



Titre: Utilisation d'une hybridation de la recherche opérationnelle et de l'apprentissage automatique pour injecter de l'émotion dans un agent conversationnel
Title:

Auteur: Dionisie Calistru
Author:

Date: 2022

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Calistru, D. (2022). Utilisation d'une hybridation de la recherche opérationnelle et de l'apprentissage automatique pour injecter de l'émotion dans un agent conversationnel [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/10350/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/10350/>
PolyPublie URL:

Directeurs de recherche: Michel Gendreau, & Nikolaj Van Omme
Advisors:

Programme: Maîtrise recherche en mathématiques appliquées
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Utilisation d'une hybridation de la recherche opérationnelle et de
l'apprentissage automatique pour injecter de l'émotion dans un agent
conversationnel**

DIONISIE CALISTRU

Département de de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Mathématiques appliquées

Mars 2022

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Utilisation d'une hybridation de la recherche opérationnelle et de
l'apprentissage automatique pour injecter de l'émotion dans un agent
conversationnel**

présenté par **Dionisie CALISTRU**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Antoine SAUCIER, président

Michel GENDREAU, membre et directeur de recherche

Nikolaj VAN OMME, membre et codirecteur de recherche

Yabo Octave NIAMIÉ, membre

DÉDICACE

*À ma mère, ma famille et mes amis qui m'ont soutenu.
À Nina Fu qui a toujours été à mes côtés, tes conseils m'ont été précieux.*

REMERCIEMENTS

À mon directeur Michel Gendreau et mon codirecteur Nikolaj Van Omme, sans qui ce rêve ne se serait jamais réalisé. Je vous dis merci du fond du coeur.

RÉSUMÉ

L'émergence d'agents conversationnels occupe une place de plus en plus importante dans notre quotidien. En passant par le soutien technique en ligne automatisé, l'incitation à la vente sur les réseaux sociaux, le clavardage général ou même un certain besoin d'avoir un compagnon d'assistance personnalisée auprès de nous en tout temps, nous nous exposons de plus en plus à une forme ou une autre d'agents conversationnels intelligents. Mais comment s'assurer que ces agents répondent à nos attentes émotionnelles et que les répliques générées soient contrôlées? Dans le cadre de ce projet, nous tentons de répondre à cette question en présentant une manière nouvelle et innovante de combiner la recherche opérationnelle et l'apprentissage machine pour optimiser et contrôler la manière dont un agent conversationnel transmet de l'émotion. Nous avons créé un système, appelé *Philia*, capable d'être intégré à n'importe quel agent conversationnel afin d'en améliorer les répliques en y injectant de l'émotion.

Le système que nous proposons est une première implémentation d'une composition complexe de plusieurs éléments :

- Une base de données qui collecte l'impact émotionnel des mots pour une émotion cible (la joie), mais qui peut être éventuellement étendue à d'autres émotions ;
- Une composante en NLP/NLU (Natural Language Processing/Natural Language Understanding) qui permet de faire des analyses émotionnelle et sentimentale de ces répliques et de mettre la base de données à jour ;
- Une composante qui recherche automatiquement sur internet des synonymes et antonymes et met à jour un lexicon ;
- Une famille de modèles en nombres entiers qui permettent d'optimiser l'injection d'émotion ;
- Deux boucles rétroactives — une automatique et l'autre impliquant un usager — qui permettent de diriger l'apprentissage.

L'optimisation de l'injection d'émotion est effectuée en remplaçant certains mots de la réplique originale de l'agent conversationnel par des mots qui élicitent l'émotion cible de manière plus prononcée. Les premiers résultats obtenus démontrent la pertinence de notre approche.

ABSTRACT

Chatbots are occupying more and more space in our daily lives. Beginning with technical support chatbots, chatbots that are trying to sell merchandise through social media, general conversation chatbots or even our own little assistants following us everywhere through our phones, we expose ourselves to one form or another of these intelligent interactive bots. But how can we ensure that these chatbots are on par with our emotional expectations and that the replies they generate are not getting out of control?

In this project, we attempt to answer this question by presenting a new and innovative way of combining operations research and machine learning to optimize and control the way a conversational agent conveys emotion. We have created a system, called *Philia*, that can be integrated into any conversational agent to enhance its responses by injecting emotion.

The system we propose is a first implementation of a complex composition of several elements:

- A database that collects the emotional impact of words for a target emotion (joy) but which can be extended to other emotions;
- A NLP/NLU (Natural Language Processing/Natural Language Understanding) component that allows for emotional and sentimental analysis of the agent's responses and update of the database;
- A component that automatically searches the internet for synonyms and antonyms and updates a lexicon;
- A family of mixed integer models to optimize the injection of emotion;
- Two feedback loops - one automatic and one involving a user - that control the learning process.

The optimization of the emotion injection is performed by replacing some words of the original responses of the conversational agent by words that elicit the target emotion in a more pronounced way. The first results obtained demonstrate the relevance of our approach.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xiv
LISTE DES SIGLES ET ABRÉVIATIONS	xv
LISTE DES ANNEXES	xvi
CHAPITRE 1 INTRODUCTION	1
1.1 Introduction du contexte et des éléments de la problématique	1
1.2 Objectifs de recherche	2
1.3 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Les théories de l'émotion	4
2.1.1 La théorie des émotions de James-Lange	4
2.1.2 La théorie des émotions de Cannon-Bard	5
2.1.3 La théorie des émotions de Schachter-Singer	6
2.1.4 La théorie des appraisals d'émotions	7
2.1.5 La théorie de l'expression faciale des émotions	8
2.1.6 Les représentations des émotions	9
2.2 Analyse émotionnelle et sentimentale	10
2.2.1 Analyse émotionnelle et sentimentale par règles	10
2.2.2 Analyse émotionnelle et sentimentale par apprentissage machine classique	16
2.2.3 Analyse émotionnelle et sentimentale par apprentissage machine profond	21
2.2.4 Analyse émotionnelle et sentimentale hybride	25

2.3	Génération de textes émotifs	26
2.3.1	Encodeurs de réponses émotives	27
2.3.2	Décodeurs de réponses émotives	28
2.4	Les émotions dans l'apprentissage par renforcement	30
2.4.1	Les implémentations par apprentissage par renforcement suivant les élicitations émotives homéostatiques et extrinsèques	31
2.4.2	Les implémentations par apprentissage par renforcement suivant les élicitations basées sur la théorie des appraisals des émotions	33
2.4.3	Les implémentations par apprentissage par renforcement suivant des mécanismes de valeurs et récompenses	35
2.4.4	Les implémentations par apprentissage par renforcement émotionnels codées de manière rigide	36
2.5	Métriques d'évaluation automatique	37
2.5.1	Les métriques basées sur la superposition des mots	37
2.5.2	Les métriques basées sur la distance des mots similaires	39
2.5.3	Les métriques libres de références	41
CHAPITRE 3 L'INJECTION D'ÉMOTION : PHILIA		43
3.1	Bloc 1 : Les agents conversationnels	44
3.2	Bloc 2 : Compréhension du langage naturel	45
3.2.1	Décomposition des répliques	45
3.2.2	Analyses sentimentale et émotionnelle	46
3.2.3	Reconstruction de la réplique optimale	47
3.3	Bloc 3 : Vocabulaire	48
3.3.1	Base de données des synonymes et antonymes	48
3.3.2	Base de données des scores	49
3.3.3	Initialisation de la base de données des scores	50
3.4	Bloc 4 : Optimisation	52
3.4.1	Un premier modèle minimaliste	52
3.4.2	Un deuxième modèle plus sophistiqué	54
3.4.3	Le modèle implémenté	55
3.4.4	Détails sur la reconstruction de la réplique optimale	58
3.5	Bloc 5 : Centre de commande	59
3.5.1	Boucle rétroactive automatique	59
3.5.2	Boucle rétroactive usager	66
3.6	Fonctionnement de Philia	68

