeurDeNavigabilite	2,13
valeurDeContenu	2
valeurRepresentationGraphique	2,6
valeurDErgonomie	2,3
valeurImageDeMarque	2
valeurUtilisabilite	2,21

Tableau 6.9 - Valeurs des propriétés d'objet du module « **Utilisabilité** » pour le portail de l'Onisep

Dans le module « **Utilisateur** », une instance « utilisateurOnisep » du concept « Utilisateur » est créée avec les propriétés présentées dans le **Tableau 6.10**.

Propriété de l'objet	Valeur
annee	Seconde
etapeDuProcessus	Cristallisation
typeDeBaccalaureat	{Général, Technologique, Professionnel}

Tableau 6.10 - Valeurs des propriétés d'objet du module « **Utilisateur** »

Pour instancier le module « **Impact** », les mêmes questions sont posées à un autre groupe d'utilisateur de même type que pour l'instanciation de Hello Charly. Ces questions sont posées auprès de quatre-vingt-six élèves de seconde. Les moyennes de leurs réponses sont répertoriées dans le **Tableau 6.11**.

Propriété de l'objet	Valeur
valeurHedonique	4,91
valeurUtilitaire	4,85
valeurSociale	3,99
valeurGlobale	$p_1 \times 4,91 + p_2 \times 4,85 + p_3 \times 3,99$ où p_1, p_2 et p_3 sont les poids reliant la valeur globale d'impact aux valeurs des trois catégories

Tableau 6.11 - Valeurs des propriétés d'objet du module « **Impact** »

La capture d'écran de l'instanciation de l'ontologie pour le portail de l'Onisep est présentée en Annexe D.

2.3. Synthèse des instanciations

L'ontologie proposée permet de représenter les deux dispositifs considérés : Hello Charly et le portail de l'Onisep. Cette représentation permet aux professionnels de l'orientation de pouvoir prendre connaissance rapidement du contenu et du type de dispositif. L'impact est l'un des aspects que souhaitent obtenir les professionnels de l'orientation sur les dispositifs numériques. Dans cette section, nous avons présenté la formule théorique de l'impact obtenue à partir d'une régression linéaire multiple, avec la propriété « valeurGlobale » du module « **Impact** ». Nous présentons dans la section suivante l'expérimentation réalisée pour concevoir le modèle d'estimation de l'impact et la différence de valeur qu'il peut prendre selon l'approche utilisée (régression linéaire multiple, modélisation par équation structurelle, réseaux bayésiens).

3. Estimation de l'impact

3.1. Expérimentations pour créer notre modèle

Cette première expérimentation a pour objectif de concevoir le modèle d'estimation d'impact présenté dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ». Cette estimation est déterminée à partir de la valeur hédonique, de la valeur sociale et de l'utilité perçus par les lycéens.

3.1.1. Protocole de mesure

Cette expérimentation a été mise en œuvre en deux étapes :

- (1) Une expérimentation pilote, à petite échelle, pour s'assurer de la bonne compréhension des questions par les lycéens de seconde et de l'alignement des questions avec les objectifs des professionnels de l'éducation du projet AILES,
- (2) Une expérimentation pour recueillir les données quantitatives.

Des données quantitatives ont été recueillies lors de notre intervention au lycée Diderot de Langres en mai 2022. L'intervention s'est déroulée en deux temps. Les quinze premières minutes étaient destinées à la réutilisation du dispositif Hello Charly afin de s'assurer que tous les lycéens aient une vision globale claire du dispositif. Puis, quinze minutes étaient dédiées au questionnaire ; le questionnaire est composé de trente questions, deux questions sur le profil du lycéen (genre et classe) et vingt-huit affirmations présentées dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ». Cent soixante-huit questionnaires ont été recueillis, 38% d'hommes et 61% de femmes ont contribué à cette expérimentation. Au niveau national, lors de la rentrée 2021, le pourcentage

était de 46% d'hommes et 54% de femmes (<https://www.education.gouv.fr/reperes-et-references-statistiques-2022-326939>).

Pour valider le modèle d'estimation pour chacune des approches, le coefficient de détermination (R^2) de chaque modèle est calculé. Il s'agit de la corrélation au carré entre les valeurs réelles des résultats et les valeurs prédites par le modèle. Cette valeur est comprise entre 0 et 1, plus la valeur est élevée, meilleure est l'adéquation du modèle [61].

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Où \hat{y}_i est l'impact estimé, y_i est l'impact observé, et \bar{y} est la moyenne des impacts observés.

3.1.2. Résultats obtenus

Une première étape dans la création du modèle d'estimation est de comprendre la distribution des variables. Les trois premiers graphiques de la **Figure 6.10** tracent la distribution bivariable entre l'impact et chacun des trois éléments de perception considérés. Cela permet de voir la distribution des réponses au questionnaire selon l'impact et les éléments perçus. L'histogramme de la **Figure 6.10** représente le nombre d'observation relevant de chaque groupe (un : pas du tout d'accord à sept : tout à fait d'accord) pour l'impact.

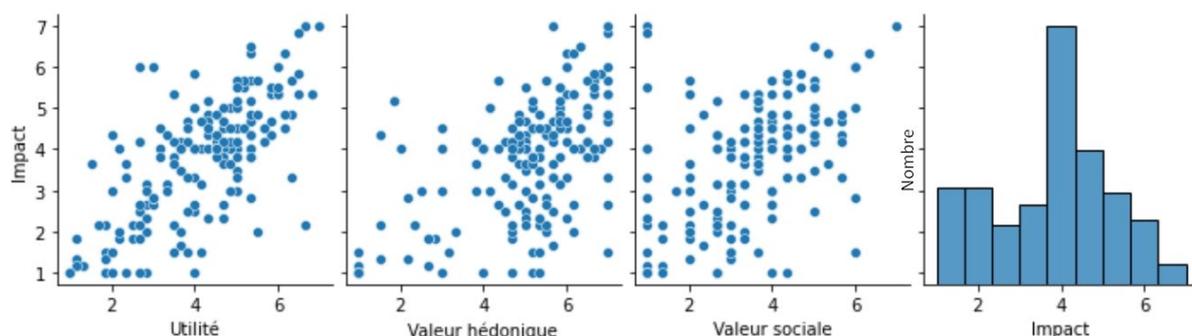


Figure 6.10 - Distribution des données collectées

3.1.2.1. Régression linéaire multiple

L'approche de régression linéaire disponible dans la bibliothèque scikit-learn (<https://scikit-learn.org>) de Python et les données présentées ci-dessus permettent d'obtenir les coefficients de corrélation du modèle suivant :

$$Impact = 0,13 \times Valeur\ Hédonique + 0,31 \times Valeur\ Sociale + 0,53 \times Utilité - 0,21$$

Ce modèle donne un plus grand poids à l'utilité, puis à la valeur sociale et enfin à la valeur hédonique. Le coefficient de détermination obtenu est 0,58.

3.1.2.2. Modélisation par équation structurelle

À partir de l'approche par équations structurelles, le modèle obtenu est présenté par la **Figure 6.11**. Il est obtenu à partir de la bibliothèque lavaan (<https://www.lavaan.ugent.be/>) de RStudio et des données présentées ci-dessus.

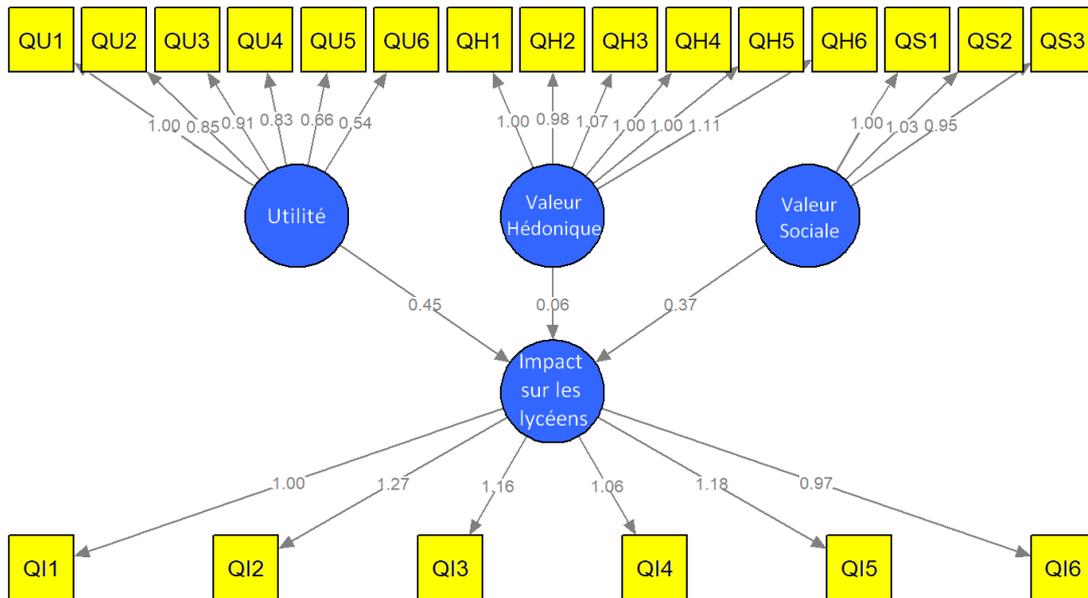


Figure 6.11 - Diagramme de cheminement PLS-SEM obtenu à partir du logiciel RStudio

Le coefficient de détermination (R^2), et l'indice comparatif d'adéquation (CFI) sont des indicateurs fréquemment utilisés pour évaluer l'adéquation d'un modèle proposé [69], [68], [152].

CFI évalue le modèle en comparant la valeur du khi-deux du modèle et le même modèle nul. Le modèle nul signifie que toutes les variables ne sont pas corrélées. Cet indice tient compte de la petite taille de l'échantillon. Une valeur seuil de 0,90 indique une bonne adéquation du modèle.

La valeur CFI du modèle est de 0,87, ce qui est inférieur au seuil de 0,90. Le coefficient de détermination obtenu est 0,73.

3.1.2.3. Réseaux bayésiens

Afin d'utiliser les réseaux bayésiens, il est nécessaire d'avoir suffisamment de données pour chaque état. Pour cette raison, nous nous plaçons dans un problème avec deux classes au lieu d'une valeur entre un et sept (1 : pas du tout d'accord à 7 : tout à fait d'accord). Ainsi, nous sommes dans un problème avec quatre variables aléatoires et deux états possibles, « Bon » si la moyenne des réponses associées à la variable aléatoire est supérieure à 4 ;

« Faible » si la moyenne des réponses associées à la variable aléatoire est inférieure ou égale à 4. Le même poids est donné à chaque affirmation du questionnaire.

Le tableau de probabilités conditionnelles obtenu à partir des données recueillies et de la librairie pyAgrum (<https://pyagrum.readthedocs.io/en/1.9.0/>) de Python est présenté dans le **Tableau 6.12**.

Utilité	Bon	Bon	Bon	Bon	Faible	Faible	Faible	Faible
Valeur hédonique	Bon	Bon	Faible	Faible	Bon	Faible	Bon	Faible
Valeur Sociale	Bon	Faible	Bon	Faible	Bon	Bon	Faible	Faible
P(Impact=Bon)	$\frac{26}{33}$	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{1}{7}$	0
P(Impact=Faible)	$\frac{7}{33}$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{4}{5}$	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{6}{7}$	1

Tableau 6.12 - Tableau de probabilités conditionnelles

Pour obtenir une unique mesure d'impact, la classe retournée par l'algorithme est la classe possédant la probabilité la plus élevée.

Précédemment, pour évaluer l'approche, nous avons utilisé le coefficient de détermination. Toutefois, ce coefficient est utilisé dans les problèmes de régression. Nous sommes ici dans un problème de classification. Ainsi, le score de précision de la classification est calculé. Ce score évalue si la classe (« Bon » ou « Faible ») prédite correspond exactement à la classe déduite de la moyenne des questions sur l'impact perçu [136]. Les valeurs possibles de ce score sont entre zéro et un. Plus le score est élevé, meilleure est la précision du modèle.

Le score de précision est calculé de la façon suivante : $\frac{\text{Nombre de classifications correctes}}{\text{Nombre total de classifications}}$.

Le score de précision obtenu pour ce modèle est de 0,81.

3.1.2.4. Discussion sur les modèles de relation entre l'impact et les indicateurs d'évaluation des dispositifs numériques

Trois modèles sont testés pour obtenir les relations entre l'impact et les indicateurs d'évaluation des dispositifs numériques. Les deux premiers modèles (régression linéaire et modélisation structurelle multiple) permettent d'avoir une valeur d'impact prédite entre un et sept, ce qui est plus précis que de caractériser l'impact uniquement à l'aide de deux classes (« Bon » ou « Faible ») comme ce qui est fait avec les réseaux bayésiens. Toutefois, l'indice comparatif d'adéquation utilisé pour vérifier l'adéquation de la modélisation structurelle multiple est inférieur au seuil. Ainsi, avec les données actuelles, ce modèle n'est pas utilisable pour prédire l'impact.

Dans la suite, les deux autres modèles sont conservés. Le modèle par régression linéaire multiple qui permet d'avoir un impact plus fin (puisque'il s'agit d'une valeur entre un et sept contre deux états pour les réseaux bayésiens), mais avec un risque d'erreur plus élevé. Les réseaux bayésiens sont moins précis avec l'utilisation de deux classes, mais avec un risque d'erreurs plus faible. De plus, les réseaux bayésiens apportent une notion intéressante de probabilité. Nous pouvons avoir la probabilité de l'état de l'impact en sachant les états de l'utilité, de la valeur sociale et de la valeur hédonique.

Finalement, dans cette section, deux modèles liant l'impact à la valeur hédonique, sociale et à l'utilité perçus sont conservés : le modèle de régression linéaire multiple et le modèle de réseaux bayésiens. Ces modèles permettent de mesurer l'impact perçu par les lycéens à partir de questions qui leur sont posées en limitant le nombre de questions et la différence potentielle entre ce qui est autodéclaré par les lycéens et ce qui est réel. L'impact est décrit de deux façons différentes afin de proposer aux professionnels de l'orientation différents éléments de comparaison : par une variable entre un et sept ou par deux états possibles (« Faible » et « Bon »).

Ces deux modèles proposés permettent de proposer une mesure de l'impact pour deux dispositifs pour lesquels il a été possible de travailler avec des lycéens : Hello Charly et le portail de l'Onisep. La section suivante présente les expérimentations réalisées et les résultats obtenus pour ces deux dispositifs.

3.2. Expérimentation de l'estimation de l'impact

À partir des deux modèles d'estimation de l'impact proposés, l'impact de Hello Charly et du portail de l'Onisep sont déterminés. Le protocole présenté dans la Section 4 du Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation » est utilisé.

Les expérimentations pour ces deux dispositifs ont été réalisées au lycée Diderot de Langres le 22 mai 2023 auprès de lycéens de seconde générale et technologique. Six classes ont été réparties sur les deux dispositifs afin que le pourcentage de filles et que la moyenne générale des classes soient les plus homogènes possible entre les deux dispositifs.

3.2.1. Mesure de l'impact du dispositif Hello Charly

L'impact est mesuré sur la version de Hello Charly accessible en mai 2023. Il s'agit de la version 10.1. Cette information est importante, car actuellement, des mises à jour sont régulièrement réalisées sur ce dispositif.

Le scénario proposé aux lycéens avant de diffuser le questionnaire est le suivant :

- *Faire des recherches sur un univers ou métier ou formation qui vous intéresse,*

- *Faire un questionnaire de 5-10 minutes (exemple : « Je mets des mots sur ma personnalité » dans la mission « Trouver mes métiers préférés »),*
- *Consulter votre profil.*

Entre quinze et vingt minutes sont utilisées pour réaliser ce scénario.

Trois classes de seconde ont participé à cette expérimentation, le pourcentage moyen de filles pour ces trois classes est de 59%, et leur moyenne générale est de 13,4/20.

A partir de ces trois classes, quatre-vingt-quatre réponses sont obtenues ; trois réponses sont éliminées, car la même réponse est donnée à toutes les questions.

Finalement, dans les quatre-vingt-une réponses conservées, 45,7% des répondants sont des hommes, 53,1% sont des femmes et 1,2% n'ont pas souhaité de préciser.

Les minimums, les premiers quartiles, les médianes, les moyennes, les troisièmes quartiles et les maximums des réponses aux affirmations pour l'utilité (1), la valeur hédonique (2) et la valeur sociale (3) sont représentées dans la **Figure 6.12**.

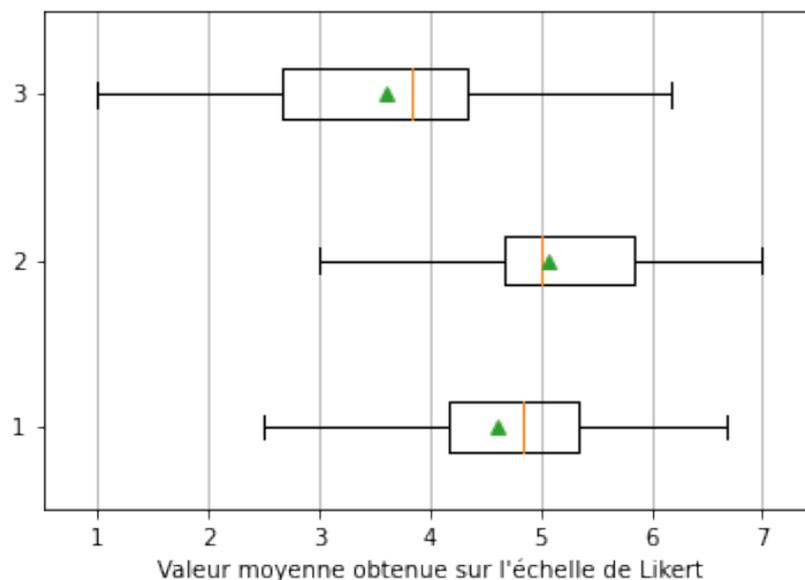


Figure 6.12 - Dispersion des moyennes pour les affirmations relatives à l'utilité (1), à la valeur hédonique (2), et à la valeur sociale (3) pour le dispositif Hello Charly

À partir de ces réponses, nous obtenons les valeurs moyennes suivantes pour le modèle de régression linéaire présenté dans la Section 3.1.2.1 :

$$\text{Utilité} = 4,62$$

$$\text{Valeur Hédonique} = 5,07$$

$$\text{Valeur Sociale} = 3,59$$

Ainsi, la valeur d'impact est calculée :

$$Impact = 0,13 \times 5,07 + 0,31 \times 3,59 + 0,53 \times 4,62 - 0,21 = 4,01$$

Pour le modèle de réseaux bayésiens présenté dans la Section 3.1.3.3, comme l'état de l'utilité, de la valeur hédonique et de la variable sociale est « Bon », la probabilité que l'impact soit « Bon » est de 0,79.