CHAPITRE 4	MÉTRIQUE D'ÉVALUATION ET RÉSULTATS	72
4.1	Quelques considérations pratiques sur les métriques	72
4.2	Métrique d'évaluation automatique	73
4.2.1	La métrique FED	73
4.2.2	Les 18 caractéristiques comparatives	75
4.2.3	Adaptation de la métrique FED	76
4.3	Évaluation et résultats	77
4.3.1	Premier contexte, le cas $\lambda_1 = 0.25$	81
4.3.2	Deuxième contexte, le cas $\lambda_1 = 0$	86
4.3.3	Impact du correcteur et incohérence sémantique	90
4.3.4	Impact de la liste des synonymes et cohérence sémantique	94
CHAPITRE 5	CONCLUSION ET RECOMMANDATION	96
5.1	Limitations de la solution proposée	97
5.2	Améliorations futures	99
RÉFÉRENCES	101
ANNEXES	119

LISTE DES TABLEAUX

Tableau 4.1	Résultats obtenus pour l'évaluation du système initialisé avec un apport sentimental égal à 0,25. Les différences, en pourcentage O , représentent la différence des scores obtenus par le système par rapport aux scores des conversations originaux.	82
Tableau 4.2	Résultats obtenus pour O_c en forçant la valeur de la variable interne émotionnelle égale à 0, 0,5, et 1. Système initialisé en pondérant l'apport sentimental au score individuel des répliques à 0.	86
Tableau 4.3	Résultats obtenus pour l'évaluation du système initialisé avec un apport sentimental nul mais avec des seuils de reconnaissance Φ_ε différents suivant les phases d'initialisation et de détection. Les différences, en pourcentage, représentent la différence O des scores obtenus par le système par rapport aux scores des conversations originaux. . .	88
Tableau 4.4	Philia en action sur quatre exemples extraits des conversations utilisées pendant son évaluation.	92
Tableau 4.5	Exemples de répliques simples en utilisant tous les synonymes.	94
Tableau 4.6	Les mêmes exemples que dans le tableau 4.5, cette fois-ci en utilisant une liste de synonymes validée manuellement.	95
Tableau A.1	Résultats obtenus en appliquant un facteur pénalisant de 0 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	119
Tableau A.2	Résultats obtenus en appliquant un facteur pénalisant de 3 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	120
Tableau A.3	Résultats obtenus en appliquant un facteur pénalisant de 6 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	121
Tableau A.4	Résultats obtenus en appliquant un facteur pénalisant de 9 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	122
Tableau A.5	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (a)	123
Tableau A.6	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (b)	123

Tableau A.7	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (c)	124
Tableau A.8	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (a)	124
Tableau A.9	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (b)	125
Tableau A.10	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (a)	125
Tableau A.11	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (b)	126
Tableau A.12	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (a)	126
Tableau A.13	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (b)	127
Tableau A.14	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (a)	127
Tableau A.15	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (b)	128
Tableau A.16	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (a)	128
Tableau A.17	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (b)	129
Tableau A.18	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (c)	129
Tableau A.19	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (a)	130
Tableau A.20	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (b)	130
Tableau A.21	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (a)	131
Tableau A.22	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (b)	131
Tableau A.23	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (a)	132
Tableau A.24	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (b)	132

Tableau A.25	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (a)	133
Tableau A.26	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (b)	133
Tableau A.27	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (a)	134
Tableau A.28	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (b)	134
Tableau A.29	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (c)	135
Tableau B.1	Résultats obtenus en appliquant un facteur pénalisant de 0 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	136
Tableau B.2	Résultats obtenus en appliquant un facteur pénalisant de 3 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	137
Tableau B.3	Résultats obtenus en appliquant un facteur pénalisant de 6 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	138
Tableau B.4	Résultats obtenus en appliquant un facteur pénalisant de 9 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.	139
Tableau B.5	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (a)	140
Tableau B.6	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (b)	140
Tableau B.7	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (c)	141
Tableau B.8	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (a)	141
Tableau B.9	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (b)	142
Tableau B.10	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (a)	142
Tableau B.11	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (b)	143

Tableau B.12	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (a)	143
Tableau B.13	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (b)	144
Tableau B.14	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (a)	144
Tableau B.15	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (b)	145
Tableau B.16	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (a)	145
Tableau B.17	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (b)	146
Tableau B.18	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (c)	146
Tableau B.19	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (a)	147
Tableau B.20	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (b)	147
Tableau B.21	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (a)	148
Tableau B.22	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (b)	148
Tableau B.23	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (a)	149
Tableau B.24	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (b)	149
Tableau B.25	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (a)	150
Tableau B.26	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (b)	150
Tableau B.27	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (a)	151
Tableau B.28	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (b)	151
Tableau B.29	Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (c)	152

LISTE DES FIGURES

Figure 3.1	Exemple de décomposition que le système effectue avec spaCy. La phrase donnée en entrée est "This is a test sentence to show the spaCy decomposition we perform for Philia."	46
Figure 3.2	Un exemple de requête graphique d'évaluation émotionnelle envoyée à l'utilisateur pour noter le score émotionnel des tuples sélectionnés dans la réplique reconstruite optimale.	67
Figure 3.3	Un exemple de requête graphique d'évaluation des synonymes et antonymes envoyée à l'utilisateur pour améliorer les synonymes et antonymes de la base de données.	68
Figure 3.4	L'organigramme complet du fonctionnement de Philia, les flèches noires représentent le cycle principal, les flèches bleues représentent la séquence de rétroaction automatique et les flèches rouges la séquence de rétroaction usager.	69
Figure 4.1	Relation entre la valeur de la variable interne émotionnelle et la précision de détection de l'émotion cible moyenne. Système initialisé en pondérant l'apport sentimental au score individuel des répliques à 0,25.	85
Figure 4.2	Relation entre la valeur de la variable interne émotionnelle et la précision de détection de l'émotion cible moyenne. Système initialisé en pondérant l'apport sentimental au score individuel des répliques à 0.	90

LISTE DES SIGLES ET ABRÉVIATIONS

BERT	Bidirectional Encoder Representations from Transformers
BoW	Bag-of-Words
CBOW	Continuous Bag-of-Words
CGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
COMET	COMmonsEnse Transformers
CVAE	Conditional Variational Autoencoder
DB	Database
DF	Document Frequency
FED	Fine-grained Evaluation of Dialog
GRU	Gated Recurrent Unit
IG	Information Gain
KNN	K-Nearest Neighbours
LSA	Latent Semantic Analysis
LSTM	Long Short-Term Memory
MI	Mutual Information
NLP	Natural Language Processing
NN	Neural Network
PMI-IR	Pointwise Mutual Information- Information Retrieval
POS	Part-of-Speech
RNN	Recurrent Neural Network
RoBERTa	Robustly Optimized BERT Pretraining Approach
RUBER	Referenced metric and Unreferenced metric Blended Evaluation Routine
S-O	Sous-Objectif
SDA	Stacked Denoising Auto-encoder
SVM	Support Vector Machines
TF-IDF	Term Frequency – Inverse Document Frequency
USR	UnSupervised and Reference-free evaluation metric

LISTE DES ANNEXES

Annexe A	Résultats détaillés obtenus lors de l'évaluation du système, initialisation faite avec un apport sentimental à 0.25	119
Annexe B	Résultats détaillés obtenus lors de l'évaluation du système, initialisation faite avec un apport sentimental nul	136
Annexe C	Pseudo-code de la boucle principale de Philia	153

CHAPITRE 1 INTRODUCTION

1.1 Introduction du contexte et des éléments de la problématique

Prenons un exemple d'actualité pour justifier l'introduction d'émotion dans les agents conversationnels. Bon nombre de personnes en détresse psychologique, nécessitant l'intervention d'une personne à l'écoute, sont incapables de recevoir l'attention qu'elles requièrent. Ceci est une problématique particulièrement alarmante chez les jeunes de 15-24 ans et les jeunes adultes de 25-44 ans dont 36% et 31% [1] respectivement ont souffert à un moment ou un autre d'une détresse psychologique de niveau dit élevé. La situation est plus catastrophique du côté des étudiants universitaires. Selon un sondage de la Fédération des associations étudiantes du campus de l'Université de Montréal [2], 55% des étudiants universitaires de 15-24 ans souffrent de détresse psychologique. Pour pallier à ce problème, on se tourne de plus en plus vers des solutions en intelligence artificielle, afin d'assister les personnes en détresse, psychologique ou autre, à l'aide d'agents conversationnels. Ces derniers peuvent aider à soulager une personne dans le besoin avant la consultation d'un expert [3]. Un problème avec les agents conversationnels actuels, c'est que les réponses qu'ils génèrent ne sont souvent pas assez émotionnellement chargées ; elles sont courtes, répétitives et dépourvues d'un sens aigu de la conversation [4]. Rendre les agents conversationnels plus émotifs est donc un besoin et, en fait, il s'agit d'un domaine très actif de recherche.

Une première difficulté est qu'il n'existe pas de réel consensus sur ce qu'est une émotion ni même sur la nature de ses apparitions. Au courant du dernier siècle, plusieurs théories de l'émotion ont fait surface, notamment la théorie de James-Lange [5], celle de Cannon-Bard [6] et les théories des « appraisals » d'émotion [7–11]. Chacune de ces théories introduit sa propre manière de comprendre les déclencheurs d'émotions et les effets physiologiques associés. Nous les reprendrons plus en détail la section 2.1.

Il existe essentiellement deux grandes approches en programmation pour simuler de l'émotion. La première approche est basée sur des architectures symboliques et cognitives et donc basée sur les différentes théories mentionnées ci-dessus. La deuxième approche, quant à elle, est basée sur l'adaptation graduelle d'un agent à son environnement en apprentissage par renforcement [12].

Une grande difficulté d'adaptation de ces connaissances théoriques par les approches cognitives et symboliques est le nombre important de règles d'inférence, parfois contradictoires, à intégrer et à faire respecter. Mais la plus grande limitation de cette approche est sans doute

le fait que ce type d'implémentations ne peut véritablement évoluer par lui-même, ce qui limite l'application qu'aux cas considérés.

La deuxième approche, basée sur l'apprentissage par renforcement, tente de dépasser cette dernière limite. Cette approche — issue de l'apprentissage automatique — consiste essentiellement à laisser un agent explorer son environnement immédiat et apprendre par lui-même comment atteindre son objectif émotif ou autre assujetti à cet environnement. Ce type d'implémentation comporte son lot de problèmes également, notamment la difficulté de contrôler un tel agent [13].

Dans le cadre de ce mémoire, nous proposons une nouvelle approche combinant la recherche opérationnelle et l'apprentissage automatique pour améliorer l'émotion d'un agent conversationnel et nous essayons de réconcilier les deux approches — l'approche cognitive et symbolique et l'approche par apprentissage par renforcement — en démontrant qu'elles sont en fait complémentaires.

1.2 Objectifs de recherche

L'objectif central de ce projet est de proposer une première implémentation concrète d'une approche hybride en recherche opérationnelle et apprentissage automatique pour injecter de l'émotion dans un agent conversationnel. De manière plus détaillée, nous pouvons décomposer cet objectif principal en quatre sous-objectifs (S-O) :

S-O1 : Créer ou adapter une métrique permettant de mesurer la consonance/dissonance émotive entre un utilisateur et un agent conversationnel.

S-O2 : Développer une base de données (BD) capable de recenser l'impact des mots sur l'élicitation d'une émotion donnée dans un contexte donné.

S-O3 : Créer ou adapter un algorithme d'analyses émotive et sentimentale.

S-O4 : Intégrer les sous-objectifs ci-dessus et développer un nouvel algorithme d'injection d'émotions cibles dans les réponses d'un agent conversationnel.

Dans le cadre de ce mémoire et comme preuve de concept, nous n'utiliserons qu'une seule émotion cible : la joie. Nous utiliserons des dialogues ouverts. Nous essayerons donc de rendre les réponses ouvertes d'un agent conversationnel plus joyeuses. Le choix de cette émotion cible est purement pratique, car la joie est hautement polarisatrice, donc facilement détectable. Nous sommes conscients que dans une application réelle, l'émotion cible appropriée serait plutôt l'empathie, surtout si notre système doit interagir avec des personnes en détresse.

1.3 Plan du mémoire

Ce mémoire comporte cinq chapitres : l'introduction (chapitre 1), la revue de littérature (chapitre 2), une description détaillée de Philia (chapitre 3), la description d'une métrique pour analyser des dialogues y compris dans leur caractère émotif suivi d'une étude des premiers résultats obtenus avec Philia (chapitre 4) et finalement notre conclusion (chapitre 5).

La revue de littérature (chapitre 2) est un survol de l'état de l'art des différents domaines que nous avons combinés pour construire Philia. Nous commençons par une présentation des principales théories émotionnelles, puis nous continuons par une présentation des différentes techniques utilisées pour classifier émotionnellement et sentimentalement des extraits textuels. Nous enchaînons avec les générateurs de textes émotifs et l'apprentissage par renforcement que nous utilisons aussi dans Philia. Finalement, nous présentons différentes métriques existantes qui permettent d'évaluer automatiquement la qualité d'un extrait textuel généré.

Philia est présenté en détail (chapitre 3), bloc par bloc. Nous consacrons une section à chacun des cinq blocs qui le composent et finalement nous les incluons tous dans une architecture générale.

Le chapitre suivant (chapitre 4) présente une métrique, FED, pour analyser automatiquement des dialogues. Nous la complétons avec des caractéristiques pour mesurer l'injection d'émotion et puis analysons nos premiers résultats issus de cette métrique augmentée.

Finalement, le dernier chapitre (chapitre 5) conclut ce mémoire par un résumé de nos apports ainsi qu'une présentation des limitations actuelles de Philia et de possibles développements futurs.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Les théories de l'émotion

Pour comprendre comment les émotions surviennent chez l'homme, plusieurs théories ont été développées. Elles se décomposent essentiellement en trois catégories : les théories physiologiques, neurologiques et cognitives [14–16]. Les théories physiologiques prétendent que les émotions sont le résultat des réactions physiques ressenties lors d'un stimulus externe. D'un autre côté, les théories neurologiques [6, 17] avancent que les émotions résultent des processus chimiques du cortex. Pour finir, les théories cognitives [7, 18–21] suggèrent que les émotions sont le fruit d'un processus mental de réflexion. Afin de mieux comprendre le déclenchement émotionnel, il est intéressant de se pencher sur les cinq théories les plus reconnues dans le domaine [14, 16, 22], soit la théorie des émotions de James–Lange, de Cannon–Bard, de Schachter–Singer, la théorie des appraisal (évaluations) des émotions et celle de l'expression faciale des émotions. Nous présentons les différentes théories émotionnelles les plus dominantes selon leur ordre d'apparition historique, ce qui permettra de comprendre comment la psychologie émotionnelle a évolué et s'est complexifiée. Nous terminons par un survol des différentes manières de représenter les émotions.