3.2.2. Mesure de l'impact du dispositif Onisep

L'impact est mesuré sur la version de l'Onisep accessible en mai 2023. Même si le portail de l'Onisep varie peu dans le temps, cette information doit être spécifiée en cas de mises à jour.

Le scénario proposé aux lycéens avant de diffuser le questionnaire est le suivant :

- *Faire des recherches sur un domaine ou métier ou activité qui vous intéresse,*
- *Chercher une formation après le bac susceptible de vous intéresser (selon un domaine ou un lieu),*
- *Aller sur le menu « Orientation » et consulter une page qui vous intéresse (exemple : « Les sections spécifiques au lycée »).*

Entre quinze et vingt minutes sont utilisées pour réaliser ce scénario.

Trois classes de seconde ont participé à cette expérimentation, le pourcentage moyen de filles pour ces trois classes est de 56%, et leur moyenne générale est de 13/20.

A partir de ces trois classes, quatre-vingt-deux réponses sont obtenues ; trois réponses sont éliminées car la même réponse est donnée à toutes les questions.

Finalement, dans les soixante-dix-neuf réponses conservées, 43% des répondants sont des hommes, 55,7% sont des femmes et 1,3% n'ont pas souhaité de préciser.

Les minimums, les premiers quartiles, les médianes, les moyennes, les troisièmes quartiles et les maximums des réponses aux affirmations pour l'utilité, la valeur hédonique et la valeur sociale sont représentées dans la **Figure 6.13**.

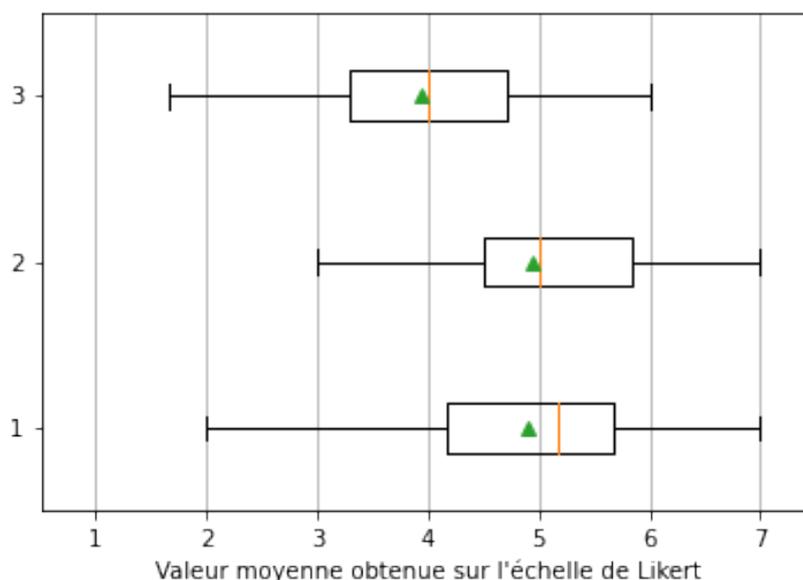


Figure 6.13 - Dispersion des moyennes pour les questions relatives à l'utilité (1), à la valeur hédonique (2), et à la valeur sociale (3) pour le dispositif de l'Onisep

À partir de ces réponses, nous obtenons les valeurs moyennes suivantes pour le modèle de régression linéaire présenté dans la Section 3.1.3.1 :

$$Utilité = 4,89$$

$$Valeur Hédonique = 4,94$$

$$Valeur Sociale = 3,94$$

Ainsi, la valeur d'impact est calculée :

$$Impact = 0,13 \times 4,94 + 0,31 \times 3,94 + 0,53 \times 4,89 - 0,21 = 4,25$$

Pour le modèle de réseaux bayésiens présenté dans la Section 3.1.3.3, l'état de l'utilité, de la valeur hédonique et de la variable sociale est « Bon », alors, à partir du **Tableau 6.12**, nous obtenons que la probabilité que l'impact soit « Bon » est de $\frac{26}{33} = 0,79$.

Ces deux expérimentations pour Hello Charly et pour le portail de l'Onisep ont permis avec le modèle d'équations linéaires multiples d'obtenir les valeurs de l'impact sur le processus d'orientation de lycéens de seconde pour ces deux dispositifs. La valeur d'impact obtenu pour le portail de l'Onisep est plus élevée que celle de Hello Charly.

Ce calcul de l'impact est réalisable grâce à la disponibilité de ces dispositifs pour les lycéens et grâce à la possibilité d'avoir accès à des classes pour diffuser les questionnaires.

La section suivante traite des expérimentations réalisées dans le cas où la diffusion de questionnaires auprès de lycéens ne serait pas possible, soit car le dispositif numérique n'est pas encore déployé, soit car la mise en place d'une expérimentation auprès de lycéens n'est pas possible.

4. Expérimentation des deux modèles de prédiction proposés

Les expérimentations de cette section ont pour objectif d'obtenir l'impact de dispositifs numériques pour lesquels nous ne pouvons pas le mesurer à partir de questionnaires auprès de lycéens. Afin de prédire l'impact, deux propositions sont faites dans la Section 3.2 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation », elles sont basées sur l'algorithme de RàPC. Cette section présente, dans ses deux premières sous-sections, les expérimentations réalisées afin d'obtenir les paramètres nécessaires à la mise en place des deux propositions. La prédiction de l'impact à partir de nos deux propositions est faite dans la troisième section.

4.1. Importance des caractéristiques des dispositifs numériques

Dans le Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation », dix-huit parties sont présentées pour caractériser les dispositifs numériques (Informations A propos, Informations Diplôme postbac, Informations Emploi et stage, Informations Établissement postbac, Informations Lycée, Informations Métier, Informations Orientation, Informations Psychologie, Informations Vie étudiante, Témoignage, FAQ, Annuaire, Outil d'échange, Outil de test, Support numérique, Interface Utilisateur, Utilisabilité, Caractéristiques Utilisateur). Nous supposons que toutes ces parties n'ont pas la même importance pour les utilisateurs et que cette importance est variable selon le type d'utilisateur.

Afin d'estimer l'importance de chacune de ces parties en excluant la partie Caractéristiques de l'Utilisateur, deux expérimentations sont réalisées. Ces expérimentations utilisent le processus d'analyse hiérarchique (« Analytic Hierarchy Process » AHP). Ce processus décompose les éléments qui affectent un système en plusieurs groupes et construit une structure hiérarchique [153]. Il est utilisé à plusieurs reprises dans le domaine de l'éducation pour établir des poids entre des critères [154]. Le poids entre les critères est obtenu en comparant deux à deux chaque critère.

Dix-sept caractéristiques sont à pondérer, ce qui induit cent-cinquante-deux questions. Afin de rendre le questionnaire plus abordable, les caractéristiques sont scindées en deux pour créer deux questionnaires AHP.

Le premier questionnaire compare les informations (Informations A propos, Informations Diplôme postbac, Informations Emploi et stage, Informations Établissement postbac, Informations Lycée, Informations Métier, Informations Orientation, Informations Psychologie, Informations Vie étudiante), il contient trente-six questions.

Le second questionnaire compare les critères restants (Témoignage, FAQ, Annuaire, Outil d'échange, Outil de test, Support numérique, Interface Utilisateur, Utilisabilité) et l'Information sur un sujet, il contient trente-six questions.

La scission en deux questionnaires permet de diminuer le nombre total de questions à soixante-douze.

La **Figure 6.14** présente un exemple de question pour comparer les critères suivant AHP. En répondant « 7 » du côté « Présence de témoignages », cela signifie que l'importance de la présence de témoignages est très forte par rapport à la présence d'une FAQ. En répondant « 1 », cela signifie que la présence de témoignages est aussi importante que la présence d'une FAQ. L'ensemble des questions posées est présenté dans l'Annexe E.

Lequel des critères suivants est le plus important pour vous ?																	
Extrême		Très forte		Forte		Modérée		Egale	Modérée		Forte		Très forte		Extrême		
9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Présence de témoignages	<input type="checkbox"/>	Présence d'une FAQ															

Figure 6.14 - Comparaison des critères « Témoignage » et « FAQ » pour le processus d'analyse hiérarchique

La première expérimentation est réalisée sous forme de discussion avec un psychologue de l'éducation nationale en visioconférence. La discussion est décomposée en deux temps :

- (1) Une explication des objectifs est donnée. Les objectifs sont de s'assurer de la pertinence des dix-sept parties comparées et d'obtenir un poids pour chacune d'entre elles. La population à laquelle nous nous intéressons est précisée auprès du psychologue, il s'agit des lycéens de seconde générale et technologique.
- (2) Un échange pour répondre au questionnaire AHP en s'assurant de la bonne compréhension des dix-sept parties.

La seconde expérimentation est réalisée avec six lycéens (trois filles et trois garçons) de seconde le 22 mai 2023 au lycée Diderot de Langres. La discussion est décomposée en deux temps :

- (1) Une explication sur l'objectif qui est d'obtenir l'importance de chacun des critères, et sur la façon de lire et de répondre au questionnaire.

- (2) Un temps de réponse aux questions en binôme, avec la possibilité de poser des questions pour les lycéens.

Afin de vérifier la cohérence entre les différentes réponses en AHP, le taux d'inconsistance est utilisé. L'estimation est acceptée si le taux d'inconsistance est inférieur ou égale à 10% [153].

Le questionnaire comparant les informations donne les résultats présentés dans le **Tableau 6.13**. Les poids d'un des groupes de lycéens n'ont pas été conservés, car le taux d'inconsistance était de 68%, les poids des deux autres groupes sont conservés, car le taux d'inconsistance est inférieur à 10%. Les poids varient selon les lycéens et le psychologue de l'éducation nationale. Toutefois, les informations les plus importantes (en rouge dans le **Tableau 6.13**) restent celles sur les métiers, sur les diplômes postbac et sur l'orientation, et les moins importantes (en bleu dans le **Tableau 6.13**) sont les informations présentant le dispositif numérique et celles sur les offres d'emploi et les stages. La conservation de la hiérarchie entre les critères d'information peut s'expliquer par le fait que les lycéens posent des questions aux psychologues de l'éducation nationale sur les mêmes sujets que ce qu'ils cherchent dans les dispositifs numériques.

Critère	Poids moyens des deux groupes de lycéens (en pourcentage)	Poids donné par un psychologue de l'éducation nationale (en pourcentage)
À propos	3,3	2,51
Informations Diplôme postbac	15,61	24,37
Emploi et stage	2,57	1,73
Établissement postbac	7,78	6,39
Lycée	9,25	11,02
Métier	35,98	20,23
Orientation	11,75	23,51
Psychologie	4,91	4,85
Vie étudiante	8,87	5,38
Taux d'inconsistance	9,7	9,5

Tableau 6.13 - Poids des différents types d'information possible

Le questionnaire comparant les éléments disponibles dans les dispositifs numériques donne les résultats présentés dans le **Tableau 6.14**. Les éléments les plus importants pour des élèves de seconde ne sont pas les mêmes selon les lycéens et le psychologue de l'éducation nationale

(en rouge dans le **Tableau 6.14**). Ces différences peuvent s'expliquer par le fait que lorsque les lycéens utilisent un dispositif numérique en étant encadrés par un psychologue de l'éducation nationale, ils ne vont pas l'utiliser de la même façon que s'ils l'utilisent seuls. Par exemple, ils ne vont pas chercher à échanger avec d'autres personnes sur un dispositif s'ils sont à côté d'un psychologue de l'éducation nationale et qu'ils ont la possibilité de lui poser leurs questions directement.

Les éléments les moins importants (en bleu dans le **Tableau 6.14**) diffèrent également sauf pour la présence d'un FAQ et de témoignages.

Critère	Poids moyens des trois groupes de lycéens (en pourcentage)	Poids donné par un psychologue de l'éducation nationale (en pourcentage)
Annuaire	2,41	3,79
FAQ	5,11	2,94
Information sur un sujet	6,17	17,81
Interface utilisateur	13,22	12,06
Outil d'échange	18,76	2,96
Outil de test	23,39	8,02
Support numérique	12,59	17,07
Témoignage	5,1	2,97
Utilisabilité	13,24	32,38
Taux d'inconsistance	8,65	9,40

Tableau 6.14 - Poids des différents éléments disponibles dans les dispositifs numériques

À partir des deux tableaux précédents, les poids obtenus pour les dix-sept parties à partir des réponses du psychologue de l'éducation nationale et des lycéens sont calculés et présentés dans le **Tableau 6.15**. Les poids moyens obtenus par les lycéens sont utilisés pour calculer les similarités dans les sections suivantes. Le choix est fait de conserver les poids des lycéens et non ceux du psychologue de l'éducation nationale, car plusieurs lycéens ont participé au questionnaire contre un seul psychologue de l'éducation nationale. De plus, les psychologues de l'éducation nationale ont la vision des lycéens qui utilisent avec eux les dispositifs numériques et il s'agit d'une minorité de lycéens.

Critère	Poids moyens des groupes de lycéens (en pourcentage)	Poids donné par un psychologue de l'éducation nationale (en pourcentage)
Annuaire	2,41	3,79
FAQ	5,11	2,94
Information A propos	0,2	0,45
Information Diplôme postbac	0,96	4,34
Information Emploi et stage	0,16	0,31
Information Établissement postbac	0,48	1,14
Information Lycée	0,57	1,96
Information Métier	2,22	3,6
Information Orientation	0,72	4,19
Information Psychologie	0,3	0,86
Information Vie étudiante	0,55	0,96
Interface utilisateur	13,22	12,06
Outil d'échange	18,76	2,96
Outil de test	23,39	8,02
Support numérique	12,59	17,07
Témoignage	5,1	2,97
Utilisabilité	13,24	32,38

Tableau 6.15 - Poids donnant l'importance des dix-sept parties des dispositifs numériques

Cette section a présenté les poids nécessaires pour prédire l'impact des dispositifs numériques à partir de nos deux propositions. La section suivante détaille les expérimentations réalisées pour obtenir des paramètres nécessaires à la prédiction.

4.2. Proximité entre les instances des concepts « interlocuteur », « support de l'information » et « type de test »

Dans cette section, nous décrivons l'expérimentation réalisée pour obtenir les matrices de similarité présentées dans la Section 3.2.1 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation ». Ces expérimentations sont nécessaires pour comparer certains éléments des dispositifs numériques (les types d'interlocuteur dans les outils d'échange, les sujets de test et les supports d'information) car ils ne peuvent pas être comparés autrement que par les utilisateurs ou des experts.

Pour construire ces matrices, le protocole suivant est réalisé :

- (1) Préparation du matériel : les trois matrices de similarité sont présentées chacune sur un paperboard afin de rendre l'expérimentation interactive,
- (2) Atelier au lycée Diderot de Langres le 22 mai 2023 avec huit lycéens de seconde (quatre filles et quatre garçons). Les lycéens remplissent chacune des matrices en binôme à l'aide de post-it. Sur chacun des post-it, ils choisissent d'écrire « c'est la même chose », « c'est très proche », « c'est proche plutôt proche », « c'est plutôt différent », c'est « très différent » ou « c'est totalement différent ». Un exemple d'une des matrices utilisées est présenté dans la **Figure 6.15**.

c'est "la même chose" que d'avoir des informations sous forme de ...
(je vais autant les regarder)

	Vidéo	Texte	Image	Audio
Vidéo	c'est la même chose	~~~~~	~~~~~	~~~~~
Texte	~~~~~	c'est la même chose	~~~~~	~~~~~
Image	~~~~~	~~~~~	c'est la même chose	~~~~~
Audio	~~~~~	~~~~~	~~~~~	c'est la même chose

AVoir des informations sous forme de ...

c'est la même chose ; c'est très proche ; c'est proche ;
c'est tout à fait différent ; c'est très différent ; c'est plutôt différent

Figure 6.15 - Image du paperboard utilisé pour obtenir la similarité entre les différents supports d'information dans les dispositifs numériques

- (3) Échange avec les huit lycéens pour comprendre et ajuster les différences éventuelles de réponses,

- (4) Analyse des données pour obtenir une unique matrice pour chacun des trois éléments. La réponse la plus fréquente des quatre groupes est prise pour obtenir chaque valeur de similarité. La valeur un correspond à « c'est la même chose », et la valeur zéro correspond à « c'est totalement différent ».

À l'issue de l'expérimentation, le **Tableau 6.16** est obtenu pour la similarité entre les différents supports d'information.

	Vidéo	Texte	Image	Audio
Vidéo	1			
Texte	0,4	1		
Image	0,4	0,2	1	
Audio	0,6	0,4	0,4	1

Tableau 6.16 - Valeurs de similarité entre les différents supports d'information

Par exemple, en se basant sur le **Tableau 6.16**, il est très différent pour des lycéens de seconde d'avoir sur un dispositif numérique d'aide à l'orientation des informations sous forme de texte et sous forme d'image.

Le **Tableau 6.17** est obtenu pour la similarité entre les différents interlocuteurs possibles dans les outils d'échange.

	Lycéen	Étudiant	Professeur lycée	Professeur postbac	Professionnel	Conseiller orientation	Robot
Lycéen	1						
Étudiant	0,4	1					
Professeur lycée	0	0	1				
Professeur postbac	0	0	0,6	1			
Professionnel	0	0,4	0,4	0,4	1		
Conseiller orientation	0	0	0,6	0	0	1	
Robot	0	0	0	0	0	0	1

Tableau 6.17 - Valeurs de similarité entre les différents interlocuteurs des outils d'échange

Par exemple, en se basant sur le **Tableau 6.17**, c'est « plutôt proche » (0,6/1) pour des lycéens de seconde d'échanger via un dispositif numérique avec des professeurs de lycée ou des professeurs de l'enseignement postbac. Alors que c'est « totalement différent » (0/1) pour eux d'échanger avec un étudiant et un professeur de lycée.