2.1.1 La théorie des émotions de James-Lange

Pour commencer, la première théorie à être introduite est celle des émotions de James-Lange. Cette théorie physiologique est le produit de deux théories développées indépendamment à la fin des années 1800. D'abord introduite en 1884 par James [5], sa théorie dit que les émotions sont le résultat d'une réaction physiologique face à un stimulus et que sans ceux-ci, une émotion ne pourrait être ressentie. Selon James, la réaction physiologique est l'émotion en soi, donc les deux sont indissociables. En 1885, indépendamment de James, Lange développe des idées similaires à James, mais contrairement à lui, Lange avance que l'émotion est le résultat d'une réflexion mentale sur les changements physiologiques observés et non uniquement une réaction face à ces changements [6]. Les deux auteurs sont d'accord sur le fait que l'émotion ne peut exister sans réaction physique du corps [5, 23–25]. Avec le temps, ces deux théories ont été reconnues comme étant une seule théorie, que l'on connaît aujourd'hui sous le nom de la théorie des émotions de James-Lange. La ligne finale que cette théorie emprunte dit que les changements physiologiques sont dépendants du stimulus externe et que les émotions sont dépendantes de l'analyse effectuée à l'égard de ces réactions. À titre d'exemple, si un

individu fait face à un danger et que le corps se met à trembler, l'individu ressentira de la peur, car il identifie le fait de trembler à de la peur. L'individu ne tremble pas parce qu'il a peur. Plusieurs arguments soutiennent cette théorie, comme la suppression d'une expression physique peut réduire l'intensité de l'émotion ressentie. Par exemple, lorsqu'on essaie de faire rire quelqu'un qui pleure pour le rendre heureux. Même un changement physique intentionnel peut augmenter ou diminuer l'intensité d'une émotion. Par exemple, un sourire, même forcé peut améliorer notre humeur [26]. Une des limites de cette théorie est la possibilité d'avoir une expression émotive sans une réaction physiologique. Il est même possible de dire que plus d'une émotion peut être associée à une réaction physique. Par exemple, avoir un corps qui tremble et un rythme cardiaque accéléré peut être associé à une expérience terrifiante, comme à une expérience excitante ou un défi à relever sans nécessairement déclencher de la peur chez un individu. Cette théorie ne prend pas en compte les croyances et expériences des individus, ce qui, in fine, peut contribuer à l'émotion ressentie et son intensité [16].

2.1.2 La théorie des émotions de Cannon-Bard

Suite à l'introduction de la théorie des émotions de James-Lange et sa critique, plusieurs autres théories ont été avancées dont celle de Cannon-Bard [6]. Cannon a voulu tester la théorie émotionnelle prédominante à l'époque en examinant s'il était possible de stimuler une émotion sans réaction physiologique de l'hôte, en coupant le lien viscéral de l'individu, c'est-à-dire le système nerveux instinctif. Pour confirmer sa théorie, Cannon a entrepris de couper ce lien nerveux chez plusieurs chats lors de ses expérimentations et a compilé ses résultats en 1915 [27]. Suite à ces expériences, Cannon s'est rendu compte que les réponses émotives de ces chats ne changeaient pas face à un stimulus comparativement aux chats ayant un système viscéral intact [6, 27]. Cette découverte allait à l'encontre de ce qu'avancait la théorie de James-Lange. De plus, Cannon fait aussi la remarque que différentes réactions physiologiques telles que la perspiration, les battements de cœurs plus rapides, la dilatation des pupilles et autres sont des réactions physiologiques trop répandues chez l'homme pour permettre de distinguer les émotions de cette manière. Voulant pousser ses découvertes, Cannon s'associa avec Britton [28] afin de comprendre comment les émotions pouvaient être générées. En continuant leurs expériences sur des chats, ils ont constaté qu'en enlevant la partie postérieure du thalamus, les expressions affectives des chats étaient inhibées, indiquant que le thalamus jouait un rôle primordial dans l'organisation neuronale émotionnelle de l'individu. Suivant ses découvertes, s'associant cette fois-ci avec son étudiant de l'époque Philip Bard, Cannon mit sur pied sa propre théorie des émotions, celle que l'on connaît sous le nom de Cannon-Bard [27]. La théorie des émotions de Cannon-Bard prétend que l'état émotionnel d'un individu se manifeste simultanément, et indépendamment, de sa réaction physiologique

et non séquentiellement à celle-ci. Selon cette théorie, les émotions sont générées principalement dans le thalamus, une zone cérébrale qui contient les récepteurs neuronaux nécessaires pour la génération d'émotions [6, 24, 27]. En reprenant l'exemple donné plus haut, si un individu fait face à un danger, le corps se met à trembler au même moment où l'individu ressentirait de la peur par l'entremise de l'activation des centres subcorticaux du cerveau. Un avantage de cette théorie par rapport à celle de James-Lange est qu'elle considère plusieurs états émotionnels pour les mêmes réactions physiologiques. Cette manière de faire a ses avantages, mais aussi ses inconvénients. La théorie de Cannon-Bard a fait l'objet de plusieurs critiques. Un des reproches est le fait que cette théorie donne trop d'importance au thalamus quant à son rôle dans la génération d'émotions tout en omettant l'implication d'autres zones du cerveau. L'importance du thalamus est justifiée, mais le processus de génération d'émotions est beaucoup plus complexe que ce que suggère cette théorie selon Roxo et coll. [29]. De plus, cette théorie suggère que les réactions physiologiques n'ont pas d'influence sur les émotions qui sont ressenties indépendamment, ce qui est faux selon Laird et Lacasse [26].

2.1.3 La théorie des émotions de Schachter-Singer

La théorie des émotions de Schachter-Singer, introduite en 1962, est aussi connue sous le nom de théorie à deux facteurs [18]. Schachter et Singer voulaient explorer comment la connaissance de l'état physiologique qu'un individu a sur lui-même pouvait contribuer à l'émotion déclenchée chez celui-ci. À travers diverses expériences conçues pour tromper les participants, Schachter et Singer ont voulu montrer que les connaissances situationnelles des individus pouvaient influencer la manière dont les émotions sont générées. Ils ont pu montrer que pour un même stimulus, si les connaissances à priori des participants sur la nature du stimulus étaient différentes, l'émotion générée différait d'un individu à un autre [18]. Selon cette théorie, les émotions sont le fruit d'un travail interactif entre les réactions physiologiques et cognitives face à un stimulus. En des termes plus exacts, face à un stimulus, une réaction physiologique autonome se produit qui est interprétée par l'individu selon le contexte auquel il fait face. Cette réflexion cognitive cause ultimement une expérience émotionnelle chez l'individu et renforce son identification dans ce contexte [24]. Il y a une ressemblance entre cette théorie et celle de James-Lange puisque les deux suggèrent que les émotions sont une réaction face aux changements physiologiques. La différence ici est que différentes émotions peuvent être ressenties pour les mêmes réactions physiologiques d'un individu, car il tient compte du contexte immédiat. Plusieurs critiques ont été avancées au courant des années par des chercheurs qui ont essayé de répliquer les expériences de Schachter et Singer. Marshall et Zimbardo en 1979 ne sont pas arrivés aux mêmes conclusions face aux réactions euphoriques et colériques des participants de l'étude originale. LeDoux [30] a aussi critiqué cette théorie

en 1995 en remarquant que la théorie de Schachter et Singer se concentre beaucoup trop sur l'importance du système autonome de l'individu face à l'émotion ressentie sans tenir suffisamment compte le rôle central du cortex lors de la génération d'émotions.

2.1.4 La théorie des appraisals d'émotions

Introduite dans les années 1950 par la psychologue Magda Arnold [7], elle suggère que la séquence menant à une élicitation émotionnelle commence par une évaluation primaire d'un stimulus externe qui stimule simultanément les zones physiologiques appropriées et l'expérience émotionnelle en soi. En 1991, Richard Lazarus développe la théorie cognitive-motivationnelle-relationnelle des émotions [8], bâtie à partir des idées avancées par Magda Arnold. Selon Lazarus, les émotions sont déterminées par l'analyse psychologique du stimulus externe. Cette évaluation se fait parallèlement et simultanément à une réponse, souvent subconsciente, de la situation peut déterminer la nature finale de l'émotion ressentie [8]. Lazarus distingue aussi entre deux types d'analyses (appraisals) du stimulus, l'analyse primaire qui cherche à donner un sens à un événement, c'est-à-dire d'en déterminer la dangerosité, et une analyse secondaire qui visualise la capacité d'un individu à surmonter un événement, donc de déterminer si l'individu se trouve en grave danger ou si l'événement est facilement surmontable. Lazarus met donc l'emphase sur le rôle du centre cognitif pour développer l'émotion face à un stimulus en analysant celui-ci face à l'historique de l'individu associé à sa capacité de surmonter ledit stimulus. Les différentes émotions pouvant être élicitées sont donc directement en lien avec ces deux analyses cognitives de l'individu face à un stimulus. Si les conditions entourant le stimulus changent, l'émotion ressentie changera aussi. Plusieurs itérations ont depuis fait leur apparition. Citons, par exemple, les raffinements apportés par Roseman [9,10]. Il propose de nouvelles idées quant au rôle d'un événement causant l'émotion. Il avance que c'est l'interprétation d'un événement et non l'événement en soi qui cause l'émotion. Pour soutenir ses idées, Roseman incorpore de nouvelles dimensionnalités à considérer, soit celles de la « consistance motivationnelle » et de la « responsabilité » de l'individu. Scherer raffine à son tour la théorie des appraisals en 2001 en introduisant un modèle séquentiel multicouche [11]. Les trois couches introduites sont la couche innée (codée génétiquement), la couche apprise (historique de l'individu) et la couche délibérative (processus cognitif). Lorsqu'un stimulus survient, l'individu voit séquentiellement à travers plusieurs points de contrôle situés dans chaque couche comment réagir face à ce stimulus, puis l'émotion est générée en conséquence. Ce qu'il faut retenir c'est qu'un individu peut ressentir plusieurs émotions et réactions physiologiques pour un même stimulus, mais que c'est le contexte, le temps, et l'interprétation (consciente ou inconsciente) que l'individu en fait, qui dictent la genèse des émotions et des réactions physiologiques ressenties. Un avantage de cette théorie est de considérer que

c'est l'analyse événementielle courante d'un individu qui dicte l'émotion ressentie et non l'évènement en soi, ce qui permet de tenir compte de la variabilité des émotions ressenties pour un même évènement [31]. Un autre aspect intéressant de cette théorie est que pour les mêmes patrons analytiques, un individu ressentira la même émotion, ce qui rend cette théorie hautement mathématiquement représentable. Plusieurs modèles en analyse sentimentale et émotionnelle sont basés sur la théorie des appraisals [32–35]. Ces modèles tirent avantage de l'architecture hiérarchique avancée par les différentes théories des appraisals pour identifier les émotions dans différents extraits textuels. D'autres auteurs comme Rajendran et coll. [36] ou Jin et coll. [37] utilisent même de l'apprentissage profond pour détecter automatiquement les différents indices syntaxiques qui permettent de comprendre la hiérarchie de la causalité émotive dans différents extraits textuels. En apprentissage par renforcement on fait même une analogie directe avec les mécanismes des différentes théories des appraisals pour faire du biomimétisme et ainsi rendre les différents agents capables de changer dynamiquement leurs buts immédiats en changeant la fonction de valeur selon les différentes variables dimensionnelles tirées de ces théories psychologiques [38–42]. Différentes théories des appraisals sont toujours en vogue et suscitent beaucoup d'intérêt à cause de leur structure hiérarchique qui facilite leur intégration dans des architectures computationnelles.

2.1.5 La théorie de l'expression faciale des émotions

Cette théorie plus récente des émotions n'explique pas entièrement la génération des émotions, mais plutôt comment les expressions faciales peuvent influencer l'intensité des émotions que nous ressentons [43]. En effet, des expériences effectuées par Havas et coll. [44] ont démontré que lorsque des individus dépressifs ont vu leurs muscles abaisseurs sourciliers paralysés à l'aide d'injections de Botox, l'intensité de leurs symptômes dépressifs s'est vue diminuer. En 1988, Strack et coll. [45] ont mis sur pied une expérience pour démontrer que les émotions ne sont pas uniquement élicitées par des réactions physiologiques et des processus cognitifs ou neurologiques, mais qu'elles peuvent aussi être influencées, voir même générées par des contractions des muscles faciaux. Comprendre comment les expressions faciales jouent un rôle sur l'expression ressentie permet d'identifier les émotions automatiquement chez les humains. Certains chercheurs se sont penchés sur la question en intégrant des modèles multimodaux capables de détecter les émotions chez les humains pour des agents conversationnels plus réactifs [46, 47].

2.1.6 Les représentations des émotions

Comprendre les différentes théories psychologiques et neurologiques qui essaient d'expliquer l'élicitation émotionnelle chez les humains peut nous aider grandement à les adapter dans nos modèles computationnels afin de leur donner un avantage qui pourrait être qualifié de biologique. Mais comment représente-t-on ces émotions? Plusieurs modèles de représentation des émotions existent et cette section va faire un survol des différentes implémentations populaires. On identifie deux familles de modèles de représentations émotive, les modèles catégoriques [5,48–50] qui font un découpage distinctif entre les différentes émotions qu'ils présentent, et les modèles dimensionnels [17,51–54] qui ressemblent plus à plusieurs ensembles de représentations émotive continues sur des plans, hyperplans ou vecteurs qui s'entrecoupent pour décrire des émotions souvent plus complexes. Les représentations catégoriques émotionnelles sont assez simples à comprendre, c'est l'identification distinguée entre les émotions fortes que nous ressentons. Un des modèles les plus simples est celui introduit par James [5] qui identifie quatre émotions de base : la peur, le deuil, l'amour et la colère, toutes basées sur les réactions physiologiques du corps. Un des modèles les plus répandus et utilisés, de par sa simplicité, est celui de Paul Ekman [48]. Il identifie six émotions de base : la colère, le dégoût, la peur, la joie, la tristesse et la surprise. Ekman fait aussi le lien entre certaines expressions faciales et des émotions [55]. Un des modèles les plus récents de catégorisation émotive est celui de Cowen et Keltner [50]. Les auteurs identifient 27 différentes émotions tirées de 2185 courts-métrages qu'ils représentent visuellement par une carte émotive. À notre connaissance c'est le travail le plus important à ce jour pour essayer d'identifier les plus d'émotions catégoriques possibles. Cet ouvrage a été employé dans [56], un modèle en apprentissage profond qui fait de la reconnaissance émotionnelle textuellement. Pour les modèles dimensionnels émotifs, les représentations sont beaucoup plus vastes. Pour commencer, on peut parler du modèle de Russell [52] qui représente les émotions sur un plan bidimensionnel découpé par deux axes, l'excitation émotive et la valence. Dépendamment d'où on se trouve, une émotion non distincte peut être identifiée. Plutchik introduit en 2001 [54] un modèle tridimensionnel complexe pour représenter les émotions. Il fait une hybridation de la représentation catégorique et dimensionnelle des émotions. On peut imaginer ce modèle comme étant une sorte de toupie dans laquelle, lorsqu'on la regarde du haut, on peut voir que les émotions de base sont identifiées au milieu du cercle et les émotions complexes similaires plus à l'extrémité. Plus on se rapproche de la base de la toupie, plus les émotions identifiées diminuent en intensité et changent d'appellation. Pour la modélisation, les représentations catégoriques prévalent en majorité, car elles sont beaucoup plus simples à implémenter, mais il existe aussi des implémentations qui tirent profit des représentations dimensionnelles des émotions [57,58].