Le **Tableau 6.18** est obtenu pour la similarité entre les différents sujets de test.

	Métiers	Personnalité	Centre d'intérêt
Métiers	1		
Personnalité	0,2	1	
Centre d'intérêt	0,6	0,8	1

Tableau 6.18 - Valeurs de similarité entre les différents sujets de test

Par exemple, en se basant sur **Tableau 6.18**, c'est « très proche » (0,8/1) pour des lycéens de seconde de réaliser un test sur leur personnalité ou leurs centres d'intérêt.

Dans cette section, le protocole pour obtenir la similarité entre les types d'interlocuteur dans les outils d'échange, les sujets de test et les supports d'information est décrit ainsi que les résultats obtenus. Ainsi, tous les paramètres sont maintenant disponibles pour réaliser les prédictions de l'impact dans la section suivante.

4.3. Prédiction de l'impact

Actuellement, l'impact est connu uniquement pour deux dispositifs, pour Hello Charly et pour le portail de l'Onisep. Afin de tester les prédictions, nous supposons que la base de cas possède un unique cas Hello Charly (respectivement le portail de l'Onisep) pour prédire l'impact du portail de l'Onisep (respectivement de Hello Charly) et le comparer à l'impact estimé à partir des questionnaires.

L'impact d'un troisième dispositif est prédit à partir de la base de cas contenant Hello Charly et le portail de l'Onisep. La réalisation de cette prédiction est détaillée pour chacune de nos propositions. Le troisième dispositif choisi est Study Advisor (<https://www.studyadvisor.fr/>) car il se distingue des deux autres dispositifs. Il ne présente pas une large gamme d'informations comme l'Onisep et il permet d'échanger avec des personnes réelles à la différence de Hello Charly.

Enfin, pour observer la variabilité de l'impact, nous créons deux dispositifs fictifs, le « Dispositif fictif « bon » » possédant des éléments avec un poids important, et le « Dispositif fictif « faible » » possédant des éléments avec un poids faible. Nous posons l'hypothèse que l'impact du premier est élevé et l'impact du deuxième est faible. La description des cas pour ces deux dispositifs est faite dans le **Tableau 6.19**.

Critère	Dispositif fictif « bon » (présence du critère et ses propriétés éventuelles)	Dispositif fictif « faible » (présence du critère et ses propriétés éventuelles)
Annuaire	Oui	Non
FAQ	Oui	Non
Information A propos	Non	{Texte}
Information Diplôme postbac	{Image, Texte, Vidéo}	Non
Information Emploi et stage	Non	{Texte}
Information Établissement postbac	{Image, Texte, Vidéo}	Non
Information Lycée	{Image, Texte, Vidéo}	Non
Information Métier	{Image, Texte, Vidéo}	Non
Information Orientation	{Image, Texte, Vidéo}	Non
Information Psychologie	{Image, Texte, Vidéo}	Non
Information Vie étudiante	{Image, Texte, Vidéo}	Non
Interface utilisateur	{Interface Par Page, Permanent}	{Interface Par Page, Permanent}
Outil d'échange	{Étudiant, Permet instantanéité}	Non
Outil de test	{Métier, Personnalité, Centres d'intérêt}	Non
Support numérique	{Site Web}	{Site Web}
Témoignage	Oui	Non
Utilisabilité	3	0

Tableau 6.19 - Description des caractéristiques des cas des dispositifs fictifs

4.3.1. Prédiction de l'impact selon le modèle de prédiction en suivant le cycle du RàPC

Dans cette section, nous expérimentons la prédiction de l'impact à partir du modèle de RàPC comme décrit dans la Section 3.2.2 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation ».

Les résultats des prédictions sont présentés dans le **Tableau 6.20**.

Dispositif dont l'impact est prédit	Hello Charly	Portail de l'Onisep	Study Advisor	Dispositif fictif « bon »	Dispositif fictif « faible »
Dispositif(s) dans la base de cas	Portail de l'Onisep	Hello Charly	Portail de l'Onisep, Hello Charly	Portail de l'Onisep, Hello Charly	Portail de l'Onisep, Hello Charly
Dispositif le plus similaire ; valeur de similarité	Portail de l'Onisep ; 0,59	Hello Charly ; 0,59	Portail de l'Onisep ; 0,66	Portail de l'Onisep ; 0,76	Hello Charly ; 0,45
PR	- 0,13	0,13	-0,24	0,06	-0,55
Impact prédit en ajoutant PR	4,12	4,14	4,01	4,31	3,46
Impact prédit en utilisant PR comme un pourcentage de changement	3,70	4,53	3,25	4,50	1,82
Impact mesuré	4,01	4,25	Non mesuré	Non mesuré	Non mesuré

Tableau 6.20 - Impact prédit et mesuré pour trois dispositifs numériques basé sur le cycle du RàPC

Les étapes pour obtenir les deux valeurs de prédiction du **Tableau 6.20** pour le dispositif Study Advisor sont décrites ci-dessous :

a) Récupération et réutilisation

Les similarités, entre Study Advisor et les dispositifs de la base de cas, sont calculées. La similarité entre Study Advisor et Hello Charly est de 0,45. La similarité entre Study Advisor et le portail de l'Onisep est de 0,66.

Ainsi, le cas source ayant la valeur de similarité globale la plus élevée avec Study Advisor est le portail de l'Onisep. Alors, la solution potentielle de Study Advior est la même que celle de l'Onisep, c'est-à-dire 4,25.

b) Révision

La révision s'effectue en modifiant la solution potentielle selon un paramètre de révision PR afin de mieux correspondre au cas cible.

PR est calculé en observant les différences entre Study Advisor et le portail de l'Onisep, sa valeur est de $-0,2354$.

À partir de ce Paramètre de Révision, nous proposons deux révisions possibles sur la valeur solution $Sol(C)$:

b.1) Le Paramètre de Révision est ajouté à la valeur solution du cas source :

$$Sol(Study\ Advisor) = Sol(Onisep) + PR = 4,25 - 0,2354 = 4,01$$

b.2) Le Paramètre de Révision est utilisé comme un pourcentage de changement à effectuer sur la solution du cas source :

$$Sol(StudyAdvisor) = Sol(Onisep) + (Sol(Onisep) \times PR)$$

$$Sol(StudyAdvisor) = 4,25 - 4,25 \times 0,2354 = 3,25$$

Les propositions de prédiction de l'impact basées sur le RàPC sont utilisées dans cette section pour prédire l'impact de Hello Charly, du portail de l'Onisep et de Study Advisor.

Dans la section suivante, le Raisonnement à Partir des Attributs (RàPA) est utilisé pour prédire l'impact de ces trois dispositifs.

4.3.2. Prédiction de l'impact selon le modèle RàPA

Dans cette section, nous expérimentons la prédiction de l'impact à partir de l'RàPA comme décrit dans la Section 3.2.3 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation ».

Les résultats des prédictions sont présentés dans le **Tableau 6.21**.

Dispositif dont l'impact est prédit	Hello Charly	Portail de l'Onisep	Study Advisor	Dispositif fictif « bon »	Dispositif fictif « faible »
Dispositif(s) dans la base de cas	Portail de l'Onisep	Hello Charly	Portail de l'Onisep, Hello Charly	Portail de l'Onisep, Hello Charly	Portail de l'Onisep, Hello Charly
Impact prédit à partir de du RàPA	4,25	4,01	4,15	4,13	4,07
Impact mesuré	4,01	4,25	Non mesuré	Non mesuré	Non mesuré

Tableau 6.21 - Impact prédit et mesuré pour trois dispositifs numériques basé sur le RàPA

Pour obtenir la valeur de prédiction du **Tableau 6.21** pour le dispositif Study Advisor, les dix-sept critères des dispositifs numériques sont comparés comme présenté dans le **Tableau 6.22**.

Par exemple, pour le critère « Annuaire », la similarité entre Study Advisor et les deux cas source est la même, ainsi la part de l'impact de ce critère pour Study Advisor est calculé de la façon suivante :

$$w_{\text{Annuaire}} \times \left(\frac{\text{Impact}_{\text{Hello Charly}} + \text{Impact}_{\text{Onisep}}}{2} \right) = 0,0241 \times \left(\frac{4,01 + 4,25}{2} \right) = 0,0995$$

où w_{Annuaire} est le poids du critère « Annuaire » comme défini dans le **Tableau 6.15**.

Critère	Onisep (présence du critère et ses propriétés éventuelles)	Hello Charly (présence du critère et ses propriétés)	Study Advior (présence du critère et ses propriétés)	Cas source avec la plus grande similarité locale pour le critère considéré	Part de l'impact des cas sources utilisée pour Study Advisor
Annuaire	Non	Non	Oui	Onisep et Hello Charly	0,0995
FAQ	Oui	Non	Oui	Onisep et Hello Charly	0,2171
Information A propos	{Image, Texte}	{Image, Texte}	{Image, Texte}	Onisep et Hello Charly	0,083
Information Diplôme postbac	{Image, Texte}	{Texte}	{Image, Texte, Vidéo}	Onisep	0,0408
Information Emploi et stage	{Image, Texte}	Non	{Image, Texte, Vidéo}	Onisep	0,0068
Information Établissement postbac	{Texte}	Non	{Image, Texte, Vidéo}	Onisep	0,0204
Information Lycée	{Image, Texte}	{Texte}	{Image, Texte}	Onisep	0,0242
Information Métier	{Image, Texte}	{Image, Texte}	{Texte}	Onisep et Hello Charly	0,0917
Information Orientation	{Image, Texte}	{Image, Texte}	{Image, Texte}	Onisep et Hello Charly	0,0297
Information Psychologie	Non	{Image, Texte}	Non	Onisep	0,0128

Information Vie étudiante	{Image, Texte}	Non	{Image, Texte}	Onisep	0,0234
Interface utilisateur	{Interface Par Page, Permanent}	{Interface Par Page, Permanent}	{Interface Par Page, Permanent}	Onisep et Hello Charly	0,5460
Outil d'échange	{Psychologue de l'éducation nationale}	{Robot, Permet instantanéité, Nécessite authentification}	{Etudiant}	Onisep	0,7973
Outil de test	{Métier}	{Métier, Personnalité, Centres d'intérêt, Nécessite authentification}	Non	Onisep et Hello Charly	0,9660
Support numérique	{Site Web}	{Site Web}	{Site Web}	Onisep et Hello Charly	0,5110
Témoignage	Oui	Non	Oui	Onisep et Hello Charly	0,2167
Utilisabilité	2,21	1,9	1,63	Hello Charly	0,5309

Tableau 6.22 - Tableau présentant le détail de la prédiction de l'impact de Study Advisor par RàPA

La somme des éléments de la colonne « Part de l'impact des cas sources utilisée » donne la valeur de l'impact prédit de Study Advisor à partir de du RàPA qui est de 4,15.

En utilisant ce modèle, l'impact prédit pour un dispositif ne peut pas être plus bas (respectivement plus haut) que le dispositif dans la base de cas avec le plus petit impact (respectivement avec le plus grand impact).

4.3.3. Synthèse des prédictions

Selon la méthode de prédiction utilisée, l'impact prédit de Study Advisor va de 3,25 à 4,15 sur une échelle de un à sept. Entre le dispositif fictif supposé « bon » et le dispositif fictif supposé « faible », l'impact prédit va de 4,5 à 1,82 ; nous ne sommes pas sur une variabilité entre un et sept comme ce qui pourrait être attendu. Toutefois, cela peut s'expliquer par le manque de similarité des dispositifs de la base de cas. Afin de déterminer la pertinence des prédictions, et de déterminer la proposition de prédiction la plus adaptée, une plus grande base de données serait nécessaire en utilisant un ensemble de dispositifs de test pour comparer les prédictions à la valeur d'impact mesuré à l'aide des questionnaires.

Il est intéressant de se demander si les poids obtenus à l'aide du questionnaire AHP influencent fortement la valeur de l'impact prédit. Si c'est vrai, nous pouvons craindre que des erreurs dans les réponses fassent grandement changer la valeur de l'impact.

4.4. Étude de la sensibilité de l'impact selon le poids des critères

Une expérimentation est réalisée afin d'étudier la sensibilité de la valeur de l'impact prédit selon le poids des critères. Cette étude de sensibilité est intéressante car il est possible que les poids obtenus à l'aide d'un questionnaire AHP, comme décrit dans la Section 4.1, puissent avoir un pourcentage d'erreur.

Pour chacun des dix-sept critères, un tirage aléatoire est réalisé entre $-V$ et $+V$ (où V est successivement la valeur 5, 10, 15 et 20) pour obtenir aléatoirement des variations sur les poids des critères. Ces variations étant aléatoires, il faut normaliser les poids obtenus afin que la somme des poids soit toujours de un. Un exemple de l'élaboration d'un vecteur de poids w'' est présenté dans le **Tableau 6.23**.

Critère	w_i : poids réel obtenu par AHP	Tirage aléatoire d'une valeur entre -5% et +5%	w_i' : poids avec la variation entre -5% et +5%	w_i'' : poids normalisé	Pourcentage de modification pour w_i''
Annuaire	2,41	4	2,51	2,48	3%
FAQ	5,11	1	5,16	5,11	0%
Information A propos	0,20	-1	0,20	0,20	-2%
Information Diplôme postbac	0,96	-3	0,93	0,92	-4%
Information Emploi et stage	0,16	-4	0,1536	0,15	-5%
Information Établissement postbac	0,48	-1	0,48	0,47	-2%
Information Lycée	0,57	-1	0,56	0,56	-2%

Information Métier	2,22	5	2,33	2,31	4%
Information Orientation	0,72	-3	0,70	0,69	-4%
Information Psychologie	0,30	-4	0,29	0,29	-5%
Information Vie étudiante	0,55	4	0,57	0,57	3%
Interface utilisateur	13,22	-4	12,69	12,57	-5%
Outil d'échange	18,76	5	19,70	19,50	4%
Outil de test	23,39	2	23,86	23,6	1%
Support numérique	12,59	-4	12,09	11,97	-5%
Témoignage	5,10	1	5,15	5,10	0%
Utilisabilité	13,24	3	13,64	13,50	2%
Somme :	99,98		101	100	
Moyenne :		0			-1

Tableau 6.23 - Exemple de l'élaboration d'un vecteur de poids modifié selon l'intervalle $[-5 ; 5]$

Pour chacun des quatre intervalles possibles ($[-5 ; 5]$, $[-10 ; 10]$, $[-15 ; 15]$, $[-20 ; 20]$), trente vecteurs de poids w'' sont calculés. Les poids modifiés sont utilisés pour calculer l'impact de Study Advisor et des deux dispositifs fictifs pour nos trois propositions de prédiction. Ainsi la moyenne de l'impact est calculée et, l'écart type des trente impacts est obtenu.

Les résultats obtenus pour les trois propositions sont présentés dans le **Tableau 6.24**, le **Tableau 6.25** et le **Tableau 6.26**. Les valeurs d'écart type étant très faibles même pour l'intervalle $[-20\% ; 20\%]$, cela indique que même avec d'autres poids, si des erreurs ont été faites en répondant au questionnaire AHP, cela a peu d'influence sur la valeur d'impact.

Intervalle de variation des poids	Study Advisor		Dispositif fictif « bon »		Dispositif fictif « faible »	
	Impact moyen	Écart-type	Impact moyen	Écart-type	Impact moyen	Écart-type
[-5% ; 5%]	4,02	0,005	4,31	0,003	3,46	0,005
[-10% ; 10%]	4,01	0,012	4,31	0,002	3,46	0,009
[-15% ; 15%]	4,01	0,018	4,31	0,004	3,47	0,012
[-20% ; 20%]	4,02	0,023	4,31	0,005	3,47	0,018

Tableau 6.24 - Impact moyen de dispositifs numérique en ajoutant PR en faisant varier les poids

Intervalle de variation des poids	Study Advisor		Dispositif fictif « bon »		Dispositif fictif « faible »	
	Impact moyen	Écart-type	Impact moyen	Écart-type	Impact moyen	Écart-type
[-5% ; 5%]	3,25	0,021	4,50	0,005	1,82	0,021
[-10% ; 10%]	3,25	0,051	4,50	0,01	1,82	0,035
[-15% ; 15%]	3,25	0,075	4,50	0,018	1,83	0,048
[-20% ; 20%]	3,28	0,097	4,50	0,022	1,84	0,070

Tableau 6.25 - Impact moyen de dispositifs numérique en utilisant PR comme un pourcentage de changement en faisant varier les poids

Intervalle de variation des poids	Study Advisor		Dispositif fictif « bon »		Dispositif fictif « faible »	
	Impact moyen	Écart-type	Impact moyen	Écart-type	Impact moyen	Écart-type
[-5% ; 5%]	4,15	0,001	4,13	0,001	4,07	0,001
[-10% ; 10%]	4,15	0,001	4,13	0,002	4,07	0,002
[-15% ; 15%]	4,15	0,003	4,13	0,003	4,07	0,003
[-20% ; 20%]	4,15	0,003	4,13	0,004	4,07	0,003

Tableau 6.26 - Impact moyen de dispositifs numérique en utilisant le RàPA en faisant varier les poids

5. Conclusion

Ce chapitre présente les différentes expérimentations réalisées lors de cette thèse.

L'instanciation de l'ontologie proposée dans le Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation » est réalisée et présentée pour le dispositif Hello Charly et pour le portail de l'Onisep. Un protocole pour aider les professionnels à réaliser l'instanciation pour d'autres dispositifs a été conçu.

Le modèle d'estimation de l'impact présenté dans le Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation » est conçu à partir de l'expérimentation réalisée en mai 2022 sur Hello Charly. Sa conception est présentée dans ce chapitre. À partir de ce modèle d'estimation, l'impact d'autres dispositifs peut être estimé en utilisant un questionnaire auprès de lycéens. Ce questionnaire se compose de dix-huit affirmations afin de mesurer la valeur hédonique, la valeur sociale et l'utilité perçues.

Le modèle d'estimation est utilisé afin d'estimer l'impact de la version 10.1 de Hello Charly et du portail de l'Onisep.