2.2 Analyse émotionnelle et sentimentale

Les analyses sentimentale et émotionnelle¹ jouent un rôle crucial dans ce projet, car elles nous permettront de détecter si l'utilisateur et le système sont en consonance ou dissonance émotive en plus de reconnaître l'émotion transmise à travers les répliques. Ces différentes analyses font partie du champ d'étude des méthodes de traitement automatique du langage naturel, en anglais, « Natural Language Processing » (NLP). L'analyse sentimentale vise à déterminer si le sentiment perçu textuellement est négatif, positif ou neutre [59,60]. L'analyse émotionnelle quant à elle est beaucoup plus vaste et complexe, elle vise à comprendre quelle émotion un extrait textuel provoque chez une personne, et même à comprendre comment cette émotion est provoquée. Dans le contexte de ce travail, c'est plutôt la catégorisation émotionnelle qui nous intéresse, catégorisation qui se base sur les différentes théories existantes [5, 8, 48, 54, 61]. Pour ces deux grands domaines du NLP, il existe différents groupes de modèles similaires dans lesquels nous pouvons les regrouper. Ces modèles se catégorisent en quatre familles [62]. Tout d'abord, il y a les modèles à base de règles : [32, 63–66] pour l'analyse sentimentale, [67–70] pour l'émotionnelle. Ensuite nous retrouvons les modèles basés sur l'apprentissage machine classique : [71–74] et [75, 76]. Par la suite, on retrouve les modèles issus de l'apprentissage profond : [71, 77, 78], [78–80]. Finalement, on identifie la famille des modèles dits hybrides/mixtes [81–84] pour les modèles en analyse sentimentale et [85–88] pour l'analyse émotionnelle.

2.2.1 Analyse émotionnelle et sentimentale par règles

Les modèles à base de règles font partie d'une famille de techniques qui se servent des connaissances générales que nous avons sur la manière dont certains mots ou contextes peuvent susciter un sentiment ou une émotion chez nous. Ces techniques peuvent être à base de détection de mots-clés à forte connotation émotive/sentimentale [70,89–91], par affinité lexicale des mots [92] ou bien par heuristiques [35, 64, 93–98].

Analyse émotionnelle et sentimentale par détection de mots-clés et affinité lexicale

Les modèles à base de détection de mots-clés sont parmi les techniques les plus naïves [62] pour détecter l'émotion ou le sentiment d'un extrait, car il s'agit d'une méthode qui compte l'occurrence des mots à caractère sentimental pour classifier le texte [96]. Plusieurs travaux

1. Nous traduisons par *analyses émotionnelle et sentimentale* les termes bien connus *emotional analysis* et *sentiment analysis* du domaine de l'apprentissage machine.

exploitent cette méthode. Regardons d'abord ce qui s'est fait au niveau de l'analyse émotionnelle. Tao [69] sépare les mots-clés pouvant imbiber une phrase d'émotion en trois catégories différentes, les mots-clés à caractère émotif (aimer, haïr, adorer, etc.), les mots modificateurs, tels que les adverbes, et les mots métaphoriques pouvant aider à identifier l'intention émotive d'un texte. L'auteur a ensuite développé des modules d'extraction basés sur les contextes textuels afin de leur permettre de catégoriser émotivement un extrait. Ma et coll. [70] se sont basés sur le système « Empathy Buddy » [99]. Ils utilisent un système de détection de mots-clés émotifs combinés à une analyse syntaxique pour déterminer la valeur émotive d'une phrase pour ensuite faire une animation 2D d'un agent qui représente l'émotion détectée à l'aide de gestes et de voix synthétiques. C'est une belle manière de combiner la détection textuelle émotionnelle avec la représentation d'émotions en « personne ». Shivhare et coll. [89] ont développé une technique de détection émotive textuelle en détectant les mots-clés à plusieurs niveaux pour ensuite catégoriser l'extrait dans l'une des six émotions d'Ekman [48]. En catégorisant le texte à plusieurs niveaux émotifs, les auteurs peuvent effectuer une comparaison directe entre les différentes émotions détectées et ainsi choisir de catégoriser l'extrait selon l'émotion ayant reçu le score le plus élevé. Pour compléter la section sur les méthodes à base de détection de mots-clés, regardons aussi ce qui s'est fait au niveau de l'analyse sentimentale ces dernières années. Le raisonneur affectif développé par Elliott [90] détecte 198 mots à connotation affective en plus de chercher des modificateurs lexicaux, des adverbes et autres mots pouvant intensifier la charge émotionnelle des mots recherchés, permettant ainsi de détecter la polarité sentimentale d'une phrase. Nasukawa et coll. [100] détectent les sentiments perçus à l'endroit d'un sujet donné identifié dans un texte. Ils font une analyse des mots entourant le sujet et déterminent, grâce à un lexique construit manuellement, la polarité sentimentale perçue.

Ces types de modèles pour effectuer de l'analyse émotionnelle/sentimentale sont relativement faciles à implémenter, ce qui les rend accessibles et économiques. Par contre, ce genre d'approche a son lot de défauts en commençant par une mauvaise reconnaissance émotive lorsque des attributs de négation sont utilisés. Une phrase comme « Il fait beau aujourd'hui » va effectivement être qualifiée de positive par ce genre d'approche, mais une phrase comme « J'aimerais qu'il fasse beau aujourd'hui » risque d'être aussi perçue comme positive malgré qu'elle ne le soit pas nécessairement [62]. Une deuxième faiblesse de ce type de modèles est qu'ils ne détectent pas l'intention approfondie d'une phrase, et en l'absence de mots-clés à identifier, sont incapables de déterminer le sentiment d'un extrait [62]. Pour essayer de mitiger un peu les défauts de la méthode par détection de mots-clés émotifs, il existe une variation de cette technique qu'on qualifie de modélisation par affinité lexicale. Similairement aux modèles utilisant la détection de mots-clés pour déterminer l'émotion perçue textuelle-

ment, les modèles qui utilisent l’affinité lexicale détectent eux aussi les mots-clés chargés émotionnellement. L’avantage qu’ils exploitent est la détection de certains mots qui ne sont pas directement chargés émotionnellement, mais qui peuvent laisser paraître un certain sentiment lorsqu’ils sont employés [62]. Pour mieux comprendre le concept, considérons ces deux exemples : « j’ai découvert par *accident* une nouvelle compagnie en bourse qui va me faire faire beaucoup d’argent » et « j’ai survécu à un *accident* ce matin. » Dans la première phrase, le mot *accident* est mis dans un contexte positif, il décrit une surprise qui est bien vue. Dans le deuxième exemple, le mot *accident* décrit plutôt une négation, car il est en lien avec un possible drame. On voit donc bien que cette méthode détecte aussi, en plus de mots-clés à connotation émotive claire, des mots-clés à connotation émotive/sentimentale qui ont une certaine probabilité de faire basculer le sentiment perçu par l’extrait. Somasundaran et coll. [92] attaquent la classification sentimentale sous cet angle ; ils essaient de détecter les sentiments dégagés par la relation d’opinions et de leurs sujets référés dans le texte. Pour donner un exemple, prenons l’extrait suivant tiré de l’article : « Et j’ai pensé à *une forme plus organique, sans être trop pointue* et sans *ressembler à une boîte*, mais **plutôt** à une sorte *d’appareil portable, pas aussi informatique*. Des designs simples, comme le dernier qu’on vient de voir, **pas trop** de *boutons*. » On peut voir que les **opinions** décrivent directement les *sujets* et procurent une idée de ce à quoi s’attend l’auteur et le sentiment qu’il a face au sujet, en exemple **pas trop** qui décrit une négation face au sujet *informatique*. Les auteurs utilisent la base de données « The AMI meeting corpus » [101] qui décrit de manière syntaxiquement riche plusieurs interactions interusagers. Les auteurs se servent de toutes les opinions détectées sur les différents sujets présents pour déterminer la polarité de l’extrait en les classifiant dans différents *cadres d’opinions*, c’est-à-dire savoir si les opinions sont en contradiction positives/négatives avec leurs sujets ou si elles sont similaires positivement/négativement. Ils ont découvert que certains *cadres d’opinions* peuvent se renforcer ou non ; s’ils se renforcent, leurs relations face au sujet peuvent faire un effet de levier face au sentiment perçu. Le problème avec les méthodes par affinité lexicale est qu’elles aussi ont du mal à capturer la négation dépendamment des contextes. De plus, les probabilités associées aux mots indicateurs d’émotions peuvent ne pas se transférer aisément d’un domaine à un autre [62].

Analyse émotionnelle et sentimentale par heuristiques L’analyse émotionnelle et sentimentale par heuristiques consiste à utiliser un amalgame de connaissances sur la nature émotionnelle des mots, les structures syntaxiques et même les connaissances générales que nous avons sur la manière dont certaines situations peuvent faire ressortir certaines émotions/sentiments chez l’humain. À l’aide de ces connaissances, on construit soigneusement

des règles capables de catégoriser émotionnellement des extraits textuels. On utilise donc la connaissance générale des mots dans leurs contextes pour créer des règles d'inférence, chose que Liu et coll. [96] exploitent bien. Les auteurs utilisent la connaissance générale du contenu sémantique d'un extrait [102–104] pour en évaluer l'émotion sans nécessairement se concentrer sur la détection de mots-clés à caractère émotif. Par exemple, ils peuvent catégoriser émotionnellement l'extrait « je me suis fait virer », car on sait que ce genre de situation peut générer du stress ; donc le sujet ressent probablement de la colère, de la peur et de la tristesse [105]. Iliadis et coll. [93] utilisent une approche plus classique : ils détectent les mots-clés affectifs, puis détectent les mots-clés qui peuvent faire pencher l'extrait dans une des six émotions de base de Ekman [48]. Les auteurs comprennent bien que l'ajout d'une composante sensible à l'affinité lexicale des mots compense le manque de mots-clés chargés émotivement. Les phrases sont décomposées à l'aide du parseur de Stanford [106] et utilisent la base de données « WordNet Affect » [107] pour détecter les mots capables de transmettre une ou des émotions et ensuite classifier l'extrait selon un graphe de dépendance émotionnel. Udochukwu et He [94] ont développé une approche basée sur le modèle OCC de Ortony et coll. [108] qui décrit une structure de déclenchement des émotions et les variables gouvernant leur intensité. La hiérarchie émotionnelle du modèle OCC se sépare en trois branches distinctes : les *conséquences évènementielles* représentent les émotions qui dégagent du plaisir ou le mécontentement face à un évènement, les *actions d'agents* représentent les émotions en relation avec l'agent, qu'il soit personnel ou autre, puis il y a les *aspects des objets* qui représentent les émotions face à cette catégorie. Le modèle d'Udochukwu et He n'exploite que les deux premiers groupes d'émotions. En assignant des valeurs à une liste de variables pouvant aider à inférer une émotion appartenant à ces deux groupes, les auteurs ont fabriqué une série de règles logiques qui, en détectant les mots émotifs et polarisateurs de [59, 109, 110], sont capables de détecter les émotions appartenant soit au groupe d'émotions évènementielles ou d'agents. Un exemple de règle tiré de l'article donnera une meilleure idée du type de règles fabriquées : « **If** Direction = « Self » **and** Tense = « Future » **and** Overall Polarity = « Positive » **and** Event Polarity = « Positive , » **then** Emotion = « Hope » [94]. Lee et coll. [97] présentent un algorithme à base de règles capable de détecter la causalité émotive dans un texte. Comme c'est un travail avant-gardiste selon les auteurs, ils n'essaient pas de détecter les émotions implicites présentes dans le texte, mais seulement de détecter les causes à effets dits explicites. C'est-à-dire que l'inférence émotive peut se faire, car il y a une détection directe de mots transmettant l'émotion. Par exemple, la phrase « ils aiment cela » est la cause directe de l'élicitation de l'émotion de la joie. Les indices syntaxiques que Lee et coll. essaient de détecter sont des verbes de causalité, des verbes d'action, des marqueurs épistémiques, des conjonctions, les prépositions, ainsi que la position des causalités émotives par rapport

au sujet/expérimentateur de celles-ci. À l'aide de ces indices syntaxiques, les auteurs ont développé une série de règles capables de détecter les causalités émotives et inférer l'émotion convoquée. Ils ont découvert, en étudiant 6058 instances émotives du corpus de Sinica [111], que la position et la manière dont la cause est emmenée dans le texte (la position : avant/après le sujet et si la cause a été avancée par un verbe ou de manière nominale) influence énormément l'émotion détectée. Les règles prennent donc avantage de ces connaissances et selon la position de la cause, du sujet, des mots-clés affectifs détectés, ils arrivent à inférer l'émotion et ses déclencheurs. Li et Xu [67] classifient des microblogues émotionnellement en analysant des catalyseurs émotionnels à l'intérieur du texte. En étudiant les blogues présents sur le site de Weibo, le plus important site de « microblogging » en Chine, et se basant sur la théorie de l'émotion narrative de Kleres [112], ils développent une base de données de mots affectueux propres aux microblogues et collectent une série d'indices syntaxiques leur permettant de détecter les déclencheurs émotifs présents dans le texte. Leurs travaux se basent directement sur ce que les auteurs de [97] ont développé, mais adapté à leurs contextes moins formels. Le modèle se base sur la détection de groupes « nom commun, verbe, nom commun », sans toutefois être complètement rigide vu la nature informelle et souvent incomplète des microblogues, et essaie de trouver comment les indices syntaxiques ou mots affectueux sont reliés au groupe pour en faire une compilation vectorielle qu'ils utilisent pour entraîner un classificateur SVM.

Du côté de l'analyse sentimentale, les travaux effectués s'enlignent dans la même direction, et certains exploitent même des concepts de la théorie des appraisals des émotions [8,21,113]. Whitelaw et coll. [35] exploitent aussi cette idée. Étant conscients de la limitation de simplement regarder la fréquence d'apparition de mots à connotation sentimentale pour catégoriser la polarité du texte, ils implémentent un système capable d'analyser la sémantique d'un extrait textuel grâce à des indicateurs taxonomiques tirés d'un lexique créé par les auteurs en comptant 1329 adjectifs et modificateurs syntaxiques. Ces indicateurs taxonomiques sont effectivement des groupes de mots associés à des groupes d'appraisals de Martin [113] (attitude, gradation, orientation, polarité) qui peuvent influencer le sentiment perçu textuellement. Par exemple, « Très ennuyant » est un groupe de mots qui porteraient une attitude plutôt négative. En plus de détecter ces indicateurs textuels, les auteurs détectent aussi des adjectifs (beau, ennuyant, beaucoup, peu, etc.) qui viennent donner un degré d'importance au sentiment perçu afin de mettre à l'échelle la polarité détectée. Le système consiste alors à encoder ces indicateurs, déjà évalués heuristiquement sur leurs sentiments dégagés, et à encoder des mots-clés polarisateurs à l'aide d'un « Bag-of-Words » [114] classique pour entraîner un SVM [115] capable de classifier la polarité de l'extrait. Benamara et coll. [95] utilisent une approche similaire, ils étudient l'influence d'adverbes et d'adjectifs, ou de leurs combinai-