Pour les dispositifs numériques pour lesquels il n'est pas possible d'estimer l'impact à l'aide d'un questionnaire, des propositions de modèle de prédiction sont faites dans le Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation ». Afin de les mettre en œuvre, ce chapitre présente les protocoles mis en place pour obtenir les paramètres nécessaires à la prédiction. Ces paramètres sont le poids des dix-sept critères, ils montrent l'importance des différentes parties des dispositifs numériques, et la similarité entre des éléments des dispositifs numériques. Ces protocoles sont mis en place pour les lycéens de seconde générale et technologique. Pour connaître l'impact de dispositifs numériques pour d'autres profils de lycéens, il est nécessaire de mettre en place ces protocoles auprès de lycéens du profil souhaité.

Enfin, la prédiction de l'impact de dispositifs numériques est réalisée pour Hello Charly avec une base de cas contenant le portail de l'Onisep, pour le portail de l'Onisep avec une base de cas contenant Hello Charly, pour Study Advisor et deux dispositifs fictifs avec une base de cas contenant Hello Charly et le portail de l'Onisep.

CONCLUSION ET PERSPECTIVES

1. Contexte et questions de recherche

L'utilisation croissante du numérique dans le domaine de l'éducation auprès des élèves a également touché l'orientation scolaire. Un grand nombre de dispositifs dédiés à ce domaine émerge. Face à la diversité des dispositifs numériques d'aide à l'orientation, il est difficile pour les professionnels de l'orientation d'identifier quel(s) dispositif(s) est (sont) le(s) plus approprié(s) pour un lycéen. Pour répondre à cette question, nous devons faire face à plusieurs verrous :

- (1) L'hétérogénéité des dispositifs numériques,
- (2) L'hétérogénéité des profils des lycéens,
- (3) La multiplicité des sources pouvant impacter le processus d'orientation.

La littérature décrit des dispositifs numériques dédiés à l'orientation existants ou en développement en les situant par rapport à l'étape du processus d'orientation du lycéen. Cependant, ces travaux sont spécifiques à un dispositif particulier, il n'existe pas de représentation uniforme permettant de faire ressortir les spécificités de chaque dispositif (L1). Afin de comparer les dispositifs entre eux, la mesure de leur impact apparaît comme une solution. Toutefois, lors de la mesure de l'impact de dispositifs numériques, les indicateurs utilisés sont dépendants de l'objectif de l'évaluation et des spécificités du dispositif numérique. Il n'existe pas de mesure de l'impact des dispositifs numériques d'aide à l'orientation sur le processus d'orientation des lycéens (L2). Ce manque entraîne l'absence de données à grande échelle de l'impact d'un dispositif ou d'un autre. Cela empêche l'utilisation d'algorithmes d'apprentissage qui pourraient fournir des recommandations sur les dispositifs les plus appropriés à chaque lycéen. Mesurer l'impact des dispositifs auprès des lycéens nécessite de les mobiliser spécifiquement pour notre étude, ce qui limite la possibilité de connaître l'impact pour un grand nombre de dispositifs ou un grand nombre de lycéens (L3).

L'objectif de cette thèse est de pallier ces limites en proposant une ontologie et une approche pour évaluer et prédire l'impact sur les lycéens des dispositifs numériques d'aide à l'orientation. Nous avons décomposé cet objectif en trois sous-objectifs.

Le premier consiste à analyser l'architecture des dispositifs numériques d'aide à l'orientation des lycéens afin de proposer un vocabulaire et une sémantique pour les représenter de façon uniforme, qui répond à la limite (L1) et entraîne la question de recherche :

QR1

Comment modéliser et stocker les caractéristiques des dispositifs numériques d'aide à l'orientation tout en étant robuste à l'apparition de nouveaux dispositifs ?

Le deuxième sous-objectif vise à proposer une mesure de l'impact des dispositifs numérique sur le processus d'orientation des lycéens afin de proposer un indicateur utile aux professionnels de l'éducation pour choisir les dispositifs les plus adaptés selon la classe. Il répond à la limite (L2) et renvoie à la question de recherche :

QR2

Comment mesurer l'impact des dispositifs numériques d'aide à l'orientation sur les lycéens dans leur démarche d'orientation ?

Enfin, le dernier sous-objectif est de proposer un système prédictif permettant aux professionnels de l'orientation de connaître la valeur de l'impact d'un dispositif alors qu'il n'est pas possible de le mesurer, en utilisant les caractéristiques (QR1) et en utilisant les dispositifs pour lesquels l'impact a été mesuré (QR2). Il répond à la limite (L3) et amène à la question de recherche :

QR3

Comment concevoir un système prédictif pour obtenir l'impact de dispositifs numériques pour lesquels il n'est pas possible de le mesurer et en s'appuyant sur la connaissance d'un nombre limité de dispositifs dont la valeur d'impact est connue ?

Nous résumons maintenant les contributions de cette thèse, associées aux questions de recherche.

2. Contributions

2.1. ODNAO : Une ontologie pour représenter de façon formelle les dispositifs numériques d'aide à l'orientation

Pour répondre à la question de recherche (QR1), nous avons choisi d'utiliser une ontologie pour capitaliser et stocker les données à propos des dispositifs numériques. Les ontologies apparaissent comme le modèle le plus complet pour représenter les connaissances en permettant d'ajouter des contraintes sur les concepts ou sur leurs relations et d'effectuer une vérification de la cohérence du modèle construit (Chapitre « État de l'art »). En revanche, cette approche nécessite de concevoir entièrement une ontologie, car il n'existe pas d'ontologie répondant à cette question de recherche (Section 2.3 du Chapitre « ODNAO : Une ontologie pour représenter les dispositifs numériques d'aide à l'orientation »). Par conséquent, nous avons utilisé la méthodologie « *Ontology Development 101* » [27] pour implémenter l'ontologie. Cette méthodologie se compose des étapes suivantes : la définition du domaine et de la portée, l'identification des ontologies existantes, l'élaboration d'une liste de termes, la création d'une hiérarchie de classes, la définition de la structure interne des concepts et la création des instances.

Ainsi, l'ontologie ODNAO est conçue. L'objectif d'ODNAO est de répondre aux questions de compétences suivantes :

- Quels sont les différents types de dispositifs numériques dédiés à l'orientation ?
- Quelles sont les différentes fonctionnalités des dispositifs numériques d'aide à l'orientation ?
- Quelle est la nature de l'information proposée par ces dispositifs ?
- Comment l'utilisateur interagît-il avec le dispositif ?
- Quels services supplémentaires le dispositif fournit-il ?
- Qui est l'utilisateur du dispositif ?
- Quel est l'impact de l'utilisation du dispositif sur l'utilisateur ?
- Quelle est la facilité d'utilisation du dispositif ?

Ces questions permettent de spécifier les exigences de l'ontologie.

L'ontologie conçue est une ontologie modulaire afin de faciliter la réutilisation des connaissances, d'être plus facile à construire, maintenir et remplacer, d'avoir des modules distincts selon les domaines d'expertise et de permettre une gestion et une navigation entre les modules. Elle est composée de six modules : (1) le **module Nature de l'information** permet une représentation cohérente des sujets couverts par un dispositif numérique, par exemple, des informations sur la vie étudiante fournies par des images et du texte ; (2) le **module**

Interaction de l'utilisateur avec le dispositif vise à décrire le type d'interface et de support utilisé pour accéder au dispositif numérique ; (3) le **module Services additionnels** fournit des concepts et des relations qui permettent de représenter des outils qui interagissent avec l'utilisateur et ne présentent pas d'informations directement ; (4) le **module Utilisateur** prend en compte les connaissances relatives à la période de l'année et à la classe de l'élève qui utilise le dispositif numérique ; (5) le **module Facilité d'utilisation (ou Utilisabilité)** vise à représenter la facilité d'utilisation d'un dispositif numérique ; (6) le **module Impact** donne une représentation formelle de l'impact en termes de comportement, d'attitudes et de connaissances sur le processus d'orientation des élèves.

La syntaxe et la cohérence de l'ontologie sont vérifiées à l'aide de l'utilisation d'un raisonneur présent dans Protégé et par l'outil OOPS!. L'intérêt d'ODNAO a été expérimenté lors de l'instanciation de l'ontologie pour deux dispositifs numériques (Section 2 du Chapitre « Expérimentations »), Hello Charly et le portail de l'Onisep. Ces deux instanciations permettent de valider l'ontologie en répondant aux questions de compétences pour ces deux dispositifs. Un protocole d'instanciation est également proposé afin de rendre cette tâche plus aisée pour des professionnels de l'orientation.

Cette contribution a des implications scientifiques et applicatives. D'un point de vue scientifique, l'ontologie contribue au développement des connaissances dans le domaine des dispositifs numériques en proposant une structure sous forme de modules pour représenter les dispositifs numériques. Cette structure permet de mettre en avant les principaux éléments d'un dispositif numérique, indépendamment de son domaine.

D'un point de vue applicatif, cette contribution dans le domaine de l'orientation aide les professionnels de l'orientation à avoir une vision plus claire des dispositifs numériques existants en leur proposant une structure normalisée et structurée des informations sur les dispositifs numériques.

2.2. Modèle de mesure de l'impact de dispositifs numériques d'aide à l'orientation

Pour répondre à la question (QR2) nous avons défini l'impact à partir du modèle « Knowledge-Attitude-Behaviour » (KAB). Ce modèle divise l'impact en trois processus : l'acquisition de connaissances, la création de croyances et la formation de comportement [40] (Section 3.1 du Chapitre « État de l'art » et Section 2 du Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation »). Ces trois processus peuvent être difficiles à mesurer, il existe un risque d'écart entre ce qui est déclaré par les lycéens et ce qui est réel.

Une identification des indicateurs d'évaluation des dispositifs numériques est également réalisée. À partir de cette identification, nous proposons une synthèse des indicateurs regroupés en neuf catégories : l'utilisateur, le ressenti global, la valeur perçue, la valeur émotionnelle provoquée (hédonique), la valeur sociale induite par le dispositif, l'utilité perçue du dispositif, l'utilisation du dispositif, la valeur sensorielle provoquée et le comportement futur de l'utilisateur (Section 3.1 et 3.2 du Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation »). L'utilisation de chacune de ces catégories ainsi que les trois processus du KAB doivent être analysés et définis selon le contexte de l'évaluation.

À partir de cette analyse, trois catégories d'indicateurs sont ressorties dans le contexte de l'orientation scolaire : la valeur hédonique provoquée, la valeur sociale induite et l'utilité perçue du dispositif. Chacune de ces trois catégories ainsi que l'impact KAB sont évalués à l'aide de six affirmations représentées à l'aide d'une échelle de Likert en sept points.

Pour concevoir le modèle de mesure de l'impact, nous avons fait l'hypothèse que la « Valeur hédonique provoquée », la « Valeur sociale induite » et l'« Utilité perçue » sont corrélées à l'impact. La connaissance de cette corrélation permet de limiter l'écart entre ce qui est déclaré par les lycéens pour les trois processus en estimant l'impact à partir des mesures de la valeur hédonique provoquée, de la valeur sociale induite et de l'utilité perçue du dispositif.

La recherche de corrélation est réalisée à partir de trois approches d'analyse de relation entre les variables : la régression linéaire multiple, la modélisation par équation structurelle et les réseaux bayésiens. Ces approches ont l'avantage de justifier l'estimation auprès des personnes utilisant la valeur de l'impact.

La conception du modèle de corrélation est réalisée à partir de données issues d'une première expérimentation réalisée sur la version 2022 de Hello Charly. À partir des résultats obtenus pour les trois approches, la régression linéaire multiple et les réseaux bayésiens sont retenus.

Cette mesure d'impact a été expérimentée sur deux dispositifs numériques, la version 10.1 de Hello Charly et le portail de l'Onisep accessible en mai 2023. Ces deux expérimentations permettent d'obtenir l'impact de deux dispositifs numériques et donc d'augmenter les connaissances que nous avons dessus.

Un protocole de mesure de l'impact est également proposé (Section 4 du Chapitre « Modèle de mesure de l'impact des dispositifs numériques d'aide à l'orientation ») afin d'accompagner les professionnels de l'orientation pour réaliser cette mesure et de s'assurer du caractère uniforme des mesures pour permettre une comparaison entre elles.

Cette contribution a des implications scientifiques et applicatives. D'un point de vue scientifique, ce modèle de mesure de l'impact contribue au développement de la mesure de l'impact des dispositifs numériques en proposant un ensemble de catégories possibles pour

évaluer les dispositifs numériques, permettant de distinguer les catégories pertinentes selon le contexte de l'évaluation.

D'un point de vue applicatif, cette contribution dans le domaine de l'orientation aide les professionnels de l'orientation en leur proposant un indicateur commun à tous les dispositifs numériques afin de les comparer entre eux.

2.3. Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation

Pour répondre à la question (QR3) nous avons proposé deux approches basées sur le RàPC afin de prédire l'impact de dispositifs numériques. Le RàPC a l'avantage d'avoir la capacité d'augmenter sa base de connaissances à chaque nouvelle prédiction, de pouvoir réaliser des prédictions même avec une faible quantité de données, et de pouvoir considérer les informations des experts. Chaque dispositif numérique est représenté par un cas. Cette représentation sous forme de cas permet de structurer les problèmes et de faciliter les calculs du RàPC. Les cas sources sont les cas dont la solution est connue, les cas cible sont ceux dont la solution est recherchée. La structure des cas utilisée dans cette thèse est basée sur les concepts et propriétés présentes dans l'ontologie ODNAO. Elle se compose de quarante-cinq attributs (Section 3.1 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation »).

Le RàPC fonctionne de façon cyclique, un cycle est réalisé pour prédire la solution d'un cas cible. De façon classique, ce cycle se compose de quatre étapes : (1) la récupération afin d'identifier les cas sources les plus similaires au cas cible, (2) la réutilisation des solutions des cas sources pour prédire le cas cible, (3) la révision de la solution en adaptant la solution réutilisée afin de mieux correspondre au cas cible, (4) la mémorisation éventuelle du cas cible et de sa solution dans la base de cas.

Selon le contexte et le type de données représentant les cas, les étapes de récupération, réutilisation et révision sont variables. L'utilisation du RàPC peut être problématique si le nombre de cas source est très faible, car il est alors possible que le cas cible n'ait pas de cas source similaire. Le processus de révision peut pallier à ce problème en modifiant la solution prédite. Toutefois, dans la littérature, cette étape est peu utilisée et présentée, car elle est dépendante du contexte.

Dans notre contexte, le processus de révision est important. Ainsi, trois propositions de prédiction sont réalisées. Les deux premières reprennent le cycle du RàPC en utilisant un paramètre de révision (PR) pour l'étape de révision (Section 3.2 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation »). Ce PR est calculé en observant les différences entre le cas cible et le cas source similaire. La troisième proposition

s'inspire du RàPC mais en raisonnant à l'échelle des attributs et non à celle des cas (Section 3.2.3 du Chapitre « Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation »).

Lors de l'expérimentation, cinq cas sont considérés, Hello Charly, le portail de l'Onisep, Study Advisor, et deux dispositifs fictifs, le premier pour lequel nous supposons que l'impact est élevé, le second pour lequel nous supposons que l'impact est faible. Les trois propositions de prédiction sont expérimentées pour prédire différents cas cibles en utilisant différentes bases de cas. Les trois propositions donnent différentes valeurs d'impact prédites.

Afin de déterminer la pertinence des prédictions, une plus grande base de données serait nécessaire en utilisant un ensemble de dispositifs de test pour comparer les prédictions à la valeur d'impact mesuré à l'aide du modèle de mesure d'impact.

Cette contribution a des implications scientifiques et applicatives. D'un point de vue scientifique, ce modèle de mesure de l'impact contribue au développement du RàPC :

- En proposant l'utilisation d'un paramètre de révision pour l'étape de révision, ce paramètre peut être utilisé pour prédire un impact indépendamment du domaine. Il nécessite de supposer que si le cas possède des éléments en plus, cela augmente son impact et inversement.
- En utilisant le RàPC avec une base de cas limitée.

D'un point de vue applicatif, cette contribution dans le domaine de l'orientation permet de prédire l'impact d'un dispositif numérique s'il n'est pas possible de mesurer son impact auprès des élèves.

3. Perspectives de recherche

3.1. Perspectives à court terme

3.1.1. ODNAO : Une ontologie pour représenter de façon formelle les dispositifs numériques d'aide à l'orientation

Une des limites de l'ontologie ODNAO est la nécessité de l'instancier manuellement pour chaque dispositif numérique à analyser. Afin de faciliter cette instanciation, un protocole est proposé.

Les travaux futurs peuvent être orientés sur l'enrichissement de l'ontologie. Ces travaux se sont concentrés sur les applications et sites web classiques. Il serait intéressant d'instancier des dispositifs de réalité virtuelle, il est possible que leur instanciation nécessite l'enrichissement de la sémantique et des relations de l'ontologie.

3.1.2. Modèle de mesure de l'impact de dispositifs numériques d'aide à l'orientation

Un axe d'amélioration du modèle de mesure de l'impact est le modèle d'estimation basé sur la régression linéaire multiple. Les corrélations de ce modèle sont obtenues à partir d'un dispositif numérique. Il sera intéressant de les déterminer à partir d'autres dispositifs et de comparer si les corrélations sont les mêmes ou non. Si elles sont proches, le modèle d'estimation sera vérifié et pourra être utilisé. Si les corrélations diffèrent d'un dispositif à un autre ou d'un groupe d'étudiant à un autre, cela remettra en question le modèle d'estimation, il faudra envisager d'utiliser directement un questionnaire pour mesurer l'impact en l'associant à des entretiens par exemple.

3.1.3. Modèle de prédiction de l'impact des dispositifs numériques d'aide à l'orientation

Nous avons noté deux axes d'amélioration.

Le premier concerne les poids obtenus pour représenter l'importance de chaque critère (présence d'informations sur les métiers, présence de témoignage, ...) dans le calcul de similarité. Ces poids sont obtenus pour des élèves de seconde au mois de mai. Il sera intéressant d'obtenir ces poids pour d'autres types de lycéens afin de savoir s'il existe une différence ou non dans ce qui est important dans les dispositifs numériques pour différents types de lycéens.

Un second axe d'amélioration est l'enrichissement de la base de cas. L'augmentation du nombre de cas dont l'impact est connu permettra dans un premier temps de vérifier les trois propositions faites pour réaliser la prédiction de l'impact. Et, dans un deuxième temps, une plus grande base de cas source permet d'avoir une plus grande possibilité de cas sources similaires au cas cible.

3.2. Perspectives à long terme

Plusieurs pistes de recherche ont été proposées dans la section ci-dessus. Nous proposons ci-dessous des perspectives nécessitant une temporalité plus grande et des recherches complémentaires.

Les profils des lycéens considérés dans ces travaux sont génériques, ils sont différenciés par le cursus choisi au lycée et le moment de l'année. L'obtention de la valeur d'impact mesuré pour un plus grand nombre d'élèves permettrait d'affiner l'impact moyen selon le profil. Il serait intéressant de distinguer l'impact selon des profils plus précis, par exemple le genre, la catégorie socioprofessionnelle de leur famille. Il serait dans ce cas nécessaire de rechercher

les caractéristiques pertinentes pour décrire les profils. Ces informations pourraient être intéressantes à l'échelle des professionnels au sein des lycées, car selon s'il s'agit d'un lycée rural ou urbain, l'impact d'un dispositif peut varier. Mais aussi à l'échelle des psychologues de l'éducation nationale selon le lycéen qui vient le consulter.

Nous nous sommes concentrés dans ce travail à la mesure de l'impact juste après l'utilisation du dispositif numérique en cherchant à savoir si l'utilisation du dispositif permet au lycéen de démarrer ou de poursuivre sa démarche d'orientation de façon proactive. Il serait intéressant de mesurer l'impact à plus long terme, au cours ou à la fin des études post-baccalauréat en interrogeant l'élève pour savoir s'il a eu une « bonne orientation » ou non. En effet, nous pouvons supposer que si l'élève a été proactif dans son processus d'orientation, la probabilité que son orientation soit « bonne » est plus importante que s'il ne l'a pas été. Toutefois, le processus d'orientation est très complexe et des facteurs différents de la proactivité dans ce processus influencent l'orientation. Par exemple, la situation économique familiale ou les métiers des parents peuvent influencer l'orientation.

Dans ces travaux, afin de prédire l'impact, nous avons fait l'hypothèse que l'ajout de fonctionnalités au dispositif augmente son impact par rapport à un dispositif qui ne la possède pas. Toutefois, l'augmentation du nombre de fonctionnalités peut rendre plus complexe l'utilisation du dispositif. En mesurant l'impact d'une grande variété de dispositifs numériques, il serait intéressant d'identifier le nombre de fonctionnalités qu'il est possible d'ajouter afin d'augmenter l'impact, et s'il existe un seuil de complexité à partir duquel l'impact diminue.

BIBLIOGRAPHIE

- [1] H. F. El-Sofany et S. A. El-Seoud, « Educational and Career Guidance Cloud-Based System to Improve Learning Outcomes », in *The Impact of the 4th Industrial Revolution on Engineering Education: Proceedings of the 22nd International Conference on Interactive Collaborative Learning (ICL2019)*, 2020, p. 346-357, doi: 10.1007/978-3-030-40274-7_35.
- [2] H. El Mrabet et A. A. Moussa, « Smart School Guidance and Vocational Guidance System Through the Internet of Things », in *Proceedings of the 2nd International Conference on Networking, Information Systems & Security - NISS19*, 2019, p. 1-5, doi: 10.1145/3320326.3320404.
- [3] W. L. Greenidge, « Using Virtual Reality Environments to Improve the Career Self-Efficacy of Minority Students: An Introduction », *Am. Couns. Assoc.*, p. 1-12, 2013.
- [4] I. Ruiz-Mallén, M. Heras, et K. Berrens, « Responsible research and innovation in science education: insights from evaluating the impact of using digital media and arts-based methods on RRI values », *Res. Sci. Technol. Educ.*, vol. 00, n° 00, p. 1-22, 2020, doi: 10.1080/02635143.2020.1763289.
- [5] N. Karacapilidis, *Solutions and innovations in web-based technologies for augmented learning: Improved platforms, tools, and applications*. IGI Global, 2009, doi : 10.4018/978-1-60566-238-1.
- [6] T. Benson, « Digital innovation evaluation: User perceptions of innovation readiness, digital confidence, innovation adoption, user experience and behaviour change », *BMJ Heal. Care Informatics*, vol. 26, n° 1, 2019, doi: 10.1136/bmjhci-2019-000018.
- [7] S. Fokkinga, P. Desmet, et P. Hekkert, « Impact-centered design: Introducing an integrated framework of the psychological and behavioral effects of design », *Int. J. Des.*, vol. 14, p. 97, 2020.
- [8] N. Delestre *et al.*, « NoDEfr-1 : Métadonnées pédagogiques et Web des données », 2019, [En ligne]. Disponible sur: <https://hal.archives-ouvertes.fr/hal-02161156>.
- [9] J. Bourdeau, R. Mizoguchi, Y. Hayashi, V. Psyché, et R. Nkambou, « When the Domain of the Ontology is Education », 4rth Conference on Intelligent, Interactive Learning Objects Repository Networks, Montreal, in: Proc. of the 4th Conf. on Intelligent, (I2LOR'07). pp. 1–9 2007.
- [10] M. Gribouval, C. Zanni-Merk, et D. Monticolo, « Ontology-based Representation of Digital Devices Used in Educational Guidance », in *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2022, p. 200-206, doi: 10.5220/0011551700003335.
- [11] M. Gribouval, D. Monticolo, et C. Zanni-Merk, « Assessment Methods to ICT Platform Used in Educational Guidance: Application to an Immersive World », Ireland International Conference on Education, 2021.

- [12] F.-E. Hamdani, M. Gribouval, B. Moissa, D. Monticolo, et V. Boly, « Computer-assisted Career Guidance Systems User Impact: An Evaluation study », in *2022 IEEE 28th International Conference on Engineering, Technology and Innovation (ICE/ITMC) & 31st International Association For Management of Technology (IAMOT) Joint Conference*, juin 2022, p. 1-6, doi: 10.1109/ICE/ITMC-IAMOT55089.2022.10033212.
- [13] R. Brachman, H. Levesque, et R. Reiter, *Knowledge representation*, MIT press. 1992.
- [14] G. Jakus, V. Milutinović, S. Omerović, et S. Tomažič, « Knowledge Representation », in *Concepts, Ontologies, and Knowledge Representation*, 2013, p. 47-62.
- [15] A. Patel et S. Jain, « Formalisms of Representing Knowledge », *Procedia Comput. Sci.*, vol. 125, p. 542-549, 2018, doi: 10.1016/j.procs.2017.12.070.
- [16] S. Grimm, « Knowledge Representation and Ontologies », in *Scientific Data Mining and Knowledge Discovery*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, p. 111-137.
- [17] M. Osama, A. Zaki-Ismail, M. Abdelrazek, J. Grundy, et A. Ibrahim, « SRCM: A Semi Formal Requirements Representation Model Enabling System Visualisation and Quality Checking », in *Proceedings of the 9th International Conference on Model-Driven Engineering and Software Development*, 2021, p. 278-285, doi: 10.5220/0010271202780285.
- [18] C. A. Lana, M. Guessi, P. O. Antonino, D. Rombach, et E. Y. Nakagawa, « A Systematic Identification of Formal and Semi-Formal Languages and Techniques for Software-Intensive Systems-of-Systems Requirements Modeling », *IEEE Syst. J.*, vol. 13, n° 3, p. 2201-2212, sept. 2019, doi: 10.1109/JSYST.2018.2874061.
- [19] V. Nazaruks et J. Osis, « A Survey on Domain Knowledge Representation with Frames », in *Proceedings of the 12th International Conference on Evaluation of Novel Approaches to Software Engineering*, 2017, p. 346-354, doi: 10.5220/0006388303460354.
- [20] L. T. Detwiler, J. L. V. Mejino, et J. F. Brinkley, « From frames to OWL2: Converting the Foundational Model of Anatomy », *Artif. Intell. Med.*, vol. 69, p. 12-21, mai 2016, doi: 10.1016/j.artmed.2016.04.003.
- [21] N. Guarino, G. Guizzardi, et J. Mylopoulos, « On the Philosophical Foundations of Conceptual Models », *Inf. Model. Knowl. Bases*, vol. 31, p. 1-15, 2020, doi: <https://doi.org/10.3233/FAIA200002>.
- [22] M. Szlenk, « Formal Semantics and Reasoning about UML Class Diagram », in *2006 International Conference on Dependability of Computer Systems*, 2006, p. 51-59, doi: 10.1109/DEPCOS-RELCOMEX.2006.27.
- [23] M. Mejhed Mkhinini, O. Labbani-Narsis, et C. Nicolle, « Combining UML and ontology: An exploratory survey », *Comput. Sci. Rev.*, vol. 35, p. 100223, févr. 2020, doi: 10.1016/j.cosrev.2019.100223.
- [24] J. F. Sowa, « Conceptual Graphs », in *Foundations of Artificial Intelligence*, 2008, p. 213-237.
- [25] Q. A. Hoang, « Using Conceptual Graphs to build failure knowledge model for forensic in civil engineering », Thèse réalisée à l'université de Bordeaux, 2021.

- [26] T. Nguyen et I. Baimuratov, « Logic Graphs: Complete, Semantic-Oriented and Easy to Learn Visualization Method for OWL DL Language », Majorov International Conference on Software Engineering and Computer Systems, 2020.
- [27] N. F. Noy et D. L. McGuinness, « A Guide to Creating Your First Ontology », *Biomed. Informatics Reseach*, p. 7-25, 2001, [En ligne]. Disponible sur: http://bmir.stanford.edu/file_asset/index.php/108/BMIR-2001-0880.pdf.
- [28] X. Zenuni, B. Raufi, F. Ismaili, et J. Ajdari, « State of the Art of Semantic Web for Healthcare », *Procedia - Soc. Behav. Sci.*, vol. 195, p. 1990-1998, juill. 2015, doi: 10.1016/j.sbspro.2015.06.213.
- [29] H. L. V. Minh et H. Quang, « Transformation of UML class diagram into OWL Ontology », *J. Inf. Telecommun.*, vol. 4, n° 1, 2020, doi: <https://doi.org/10.1080/24751839.2019.1686681>.
- [30] J. Tutcher, J. M. Easton, et C. Roberts, « Enabling Data Integration in the Rail Industry Using RDF and OWL: The RaCoOn Ontology », *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.*, vol. 3, n° 2, juin 2017, doi: 10.1061/AJRU6.0000859.
- [31] K. Sengupta et P. Hitzler, « Web Ontology Language (OWL) », *Encyclopedia of Social Network Analysis and Mining*. 2014, [En ligne]. Disponible sur: <https://corescholar.libraries.wright.edu/cse/184>.
- [32] S. Higgins, Z. Xiao, et M. Katsipataki, « The Impact of Digital Technology on Learning: A Summary for the Education Endowment Foundation », Full Report. *Education Endowment Foundation*, 2012.
- [33] T. Dienlin et N. Johannes, « The impact of digital technology use on adolescent well-being », *Dialogues Clin. Neurosci.*, vol. 22, n° 2, p. 135-142, juin 2020, doi: 10.31887/DCNS.2020.22.2/tdienlin.
- [34] N. Pongsakdi, A. Kortelainen, et M. Veermans, « The impact of digital pedagogy training on in-service teachers' attitudes towards digital technologies », *Educ. Inf. Technol.*, vol. 26, n° 5, p. 5041-5054, sept. 2021, doi: 10.1007/s10639-021-10439-w.
- [35] E. Lacka et T. C. Wong, « Examining the impact of digital technologies on students' higher education outcomes: the case of the virtual learning environment and social media », *Stud. High. Educ.*, vol. 46, n° 8, p. 1621-1634, août 2021, doi: 10.1080/03075079.2019.1698533.
- [36] S. Graham, D. McKeown, S. Kihara, et K. R. Harris, « A meta-analysis of writing instruction for students in the elementary grades. », *J. Educ. Psychol.*, vol. 104, n° 4, p. 879-896, nov. 2012, doi: 10.1037/a0029185.
- [37] S. Hearn et A. L. Buffardi, « What is impact? A Methods Lab publication ». Report, London: Overseas Development Institute, 2016. doi: <https://doi.org/10.13140/RG.2.1.4799.9766>.
- [38] G. Abramo, « Revisiting the scientometric conceptualization of impact and its measurement », *J. Informetr.*, vol. 12, n° 3, p. 590-597, août 2018, doi: 10.1016/j.joi.2018.05.001.

- [39] F. Toledo, M. Camargo, V. Boly, et D. Monticolo, « Evaluating the Impact of the Development of Innovative Products: A Multi-agent Systems Based Methodology », in *2011 Seventh International Conference on Signal Image Technology & Internet-Based Systems*, nov. 2011, p. 134-140, doi: 10.1109/SITIS.2011.44.
- [40] L. Liu, Y.-P. Liu, J. Wang, L.-W. An, et J.-M. Jiao, « Use of a knowledge-attitude-behaviour education programme for Chinese adults undergoing maintenance haemodialysis: Randomized controlled trial », *J. Int. Med. Res.*, vol. 44, n° 3, p. 557-568, juin 2016, doi: 10.1177/0300060515604980.
- [41] S. S. Zulkifly, S. N. S. Ismail, N. H. Hasan, M. R. Mahadi, et M. R. Baharudin, « Assessing the Level of Safety Knowledge-Attitude-Behaviour (Safety KAB): A Case Study in a Public Cleansing Firm », *J. Safety, Heal. Ergon.*, vol. 2, n° 1, p. 1-7, 2020.
- [42] P. G. Schrader et K. A. Lawless, « The knowledge, attitudes, & behaviors approach how to evaluate performance and learning in complex environments », *Perform. Improv.*, vol. 43, n° 9, p. 8-15, sept. 2004, doi: 10.1002/pfi.4140430905.
- [43] L. Shankar Iyer, « Knowledge, Attitude and Behaviour (KAB) of Student Community Towards Electronic Waste - A Case Study », *Indian J. Sci. Technol.*, vol. 11, n° 10, p. 1-9, févr. 2018, doi: 10.17485/ijst/2018/v11i10/109038.
- [44] I. Mir, « Effects of beliefs and concerns on user attitudes toward online social network advertising and their ad clicking behavior », *J. Internet Bank. Commer.*, vol. 20, p. 1-25, 2015.
- [45] H.-Y. Liu, J.-R. Chen, S.-Y. Hsiao, et S.-T. Huang, « Caregivers' oral health knowledge, attitude and behavior toward their children with disabilities », *J. Dent. Sci.*, vol. 12, n° 4, p. 388-395, déc. 2017, doi: 10.1016/j.jds.2017.05.003.
- [46] B. McKenzie *et al.*, « The Science of Salt: A focused review on salt-related knowledge, attitudes and behaviors, and gender differences », *J. Clin. Hypertens.*, vol. 20, n° 5, p. 850-866, mai 2018, doi: 10.1111/jch.13289.
- [47] D. D. M. Dolawattha, H. . Salinda Pramadasa, et P. M. Jayaweera, « The Impact Model: Teachers' Mobile Learning Adoption in Higher Education », *Int. J. Educ. Dev. Using Inf. Commun. Technol.*, vol. 15, n° 4, p. 71-88, 2019.
- [48] J. D. Olson, F. D. Appunn, C. A. McAllister, K. K. Walters, et L. Grinnell, « Webcams and virtual teams: an impact model », *Team Perform. Manag.*, vol. 20, n° 3/4, p. 148-177, juin 2014, doi: 10.1108/TPM-05-2013-0013.
- [49] E. Scornavacca, « An investigation of the factors that influence user acceptance of mobile information systems in the workplace », Thesis in Victoria University of Wellington, 2010.
- [50] E. Firmansyah, D. Herdiana, et D. Yuniarto, « Examining Readiness of E-Learning Implementation using Information System Readiness Impact Model », in *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, oct. 2020, p. 1-5, doi: 10.1109/CITSM50537.2020.9268921.
- [51] S. S. Noesgaard et R. Ørngreen, « The effectiveness of e-learning: An explorative and integrative review of the definitions, methodologies and factors that promote e-

- Learning effectiveness », *Electron. J. E-Learning*, vol. 13, n° 4, p. 278-290, 2015.
- [52] R. S. Steiner, J. Marciniak, C. S. Johnston, et A. Hirschi, « Career Preparedness in Adolescents: An Overview of Empirical Research and Suggestions for Practice », in *International Handbook of Career Guidance*, Cham: Springer International Publishing, 2019, p. 305-323.
- [53] M. Dziallas et K. Blind, « Innovation indicators throughout the innovation process: An extensive literature analysis », *Technovation*, vol. 80-81, n° February 2017, p. 3-29, 2019, doi: 10.1016/j.technovation.2018.05.005.
- [54] I. Miremedi, Y. Saboohi, et S. Jacobsson, « Assessing the performance of energy innovation systems: Towards an established set of indicators », *Energy Res. Soc. Sci.*, vol. 40, n° December 2017, p. 159-176, 2018, doi: 10.1016/j.erss.2018.01.002.
- [55] F. Zhang, S. Sun, C. Liu, et V. Chang, « Consumer innovativeness, product innovation and smart toys », *Electron. Commer. Res. Appl.*, vol. 41, n° March, p. 100974, 2020, doi: 10.1016/j.elerap.2020.100974.
- [56] S. F. Jahanmir et J. Cavadas, « Factors affecting late adoption of digital innovations », *J. Bus. Res.*, vol. 88, n° February, p. 337-343, 2018, doi: 10.1016/j.jbusres.2018.01.058.
- [57] J. Lessiter, S. Mitchell, E. Ferrari, P. Borden, H. Bakhshi, et J. Freeman, « Evaluating Immersive User Experience and Audience Impact », London, 2018. [En ligne]. Disponible sur: https://www.immerseuk.org/wp-content/uploads/2018/07/Evaluating_Immersive_User_Experience_and_Audience_Impact.pdf.
- [58] S. Rakic, N. Tasic, U. Marjanovic, S. Softic, E. Lüftenegger, et I. Turcin, « Student Performance on an E-Learning Platform: Mixed Method Approach », *Int. J. Emerg. Technol. Learn.*, vol. 15, n° 02, p. 187, janv. 2020, doi: 10.3991/ijet.v15i02.11646.
- [59] G. Grégoire, « Multiple Linear Regression », *Eur. Astron. Soc. Publ. Ser.*, vol. 66, p. 45-72, janv. 2014, doi: 10.1051/eas/1466005.
- [60] K. Schmidheiny, « The multiple linear regression model », Report : *Short Guid. to Microeconometrics*, 2013.
- [61] O. El Aissaoui, Y. El Alami El Madani, L. Oughdir, A. Dakkak, et Y. El Alloui, « A Multiple Linear Regression-Based Approach to Predict Student Performance », in *International Conference on Advanced Intelligent Systems for Sustainable Development.*, 2019, p. 9-23, doi: 10.1007/978-3-030-36653-7_2.
- [62] B. Sravani et M. M. Bala, « Prediction of Student Performance Using Linear Regression », in *2020 International Conference for Emerging Technology (INCET)*, juin 2020, p. 1-5, doi: 10.1109/INCET49848.2020.9154067.
- [63] O. D. Oyerinde et P. A. Chia, « Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression », *Int. J. Comput. Appl.*, vol. 157, n° 4, p. 37-44, 2017.
- [64] N. P. Y. Martin, « Acceptabilité, acceptation et expérience utilisateur : évaluation et modélisation des facteurs d'adoption des produits technologiques (Thèse) », Université

Rennes 2, 2018.

- [65] J. B. Ullman et P. M. Bentler, « Structural Equation Modeling », in *Handbook of Psychology, Second Edition*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2012.
- [66] Y. Fan *et al.*, « Applications of structural equation modeling (SEM) in ecological studies: an updated review », *Ecol. Process.*, vol. 5, n° 1, p. 19, déc. 2016, doi: 10.1186/s13717-016-0063-3.
- [67] Emmanuel Jakobowicz, « Contributions aux modèles d'équations structurelles à variables latentes », Thèse au Conservatoire national des arts et métiers Paris, 2007.
- [68] N. Nazari et D. M. Far, « The Relationship between Teaching Skills, Academic Emotion, Academic Stress and Mindset in University Student Academic Achievement Prediction: A PLS-SEM Approach », *J. Intellect. Disabil. - Diagnosis Treat.*, vol. 7, n° 3, p. 119-133, août 2019, doi: 10.6000/2292-2598.2019.07.03.9.
- [69] L. Lin, Z. Huang, B. Othman, et Y. Luo, « Let's make it better: An updated model interpreting international student satisfaction in China based on PLS-SEM approach », *PLoS One*, vol. 15, n° 7, p. e0233546, juill. 2020, doi: 10.1371/journal.pone.0233546.
- [70] S. Astuti et D. Rukmana, « Student satisfaction on the implementation of the online undergraduate thesis examination: a PLS-SEM analysis », *Qual. Assur. Educ.*, vol. 29, n° 4, p. 491-508, oct. 2021, doi: 10.1108/QAE-12-2020-0148.
- [71] R. Kartika, N. Saniyyah Putri, G. Mahesa Wiwenar, et M. Ohlyver, « Analysis of Student Satisfaction with Online Learning during the COVID-19 Pandemic using PLS-SEM », in *2021 4th International Conference on Mathematics and Statistics*, juin 2021, p. 90-96, doi: 10.1145/3475827.3475840.
- [72] C. M. Ringle, M. Sarstedt, et E. A. Mooi, « Response-Based Segmentation Using Finite Mixture Partial Least Squares », in *Data Mining, Annals of Information Systems*, 2010, p. 19-49.
- [73] P. Naïm, P.-H. Wuillemin, P. Leray, et O. Pourret, *Réseaux bayésiens*. 2007.
- [74] M. K. Eliasquevici, M. C. da R. Seruffo, et S. N. F. Resque, « Persistence in Distance Education: A Study Case Using Bayesian Network to Understand Retention », *Int. J. Distance Educ. Technol.*, vol. 15, n° 4, p. 61-78, oct. 2017, doi: 10.4018/IJDET.2017100104.
- [75] C. Chen, R. Madarshahian, J. Caicedo, C. Pierce, et G. Terejanu, « Bayesian Network Models for Student Knowledge Tracking in Large Classes », *ASEE Annual Conference & Exposition Proceedings*, 2016, doi: 10.18260/p.27282.
- [76] M. Hussain, W. Zhu, W. Zhang, et S. M. R. Abidi, « Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores », *Comput. Intell. Neurosci.*, vol. 2018, p. 1-21, oct. 2018, doi: 10.1155/2018/6347186.
- [77] E. Bonjour *et al.*, « Synthèse de l'enquête GDR MACS et IA: Cartographie des techniques utilisées, des domaines et objets de recherche concernés », 2022. [En ligne]. Disponible sur: https://hal.science/hal-04004385v1/file/Synthese_enquete_GDR_MACS_et_Intelligence_Artificielle_-_06-01-2022_V-2023.pdf.

- [78] H. J. Levesque, « Knowledge representation and reasoning », *Annu. Rev. Comput. Sci.*, vol. 1, p. 255-287, 1986.
- [79] Y. Wei *et al.*, « A review of data-driven approaches for prediction and classification of building energy consumption », *Renew. Sustain. Energy Rev.*, vol. 82, p. 1027-1047, févr. 2018, doi: 10.1016/j.rser.2017.09.108.
- [80] F. Le Ber, J. Lieber, et A. Napoli, « Les systèmes à base de connaissances », in *Encyclopédie de l'informatique et des systèmes d'information*, Vuibert, 2006, p. 1197–1208.
- [81] J. L. Kolodner, « Instructional Design : Case-Based Reasoning », *Encycl. Educ.*, vol. 2, 1999.
- [82] J. Kolodner, *Case-based reasoning*. Morgan Kaufmann, 2014.
- [83] N. Choudhury et S. A. Begum, « A survey on case-based reasoning in medicine », *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, 2016.
- [84] R. Bergmann, K. Althoff, M. Minor, M. Reichle, et K. Bach, « Case-Based Reasoning Introduction and Recent Developments », *Künstliche Intelligenz*, n° December 2014, p. 1-8, 2009.
- [85] S. Craw, « Case-Based Reasoning », in *Encyclopedia of Machine Learning and Data Mining*, Boston, MA: Springer US, 2017, p. 180-188.
- [86] F. V. Jensen, « Bayesian networks », *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 1, n° 3, p. 307-315, 2009, doi: 10.1002/wics.48.
- [87] I. Ben-Gal, « Bayesian networks », in *Encyclopedia of statistics in quality and reliability*, Wiley Online Library, 2008.
- [88] R. Cowell, « Introduction to Inference for Bayesian Networks », in *Learning in Graphical Models*, Dordrecht: Springer Netherlands, 1998, p. 9-26.
- [89] B. Efron, « Bayes' Theorem in the 21st Century », *Science (80-.)*, vol. 340, n° 6137, p. 1177-1178, juin 2013, doi: 10.1126/science.1236536.
- [90] D. Jakhar et I. Kaur, « Artificial intelligence, machine learning and deep learning: definitions and differences », *Clin. Exp. Dermatol.*, vol. 45, n° 1, p. 131-132, janv. 2020, doi: 10.1111/ced.14029.
- [91] B. Mahesh, « Machine Learning Algorithms - A Review », *Int. J. Sci. Res.*, vol. 9, n° 1, p. 381-386, 2020, doi: 10.21275/ART20203995.
- [92] S. Ray, « A Quick Review of Machine Learning Algorithms », in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, févr. 2019, p. 35-39, doi: 10.1109/COMITCon.2019.8862451.
- [93] P. Cunningham, M. Cord, et S. J. Delany, « Supervised Learning », in *Machine Learning Techniques for Multimedia*, Berlin, Heidelberg: Springer Berlin Heidelberg, p. 21-49.
- [94] V. Nasteski, « An overview of the supervised machine learning methods », *HORIZONS.B*, vol. 4, p. 51-62, déc. 2017, doi: 10.20544/HORIZONS.B.04.1.17.P05.

- [95] G. James, D. Witten, T. Hastie, et R. Tibshirani, « Unsupervised Learning », in *An Introduction to Statistical Learning: with Applications in R*, Springer US, 2021, p. 497-552.
- [96] J. E. van Engelen et H. H. Hoos, « A survey on semi-supervised learning », *Mach. Learn.*, vol. 109, n° 2, p. 373-440, févr. 2020, doi: 10.1007/s10994-019-05855-6.
- [97] R. S. Sutton et A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [98] Ministère de l'Éducation nationale et de la jeunesse et Ministère de l'enseignement supérieur de la recherche et de l'innovation, « Repères et Références statistiques : enseignements, formation et recherche », 2019.
- [99] E. Belskaya, E. Moldovanova, S. Rozhkova, O. Tsvetkova, et M. Chervach, « University Smart Guidance Counselling », in *Smart education and e-learning*, 2016, p. 39-49, doi: 10.1007/978-3-319-39690-3_4.
- [100] H. Hashim, « Application of Technology in the Digital Era Education », *Int. J. Res. Couns. Educ.*, vol. 1, n° 2, p. 1, mai 2018, doi: 10.24036/002za0002.
- [101] L. W. Foon, Z. N. Zainudin, Y. M. Yusop, et W. N. W. Othman, « E-Counselling: The Intention, Motivation and Deterrent among School Counsellors », *Univers. J. Educ. Res.*, vol. 8, n° 3C, p. 44-51, mars 2020, doi: 10.13189/ujer.2020.081605.
- [102] Y. Forner, « Career Indecision in Adolescents », *Trav. Hum.*, vol. Vol. 70, n° 3, p. 213-234, sept. 2007, doi: 10.3917/th.703.0213.
- [103] F. Thyron, M. Crahay, S. Marlair, C. Leroy, et F. Andry, « Blocus, un jeu stratégique pour mieux réussir à l'université ! », *Encycl. Creat. Invent. Innov. Entrep.*, p. 1616-1616, 2013, doi: 10.1007/978-1-4614-3858-8_100853.
- [104] L. Thomas, E. Orme, et F. Kerrigan, « Student Loneliness: The Role of Social Media Through Life Transitions », *Comput. Educ.*, vol. 146, p. 103754, mars 2020, doi: 10.1016/j.compedu.2019.103754.
- [105] R. Mizoguchi et J. Bourdeau, « Using Ontological Engineering to Overcome AI-ED Problems: Contribution, Impact and Perspectives », *Int. J. Artif. Intell. Educ.*, vol. 26, n° 1, p. 91-106, mars 2016, doi: 10.1007/s40593-015-0077-5.
- [106] S. Bennani, A. Maalel, et H. Ben Ghezala, « AGE-Learn: Ontology-Based Representation of Personalized Gamification in E-learning », *Procedia Comput. Sci.*, vol. 176, p. 1005-1014, 2020, doi: 10.1016/j.procs.2020.09.096.
- [107] M. Guinebert, A. Yessad, M. Muratet, et V. Luengo, « An ontology for describing scenarios of multi-players learning games: Toward an automatic detection of group interactions », in *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning*, 2017, vol. 10474 LNCS, p. 410-415, doi: 10.1007/978-3-319-66610-5_35.
- [108] T. R. Gruber, « A translation approach to portable ontology specifications », *Knowl. Acquis.*, vol. 5, n° 2, p. 199-220, juin 1993, doi: 10.1006/knac.1993.1008.
- [109] F. Gandon, « Ontology Engineering : a Survey and a Return on Experience », INRIA, 2006.

- [110] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, et R. García-Castro, « LOT: An industrial oriented ontology engineering framework », *Eng. Appl. Artif. Intell.*, vol. 111, p. 104755, mai 2022, doi: 10.1016/j.engappai.2022.104755.
- [111] M. Cristani et R. Cuel, « A Survey on Ontology Creation Methodologies », *Int. J. Semant. Web Inf. Syst.*, vol. 1, n° 2, p. 49-69, avr. 2005, doi: 10.4018/jswis.2005040103.
- [112] P. KOGUT *et al.*, « UML for ontology development », *Knowl. Eng. Rev.*, vol. 17, n° 1, p. 61-64, mars 2002, doi: 10.1017/S0269888902000358.
- [113] N. Drummond, M. Horridge, et H. Knublauch, « Protégé-OWL tutorial », 2005.
- [114] B. Caldwell *et al.*, « Web content accessibility guidelines (WCAG) 2.0 », *WWW Consort.*, vol. 290, p. 1-34, 2008, [En ligne]. Disponible sur: <https://www.w3.org/TR/WCAG20/>.
- [115] S. Peroni, « A Simplified Agile Methodology for Ontology Development », in *Lecture Notes in Computer Science*, 2017, p. 55-69.
- [116] D. Kless, S. Milton, E. Kazmierczak, et J. Lindenthal, « Thesaurus and ontology structure: Formal and pragmatic differences and similarities », *J. Assoc. Inf. Sci. Technol.*, vol. 66, n° 7, p. 1348-1366, juill. 2015, doi: 10.1002/asi.23268.
- [117] S. Ben Abbès, A. Scheuermann, T. Meilender, et M. D'Aquin, « Characterizing Modular Ontologies », in *7th International Conference on Formal Ontologies in Information Systems - FOIS 2012*, 2012, p. 13-25, [En ligne]. Disponible sur: <https://hal.science/hal-00710035>.
- [118] J. Han, K. H. Kim, W. Rhee, et Y. H. Cho, « Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation », *Comput. Educ.*, vol. 163, p. 104041, avr. 2021, doi: 10.1016/j.compedu.2020.104041.
- [119] A.-M. Martínez-Sala, J. Monserrat-Gauchi, et D. Alemany-Martínez, « User Usable Experience: A three-dimensional approach on usability in tourism websites and a model for its evaluation », *Tour. Manag. Perspect.*, vol. 33, p. 100579, janv. 2020, doi: 10.1016/j.tmp.2019.100579.
- [120] B. Alexandre, E. Reynaud, F. Osiurak, et J. Navarro, « Acceptance and acceptability criteria: a literature review », *Cogn. Technol. Work*, vol. 20, n° 2, p. 165-177, mai 2018, doi: 10.1007/s10111-018-0459-1.
- [121] R. Garrett, J. Chiu, L. Zhang, et S. D. Young, « A Literature Review: Website Design and User Engagement. », *Online J. Commun. media Technol.*, vol. 6, n° 3, p. 1-14, juill. 2016, [En ligne]. Disponible sur: <http://www.ncbi.nlm.nih.gov/pubmed/27499833>.
- [122] S. H. Mustafa et L. F. Al-Zoua'bi, « Usability of the academic websites of Jordans' Universities An evaluation study », *Conf. 9th Int. Arab Conf. Inf. Technol. Tunis.*, n° December 2009, p. 1-9, 2008, [En ligne]. Disponible sur: <http://faculty.ksu.edu.sa/lalzouabi/Documents/Usability%0Aof%0Athe%0AAcademic%0AWebsites%0AJordan's%0AUniversities%0AAn%0AEvaluation%0AStudy.pdf>.
- [123] L. M. Lozano, E. García-Cueto, et J. Muñiz, « Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales », *Methodology*, vol. 4, n° 2, p. 73-79, janv. 2008, doi: 10.1027/1614-2241.4.2.73.

- [124] C. C. Preston et A. M. Colman, « Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences », *Acta Psychol. (Amst.)*, vol. 104, n° 1, p. 1-15, mars 2000, doi: 10.1016/S0001-6918(99)00050-5.
- [125] M. McDaniel et V. C. Storey, « Evaluating Domain Ontologies », *ACM Comput. Surv.*, vol. 52, n° 4, p. 1-44, juill. 2020, doi: 10.1145/3329124.
- [126] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, et Z. Wang, « Hermit: An OWL 2 Reasoner », *J. Autom. Reason.*, vol. 53, n° 3, p. 245-269, oct. 2014, doi: 10.1007/s10817-014-9305-1.
- [127] M. Poveda-Villalón, A. Gómez-Pérez, et M. C. Suárez-Figueroa, « OOPS! (Ontology Pitfall Scanner!) », *Int. J. Semant. Web Inf. Syst.*, vol. 10, n° 2, p. 7-34, avr. 2014, doi: 10.4018/ijswis.2014040102.
- [128] S. Lacoste, S. Esparbès-Pistre, et P. Tap, « L'orientation scolaire et professionnelle comme source de stress chez les collégiens et les lycéens », *L'Orientation Sc. Prof.*, n° 34/3, p. 295-322, sept. 2005, doi: 10.4000/osp.617.
- [129] R. P. Oliveira, C. G. de Souza, A. da C. Reis, et W. M. de Souza, « Gamification in E-Learning and Sustainability: A Theoretical Framework », *Sustainability*, vol. 13, n° 21, p. 11945, oct. 2021, doi: 10.3390/su132111945.
- [130] A. Dickinger, M. Arami, et D. Meyer, « Reconsidering the Adoption Process: Enjoyment and Social Norms - Antecedents of Hedonic Mobile Technology Use », in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, 2006, p. 23a-23a, doi: 10.1109/HICSS.2006.410.
- [131] S. Choi, « The flipside of ubiquitous connectivity enabled by smartphone-based social networking service: Social presence and privacy concern », *Comput. Human Behav.*, vol. 65, p. 325-333, déc. 2016, doi: 10.1016/j.chb.2016.08.039.
- [132] R. Batra et O. T. Ahtola, « Measuring the hedonic and utilitarian sources of consumer attitudes », *Mark. Lett.*, vol. 2, n° 2, p. 159-170, avr. 1991, doi: 10.1007/BF00436035.
- [133] B. B. Hazarika, M. Mousavizadeh, et M. Tarn, « A Comparison of Hedonic and Utilitarian Digital Products Based on Consumer Evaluation and Technology Frustration », *J. Inf. Syst. Technol. Manag.*, vol. 16, p. 1-16, mai 2019, doi: 10.4301/S1807-1775201916008.
- [134] N. Gauducheau et M. Marcoccia, « Les forums Internet comme espaces de discussion entre adolescent.e.s sur l'orientation scolaire : vers la construction d'une contre-expertise ? », *L'Orientation Sc. Prof.*, n° 46/2, juin 2017, doi: 10.4000/osp.5398.
- [135] T. Nemoto et D. Beglar, « Developing Likert-Scale Questionnaires », in *Japan Association for Language Teaching*, 2013, p. 1-8.
- [136] L. Buitinck *et al.*, « Scikit-learn: Machine Learning in Python », *J. Mach. Learn. Res.*, vol. 12, p. 2825--2830, 2011.
- [137] Z.-H. Zhou, « Learnware: on the future of machine learning », *Front. Comput. Sci.*, vol. 10, n° 4, p. 589-590, 2016.
- [138] A. Aamodt et E. Plaza, « Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches », *AI Commun.*, vol. 7, n° 1, p. 39-59, 1994, doi:

- 10.3233/AIC-1994-7104.
- [139] J. Lieber, « Application of the revision theory to adaptation in case-based reasoning: The conservative adaptation », in *7th International Conference on Case-Based Reasoning*, 2007, p. 239-253.
- [140] M. K. Haouchine, « Remémoration guidée par l'adaptation et maintenance des systèmes de diagnostic industriel par l'approche du raisonnement à partir de cas (Thèse) », Université de Franche-Comté, 2009.
- [141] J. B. Lamy, B. Sekar, G. Guezennec, J. Bouaud, et B. Séroussi, « Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach », *Artif. Intell. Med.*, vol. 94, n° December 2018, p. 42-53, 2019, doi: 10.1016/j.artmed.2019.01.001.
- [142] H. Zhang, Z. Zhang, L. Zhou, et S. Wu, « Case-Based Reasoning for Hidden Property Analysis of Judgment Debtors », *Mathematics*, vol. 9, n° 13, p. 1559, juill. 2021, doi: 10.3390/math9131559.
- [143] A. Maedche et S. Staab, « Measuring Similarity between Ontologies », in *International Conference on Knowledge Engineering and Knowledge Management*, 2002, p. 251-263, doi: 10.1007/3-540-45810-7_24.
- [144] T. Slimani, « Description and Evaluation of Semantic Similarity Measures Approaches », *Int. J. Comput. Appl.*, vol. 80, n° 10, p. 25-33, oct. 2013, doi: 10.5120/13897-1851.
- [145] O. Okudan, C. Budayan, et I. Dikmen, « A knowledge-based risk management tool for construction projects using case-based reasoning », *Expert Syst. Appl.*, vol. 173, p. 114776, juill. 2021, doi: 10.1016/j.eswa.2021.114776.
- [146] B. Perez, C. Lang, J. Henriët, L. Philippe, et F. Auber, « Risk prediction in surgery using case-based reasoning and agent-based modelization », *Comput. Biol. Med.*, vol. 128, p. 104040, janv. 2021, doi: 10.1016/j.compbiomed.2020.104040.
- [147] P. Qi, F. Wang, Y. Huang, et X. Yang, « Integrating functional data analysis with case-based reasoning for hypertension prognosis and diagnosis based on real-world electronic health records », *BMC Med. Inform. Decis. Mak.*, vol. 22, n° 1, p. 149, déc. 2022, doi: 10.1186/s12911-022-01894-7.
- [148] A. Shabani, A. Paul, R. Platon, et E. Hüllermeier, « Predicting the Electricity Consumption of Buildings: An Improved CBR Approach », in *International Conference on Case-Based Reasoning*, 2016, p. 356-369, doi: 10.1007/978-3-319-47096-2_24.
- [149] R. L. De Mantaras *et al.*, « Retrieval, reuse, revision and retention in case-based reasoning », *Knowl. Eng. Rev.*, vol. 20, n° 3, p. 215-240, sept. 2005, doi: 10.1017/S0269888906000646.
- [150] S. H. El-Sappagh et M. Elmoghy, « Case based reasoning: Case representation methodologies », *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, p. 192--208.
- [151] A. Kaya, R. Ozturk, et C. Altin Gumussoy, « Usability Measurement of Mobile Applications with System Usability Scale (SUS) », in *Industrial Engineering in the Big Data Era: Selected Papers from the Global Joint Conference on Industrial Engineering and Its Application Areas*, 2019, p. 389-400, doi: 10.1007/978-3-030-03317-0_32.

- [152] G. Dash et J. Paul, « CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting », *Technol. Forecast. Soc. Change*, vol. 173, p. 121092, déc. 2021, doi: 10.1016/j.techfore.2021.121092.
- [153] T. L. Saaty, « How to make a decision: The analytic hierarchy process », *Eur. J. Oper. Res.*, vol. 48, n° 1, p. 9-26, sept. 1990, doi: 10.1016/0377-2217(90)90057-I.
- [154] X.-Y. Wang, G. Li, J.-F. Tu, N. T. T. Khuyen, et C.-Y. Chang, « Sustainable Education Using New Communication Technology: Assessment with Analytical Hierarchy Process (AHP) », *Sustainability*, vol. 13, n° 17, p. 9640, août 2021, doi: 10.3390/su13179640.

ANNEXES

TABLE DES MATIERES

Annexe A : Extrait de la première liste de termes	169
Annexe B : Questionnaire pour estimer l'impact.....	170
Annexe C : Protocole pour instancier l'ontologie ODNAO pour un dispositif numérique.....	173
Annexe D : Capture d'écran de l'instanciation du Portail de l'Onisep sur Protégé	181
Annexe E : Questionnaire AHP pour obtenir les poids des critères	182

Annexe A : Extrait de la première liste de termes

Terme	Définition	Classes de navigation	Sous Catégories
(page d')Accueil	Interface destinée au 1 ^{er} contact qu'a l'utilisateur avec le dispositif numérique. Correspond à l'adresse racine pour un site. Souvent conçue pour présenter le site Web et donner des hyperliens vers les principaux services et informations	Navigation structurelle	Support pour des informations/ressources
Accès au contenu	Outils permettant d'accéder aux ressources documentaire/au contenu du dispositif	Navigation outils	Outil
Application mobile	Programme téléchargeable et exécutable à partir du système d'exploitation d'un smartphone ou d'une tablette.	/	Support pour le dispositif num
Application ordinateur	Programme téléchargeable et exécutable à partir du système d'exploitation d'un PC / Linux / Mac	/	Support pour le dispositif num
Article/Billet	Texte traitant d'un sujet	Navigation structurelle	Support pour des informations/ressources
Authentification	Interface de connexion.	Navigation outils	Outil
Barre de recherche	Zone permettant à l'utilisateur de faire une recherche à partir de mots clefs dans tout le site, ou dans une partie du site	Navigation outils	Outil
Blog	Site/partie d'un site regroupant des billets/articles classés par ordre rétrochronologique.	Navigation structurelle	Support pour des informations/ressources
Bouton cliquable	Bouton cliquable qui vise à inciter le visiteur à effectuer une action précise et qui le dirige vers une page d'atterrissage.	Navigation outils	Outil
Carte géographique interactive	Carte géographique sur laquelle l'utilisateur peut zoomer/cliquer...		Support pour des informations/ressources
CGU	Modalités d'accès et de navigation sur le site Internet, et déterminent les droits et les obligations respectifs de l'utilisateur et de l'éditeur dans le cadre de l'utilisation du site.	Navigation outils	Informations/Contenu
Communauté d'entraide	Groupe de personne/personne seule pouvant aider/répondre à des questions. Il peut s'agir de professionnels, d'écoles, de jeunes, d'entreprises.	Navigation structurelle	Support pour des informations/ressources
Contact	Formulaire/adresse mail pour contacter les administrateurs du site.	Navigation outils	Outil
Contexte de l'utilisation du disp num.	Toutes les info pouvant être utilisées pour caractériser la situation d'une entité (= personne, endroit, objet considéré comme pertinent pour l'interaction entre un utilisateur et une application, incluant l'application et l'utilisateur eux-même).	/	Contexte
Espace enseignant	Espace du dispositif à destination des enseignants	Navigation profils	Support pour des informations/ressources
Espace étudiant	Espace du dispositif à destination des étudiants		Support pour des informations/ressources
Espace lycéen	Espace du dispositif à destination des lycéens		Support pour des informations/ressources
Espace parent	Espace du dispositif à destination des parents	Navigation profils	Support pour des informations/ressources

Annexe B : Questionnaire pour estimer l'impact

Etude d'impact - Contexte de l'orientation des lycéens



Évaluation de l'impact des plateformes digitales sur le projet d'orientation des lycéens : "Cas d'application 2"

Cette étude entre le cadre d'un projet d'Accompagnement à l'Intégration des Lycéens dans l'Enseignement Supérieur, L'orientation au cœur des parcours et des territoires. Elle est menée par le laboratoire de recherche ERPI (Équipe de Recherche sur les Processus Innovatifs), Université de Lorraine, Nancy.

Votre contribution en répondant aux questions suivantes nous aidera à atteindre notre objectif principal : Evaluer l'impact des dispositifs numériques sur l'accompagnement et l'orientation des lycéens et prédire l'impact de nouveaux dispositifs.

Nous vous remercions, et nous vous tiendrons informés des principaux résultats de cette recherche.

Bien à vous,

Équipe de Recherche sur les Processus Innovatifs (ERPI) Laboratory
Université de Lorraine
8, rue Bastien Lepage - BP 90647 - 54010 Nancy



Partie 1: Informations Générales

Cette catégorie regroupe des questions générales sur ton genre et ton niveau scolaire. Toutes les données sont anonymisées, et aucune information ne sera partagée, communiquée, ou publiée.

Tu es... *

- un homme
- une femme
- autre
- non précisé

Année d'étude *

- seconde
- première
- terminale

Type de bac *

- général
- technologique
- professionnel

Tu as utilisé le dispositif sur... *

- ordinateur
- mobile

40%

Partie 2 : Pertinence du dispositif numérique

Cette catégorie porte sur l'utilité de l'application, le contenu et les fonctions proposées.

Le dispositif a répondu à tes besoins pour ton projet d'orientation... *

	Pas du tout d'accord	Pas d'accord	Plutôt pas d'accord	Ni d'accord, ni pas d'accord	Plutôt d'accord	D'accord	Tout à fait d'accord
Facilement *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rapidement *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J'ai trouvé les réponses à toutes mes questions *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Le dispositif est adapté et personnalisé pour tes besoins individuels *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Le contenu et les fonctions présentés sur le dispositif étaient... *

	Pas du tout d'accord	Pas d'accord	Plutôt pas d'accord	Ni d'accord, ni pas d'accord	Plutôt d'accord	D'accord	Tout à fait d'accord
clairs *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
à jour *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

60%

Partie 3 : Design et fonctions techniques

Cette catégorie concerne le ressenti du lycéen vis-à-vis de la forme, le design et les fonctions techniques du dispositif numérique

Tu as trouvé que le dispositif ... *

	Pas du tout d'accord	Pas d'accord	Plutôt pas d'accord	Ni d'accord, ni pas d'accord	Plutôt d'accord	D'accord	Tout à fait d'accord
est agréable à utiliser *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
a un design moderne *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
est ludique (présente des outils d'animations, des vidéos, des images, ...) *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
est interactif *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
est facile à utiliser *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
est pratique à utiliser *	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

80%

Partie 4 : Expérience Sociale

Cette catégorie concerne les outils et les fonctions proposés sur le dispositif numérique pour assurer une expérience sociale lors de l'utilisation.

Tu as pu découvrir des expériences d'orientation d'autres personnes (cela peut être leurs sentiments, leurs craintes, leurs projets d'orientations, leurs feedbacks, leurs idées d'orientation) *

Pas du tout d'accord
 Pas d'accord
 Plutôt pas d'accord
 Ni d'accord, ni pas d'accord
 Plutôt d'accord
 D'accord
 Tout à fait d'accord

Tu as pu partager ton expérience (cela peut être ton sentiment au sujet de ton projet, tes craintes, ton idée d'orientation, tes métiers désirés, tes doutes...) *

Pas du tout d'accord
 Pas d'accord
 Plutôt pas d'accord
 Ni d'accord, ni pas d'accord
 Plutôt d'accord
 D'accord
 Tout à fait d'accord

Tu as été en mesure de parler de l'orientation avec d'autres utilisateurs : *

Pas du tout d'accord
 Pas d'accord
 Plutôt pas d'accord
 Ni d'accord, ni pas d'accord
 Plutôt d'accord
 D'accord
 Tout à fait d'accord

Tu as eu envie d'utiliser le dispositif ou certaines de ses fonctionnalités grâce à des personnes que tu connais (amis, famille,...) : *

Pas du tout d'accord
 Pas d'accord
 Plutôt pas d'accord
 Ni d'accord, ni pas d'accord
 Plutôt d'accord
 D'accord
 Tout à fait d'accord

Tu souhaites discuter avec tes proches (amis, familles,...) de ce que tu as lu/vu sur ce dispositif : *

Pas du tout d'accord
 Pas d'accord
 Plutôt pas d'accord
 Ni d'accord, ni pas d'accord
 Plutôt d'accord
 D'accord
 Tout à fait d'accord

Tu as plus confiance en ton projet d'orientation après avoir utilisé ce dispositif : *

Pas du tout d'accord
 Pas d'accord
 Plutôt pas d'accord
 Ni d'accord, ni pas d'accord
 Plutôt d'accord
 D'accord
 Tout à fait d'accord

Tu as quelque chose d'autre à ajouter sur le dispositif.

Annexe C : Protocole pour instancier l'ontologie ODNAO pour un dispositif numérique

PROTOCOLE POUR INSTANCIER UNE ONTOLOGIE D'UN DISPOSITIF D'ORIENTATION NUMÉRIQUE

Ce document permet d'accompagner l'instanciation de l'ontologie ODNAO pour un dispositif numérique. Il permet de rappeler le fonctionnement de Protégé et de proposer une suite d'étape pour limiter le risque d'oubli lors de l'instanciation

TABLE DES MATIÈRES

1. Utilisation de Protégé	2
1.1 Créer une instance pour une classe	2
1.2 Création des « Object property »	4
1.3 Création des « Data property »	5
1.4 Ajouter des annotations	6
1.5 Vérifier l'instanciation créée	6
2. Étapes pour instancier l'ontologie ODNAO pour un dispositif numérique ...	7

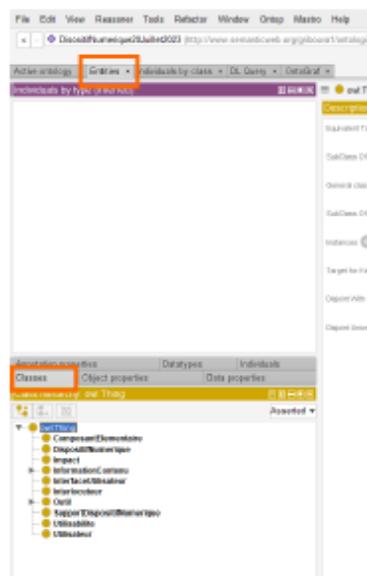
1. UTILISATION DE PROTÉGÉ

1.1 CRÉER UNE INSTANCE POUR UNE CLASSE

Afin de comprendre comment créer une instanciation d'un dispositif d'orientation numérique, nous devons présenter quelles parties de l'ontologie doivent être complétées avant d'expliquer le protocole lui-même.

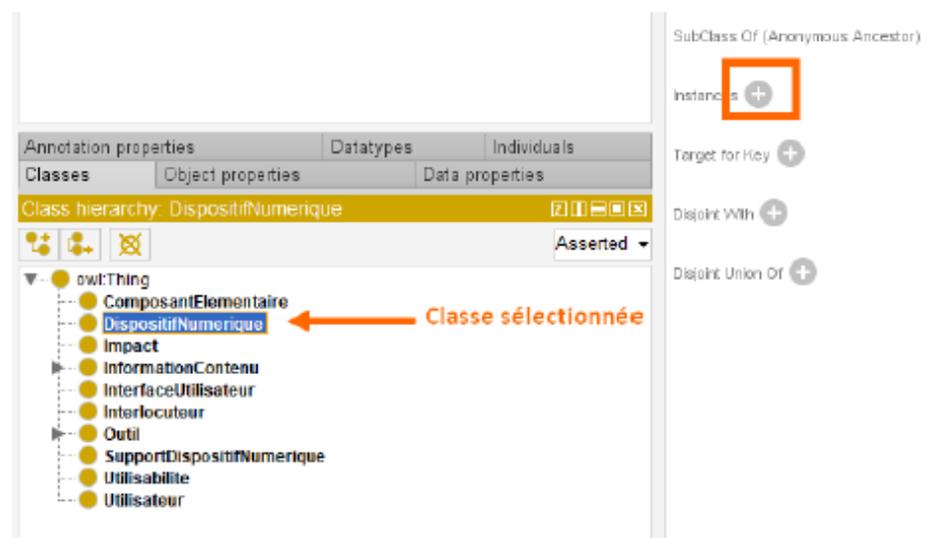
Si vous n'avez pas l'habitude d'utiliser le logiciel Protégé, logiciel utilisé pour créer et instancier nos ontologies, toutes les étapes sont guidées et expliquées.

(1) Pour afficher les classes, comme dans l'image ci-dessous, cliquez sur l'onglet « Entities » puis « Classes ».

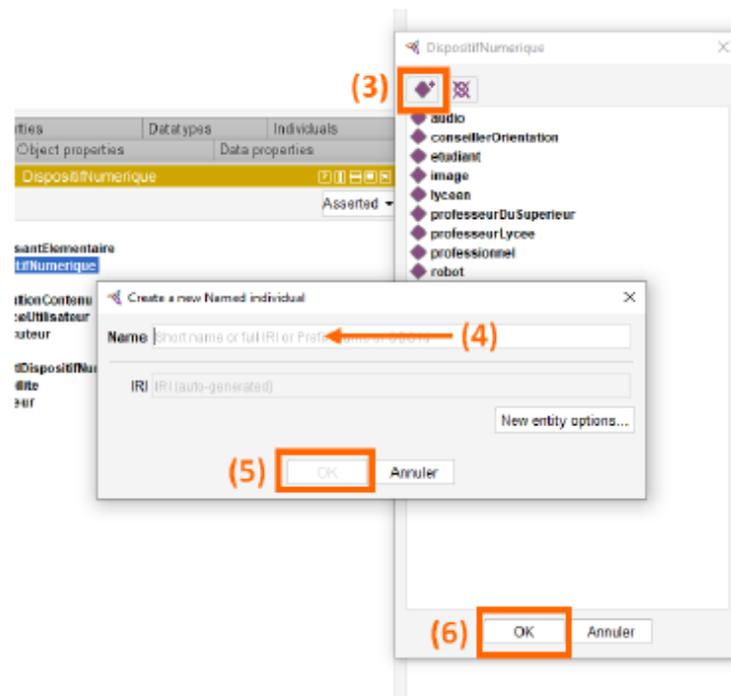


Capture d'écran de la sélection de « Entities » puis « Classes » (1)

(2) Pour créer une instance associée à une classe, cliquez sur le bouton « + » associé à Instances dans la description de la classe voulue.



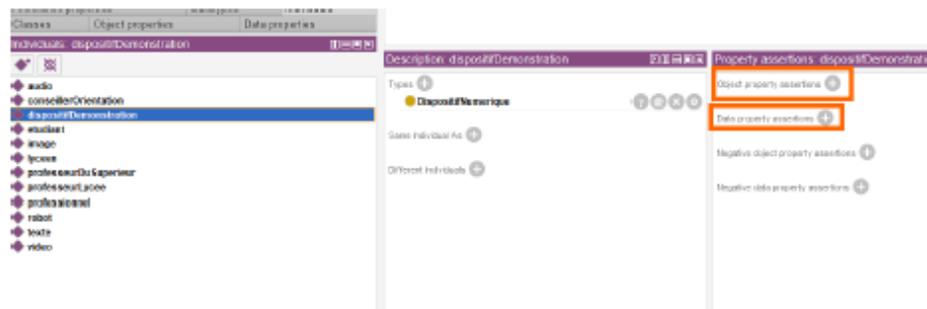
Capture d'écran de l'étape (2) de la création d'une instance pour la classe « DispositifNumerique »



Capture d'écran des étapes (3), (4), (5), (6) de la création d'une instance pour la classe « DispositifNumerique »

- (3) Cliquez sur le bouton pour créer une nouvelle instance,
- (4) Nommez votre instance et (5) confirmez en appuyant sur « OK »,
- (6) Ajoutez votre instance à l'ensemble des instances de la classe sélectionnée en cliquant sur « OK » de nouveau.,

Vous savez désormais créer une instance, l'ajouter à l'ensemble des instances d'une classe et naviguer dans l'arborescence de l'ontologie. Pour ajouter des propriétés, la méthode est identique en sélectionnant « Object property assertions » ou « Data property assertions » au sein de l'instance.



Capture d'écran présentant les boutons « Object property assertions » et « Data property assertions » pour une instance

1.2 CRÉATION DES « OBJECT PROPERTY »

Les « Object property assertions » sont les relations existantes entre les classes. Pour prendre connaissance des propriétés d'objet existantes réalisez les points 1) et 2).

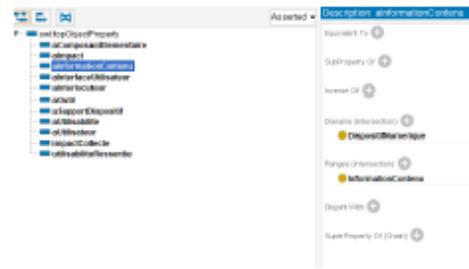
- 1) Pour chaque concept (classe) renseignez-vous grâce à sa description sur ses object properties possibles

Exemple : Sélectionnez la classe « OutilEchange », observez sa description et la rubrique, nous nous intéressons à toutes les contraintes pour la création des « Object property assertions ». Nous remarquons par exemple qu'un « OutilEchange » a exactement un interlocuteur.

Equivalent To +
 ● Outil
 and (aInterlocuteur exactly 1 Interlocuteur)
 and (aVideo only xsd:boolean)
 and (permetImmediate only xsd:boolean)

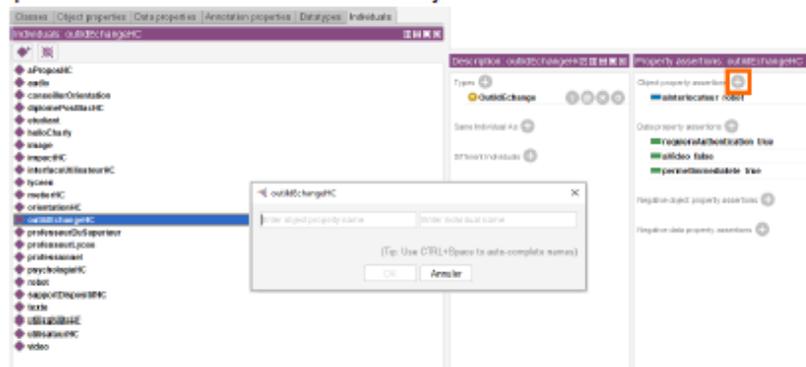
- 2) Dans l'onglet « Object properties » cherchez toutes les relations où une classe sélectionnée apparaît soit comme domaine, soit comme range

Exemple : La classe « DispostifNumerique », est le domaine dans la relation « aInformationContenu », et la classe « InformationContenu » est le range dans cette relation. Cela signifie que les instances de « DispostifNumerique » sont liées aux instances de « InformationContenu » par la relation « aInformationContenu ».



3) Ajout d'une propriété d'objet à une instance

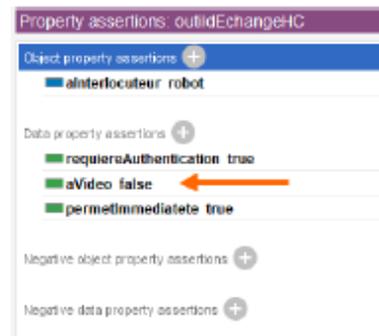
Pour ajouter une propriété d'objet à une instance, vous devez appuyer sur le bouton « + » dans la catégorie « Object property assertions » de l'instance. Ensuite entrez le nom de la relation et le nom de l'instance par lesquelles vous souhaitez relier votre objet.



1.3 CRÉATION DES « DATA PROPERTY »

Pour chaque data property, renseignez-vous sur sa description et sur le domaine auquel il appartient, puis ajoutez pour chaque instance les « data property » nécessaires.

Exemple : Pour la data property « aVideo », nous remarquons que chaque instance du type « OutidEchange » doit contenir cette data property qui est du type booléen. Si l'outil de communication permet d'utiliser une vidéo il faudra mettre « true », et « false » dans le cas inverse.



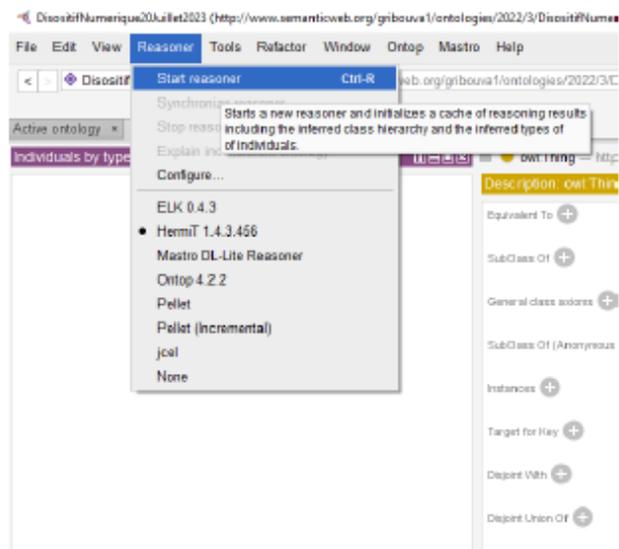
1.4 AJOUTER DES ANNOTATIONS

Pour les instances plus spécifiques, vous pouvez ajouter une annotation pour donner une explication plus claire et précise concernant une instance sans laisser aucune ambiguïté.



1.5 VÉRIFIER L'INSTANCIATION CRÉÉE

- 1) Vous pouvez utiliser le « Reasoner » qui permet de créer des inférences et de vérifier la présence d'erreurs logiques. Il est intégré directement à Protégé.



- 2) Vous pouvez utiliser le site de « Oops Ontology Pitfall Scanner ! » (<https://oops.linkeddata.es/>). L'outil OOPS! identifie des erreurs dans une ontologie. Il détecte les erreurs structurelles (la syntaxe et la sémantique formelle), les erreurs fonctionnelles (en considérant l'utilisation et la fonctionnalité prévues de l'ontologie proposée) et les erreurs de profilage de l'utilisabilité (contexte de communication d'une ontologie).

2. ÉTAPES POUR INSTANCIER L'ONTOLOGIE ODNAO POUR UN DISPOSITIF NUMÉRIQUE

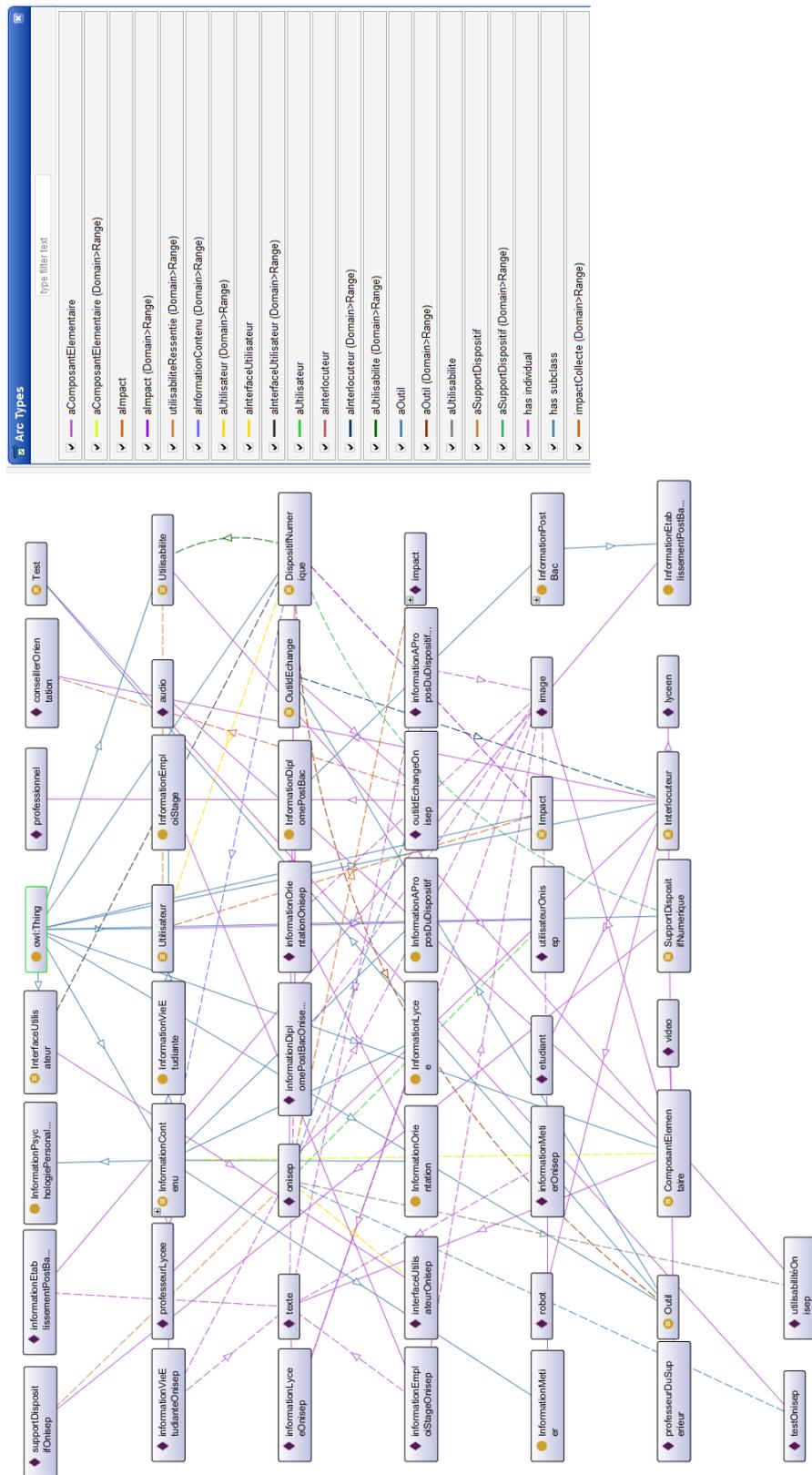
Remarque préliminaire : le nom choisi pour une instance doit commencer par une minuscule et avoir un nom compréhensible indiquant l'appartenance au dispositif analysé. Exemple : `emploiStagelInformationOnisep`

Pour instancier l'ontologie, vous devez créer des instances et définir les propriétés de ces dernières dans 8 classes différentes : « Dispositif Numérique », « Information Contenu » (qui possède 9 sous-classes), Outil (qui possède 2 sous-classes), « Interface Utilisateur », « Support Dispositif Numérique », « Utilisabilité », « Utilisateur » et « Impact ».

- 1) Dans l'onglet des « Classes », créez et nommez l'instance qui correspond à la classe « **DispositifNumérique** ». Exemple : `onisep` ; `wilbi` ; `helloCharly`.
- 2) À partir de cette instance, précisez si le dispositif contient des témoignages (data property : « `contientTemoignage` »), une FAQ (data property : « `contientFaq` ») et un annuaire (data property : « `contientAnnuaire` »). (Voir rubrique « Création des "Data Properties" »)
- 3) Après il faut s'intéresser à la classe de « **InformationContenu** ». Elle contient 9 sous-classes. Si une sous-classe existe dans le dispositif numérique, créez une instance correspondante. Exemple : `emploiStagelInformationOnisep`, classe : `EmploiStagelInformation`.
- 4) Pour chaque instance présente dans les sous-classes de « **InformationContenu** », il faut les relier aux "ComposantsElementaires" utilisés grâce à la relation "`aComposantElementaire`". Exemple : l'instance `emploiStagelInformationOnisep` est reliée à « `image` » et « `texte` ». (Voir rubrique « Création des "Object Properties" »)
- 5) Nous nous intéressons à la classe « **Outil** » qui contient 2 sous-classes. Si un outil d'échange et/ou un test existe dans le dispositif numérique, créez une instance. Exemple : `testOutilOnisep`. Pour chaque instance créée, précisez si pour accéder à l'outil, il faut s'authentifier ou non (data property : « `requiereAuthentification` »).
Si une instance de la classe « **Test** » est créée, il faut spécifier le sujet du test avec la data property « `aSujetDeTest` ».
Si une instance de la classe « **OutilEchange** » est créée, il faut spécifier l'interlocuteur (object property « `alinterlocuteur` »), l'utilisation possible d'une vidéo ou non (data property « `aVideo` »), la possibilité d'avoir une conversation instantanée (data property « `permetImmediatete` »).

- 6) Créez et nommez l'instance qui correspond à la classe « **InterfaceUtilisateur** » en précisant si l'interface est toujours disponible ou non (data property « **disponibilite** »), et le type de l'interface (data property « **typeInterface** »).
- 7) Créez et nommez l'instance qui correspond à la classe « **SupportDispositifNumerique** » en précisant le type de support (data property « **typeSupport** »), et le support physique utilisé (data property « **supportPhysique** »).
- 8) Créez et nommez l'instance qui correspond à la classe « **Utilisabilite** » en précisant les 6 valeurs associées comprises entre 0 et 3 (data properties : « **valeurDeContenu** », « **valeurDeNavigabilite** », « **valeurDErgonomie** », « **valeurImageDeMarque** », « **valeurRepresentationGraphique** », « **valeurUtilisabilite** »).
- 9) Créez et nommez l'instance qui correspond à la classe « **Utilisateur** » en précisant en quelle année au lycée se situe l'élève (data properties : « **annee** »), l'étape du processus d'orientation (data properties : « **etapeDuProcessus** ») et le type de baccalauréat (data properties : « **typeDeBaccalaureat** »).
- 10) Si l'impact du dispositif est mesuré, créez et nommez l'instance qui correspond à la classe « **Impact** » en précisant les 4 valeurs associées comprises entre 1 et 7 (data properties : « **valeurGlobale** », « **valeurHedonique** », « **valeurSociale** », « **valeurUtilitaire** »).
Si l'impact du dispositif est prédit, créez et nommez l'instance qui correspond à la classe « **Impact** » en précisant la valeur d'impact global associée comprise entre 1 et 7 (data property : « **valeurGlobale** »).

Annexe D : Capture d'écran de l'instanciation du Portail de l'Onisep sur Protégé



Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême		
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9		
Présentation du dispositif	<input type="checkbox"/>	Psychologie et Personnalité																	
Présentation du dispositif	<input type="checkbox"/>	Vie étudiante																	
Présentation du dispositif	<input type="checkbox"/>	Etablissement Post-Bac																	
Présentation du dispositif	<input type="checkbox"/>	Diplômes Post-Bac																	
Présentation du dispositif	<input type="checkbox"/>	Orientation																	
Présentation du dispositif	<input type="checkbox"/>	Métier																	
Présentation du dispositif	<input type="checkbox"/>	Lycée																	

Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême	
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Psychologie et Personnalité	<input type="checkbox"/>																	
Psychologie et Personnalité	<input type="checkbox"/>																	
Psychologie et Personnalité	<input type="checkbox"/>																	
Psychologie et Personnalité	<input type="checkbox"/>																	
Psychologie et Personnalité	<input type="checkbox"/>																	
Psychologie et Personnalité	<input type="checkbox"/>																	
Psychologie et Personnalité	<input type="checkbox"/>																	

Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême	
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Vie étudiante	<input type="checkbox"/>																	
Vie étudiante	<input type="checkbox"/>																	
Vie étudiante	<input type="checkbox"/>																	
Vie étudiante	<input type="checkbox"/>																	
Vie étudiante	<input type="checkbox"/>																	

Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême	
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Etablissement Post-Bac	<input type="checkbox"/>																	
Etablissement Post-Bac	<input type="checkbox"/>																	
Etablissement Post-Bac	<input type="checkbox"/>																	
Etablissement Post-Bac	<input type="checkbox"/>																	

Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême	
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	
Diplômes Post-Bac	<input type="checkbox"/>																	
Diplômes Post-Bac	<input type="checkbox"/>																	
Diplômes Post-Bac	<input type="checkbox"/>																	

Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême		
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9		
Orientation	<input type="checkbox"/>																		
Orientation	<input type="checkbox"/>																		
																		Métier	
																			Lycée

Laquelle des informations suivantes est la plus importante pour vous ?

	Extrême		Très forte		Forte		Modérée		Egale		Modérée		Forte		Très forte		Extrême		
	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9		
Métier	<input type="checkbox"/>																		
																			Lycée