sons, telles que *très mauvais*, sur le sentiment dégagé d'un texte. L'équipe de Benamara définit un ensemble d'axiomes généraux basés sur la classification des adverbes en cinq catégories différentes : les adverbes d'affirmation, de doute, fortement intensificateurs, faiblement intensificateurs et finalement de négation/minimiseurs. Ces axiomes régissent comment les scores sentimentaux des groupes d'adjectifs-adverbes sont bornés et déterminés. Par exemple, un des axiomes dit que « chaque adverbe faiblement intensificateur et chaque adverbe de doute a un score qui est inférieur ou égal au score de chaque adverbe fortement intensificateur » [95]. À l'aide de ces axiomes, les auteurs ont développé trois algorithmes différents (score variable, score à priorité adjectivale, score à priorité adverbiale) pouvant scorer un extrait textuel entre -1 et +1, et ainsi le catégoriser en tant que texte à sentiment négatif, positif ou neutre. Kontopoulos et coll. [116] présentent un modèle d'inférence sentimentale basée sur l'ontologie autour des microblogues sur Twitter. L'ontologie est une manière de représenter les concepts, et leurs interactions, autour d'un domaine, tel que les microblogues de Twitter, afin de les rendre lisibles par des machines, que ce soit sous forme de graphe ou autre [117]. Les auteurs ont alors eux-mêmes développé leur propre base ontologique dans un premier temps et, dans un deuxième temps, ils effectuent de l'analyse sentimentale grâce à ces connaissances. La base ontologique utilisée est une table binaire ayant comme lignes les objets d'intérêt et comme colonnes les différents qualificatifs pouvant être associés aux termes de gauche, puis on attribue un « + » si une relation existe entre le terme et le qualificatif. De cette manière les auteurs peuvent dresser un graphe relationnel permettant de représenter les relations sémantiques entre plusieurs indices syntaxiques ou des mots d'intérêt. Ils identifient alors les microblogues d'intérêt et effectuent l'analyse sentimentale sur eux à l'aide du service OpenDover². Un des articles les plus importants sur l'analyse sentimentale par heuristiques provient de Hutto et Gilbert [63], les auteurs de Vader. Vader est un modèle à base de règles qui utilise une combinaison de méthodes qualitatives et quantitatives pour effectuer de l'analyse sentimentale. Premièrement, à l'aide d'un lexique construit pour représenter les mots d'intérêt polarisateurs dans un contexte de « microblogging » et leurs intensités, puis combiné à un ensemble de cinq règles d'analyse syntaxique et grammaticale pour produire un score de polarité pour chaque extrait textuel. En incorporant des indices syntaxiques autres que les mots polarisateurs, tels que les points d'exclamation, les mots écrits tout en majuscules, les mots intensificateurs comme les adverbes, les conjonctions et les groupes de mots qui précèdent l'introduction des mots à caractère sentimental, les auteurs peuvent identifier les degrés d'intensité réelle des sentiments perçus ainsi que fortement diminuer les cas où la négation n'est pas détectée, ce qui mène à une mauvaise classification sentimentale.

Les modèles issus de la technique par heuristique ajoutent une surcouche, en général, aux

2. OpenDover online sentiment analysis service : <https://opendover.nl>

modèles qui se servent principalement de la détection de mots-clés émotifs pour en déterminer l'émotion ou le sentiment. En exploitant les différentes affinités syntaxiques entre les mots dans une phrase, ces techniques arrivent à détecter de l'émotion/sentiment où il n'y a aucune trace de mots émotifs, ce qui règle une des faiblesses des modèles usant uniquement de la détection de mots-clés. Une des faiblesses de ce genre de modèles, c'est qu'ils peuvent être coûteux computationnellement [62]. Une autre faiblesse est que ce genre d'approche est fortement dépendante du contexte d'utilisation du modèle, donc difficilement généralisable à d'autres domaines sans devoir faire des efforts substantiels pour adapter comment les mots à affinité lexicale interagissent sur la manière de percevoir une émotion/sentiment [62]. Ceci qui rend leur implémentation pour d'autres domaines coûteuse, ou même voire impossible, si les connaissances appropriées sur l'interaction syntaxique entre les différentes composantes d'une phrase est mal comprise pour le domaine d'application.

2.2.2 Analyse émotionnelle et sentimentale par apprentissage machine classique

L'analyse émotionnelle et sentimentale par apprentissage machine classique consiste généralement en la construction automatique d'une base de données de représentations vectorielles des mots d'intérêt émotif puis d'implémenter un classificateur afin d'inférer l'émotion ou le sentiment dans un extrait textuel nouveau. Parmi les techniques de vectorisation existantes, les plus populaires sont le « Bag-of-Words » [114], le « TF-IDF » [118], le « Word2Vec » [119] ou bien le « Glove » [120]. Une fois l'extraction des caractéristiques et l'encodage vectoriel effectués, il sera alors possible d'entraîner un modèle classique de type arbre de décision [75], faire de la régression linéaire [73, 121, 122] ou polynomiale, utiliser un SVM [75, 76, 123–130], des réseaux bayésiens [74–76, 91, 122, 124, 125, 127, 128, 130–133], des KNN [75, 125, 133, 134], LSA [91].

Pour commencer, nous allons couvrir les articles ayant eu le plus d'influence en analyse émotionnelle en commençant par l'article de Alm et Roth [121]. Ces auteurs explorent la possibilité de prédire l'émotion présente textuellement en utilisant des techniques d'apprentissage machine supervisée utilisant un modèle de classification multi-classes suivant l'architecture d'apprentissage SNoW de Carlson [135], un modèle de classification linéaire. Les auteurs ont annoté 185 histoires pour enfants (1580 phrases) et leur ont attribué une des six émotions de base d'Ekman [48] et la neutralité. Les variables utilisées pendant l'entraînement prenaient en compte non uniquement les mots émotifs, mais aussi le nombre de verbes présents, la position de la phrase dans l'histoire, les mots écrits en majuscule, puis les points d'exclamation pour n'en nommer que quelques-uns. Au total, trente différentes variables ont été utilisées pour l'entraînement. Les auteurs se sont rendu compte qu'à cause du nombre limité de données,

le choix des hyper paramètres faisait varier les résultats énormément dû à la « malédiction de la dimensionnalité », c'est-à-dire que le nombre de variables était trop grand pour le nombre d'entrées comptabilisées. Ceci a été mitigé par l'augmentation artificielle des données en faisant du « k-fold validation » à 10% et une recherche des paramètres par exploration séquentielle des hyper paramètres résultant en une possibilité de 4356 combinaisons différentes. Ils ont montré que l'utilisation d'un modèle linéaire pouvait emmener des sérieux progrès dans le domaine de l'analyse émotionnelle textuelle. Aman et coll. [76] présentent une nouvelle manière d'annoter les textes afin d'effectuer de la classification émotionnelle. En utilisant des blogues tirés du NET, les auteurs ont annoté chaque phrase faisant référence aux six émotions d'Ekman [48]. Pour les phrases qui ne dégageaient pas une émotion en particulier, la référence « émotion mixte » fut utilisée et pour les phrases neutres, « non émotive ». Ils ont aussi attribué une cote sur l'intensité ressentie (faible, moyenne, forte). Les indicateurs émotionnels, groupes de mots ou lettres, ont aussi été recueillis. Vu la nature subjective de l'annotation de phrases émotionnelles, les auteurs ont voulu s'assurer d'une certaine homogénéité entre les juges en se servant du kappa de Cohen [136] vu la nature mutuellement exclusive entre les classes émotionnelles. Ils se sont rendu compte qu'en général le consensus suit la même direction entre les juges ayant annoté les mêmes extraits, ce qui leur a permis de valider en quelque sorte les données acquises. Les auteurs se sont servis des bases de données de mots émotifs « The General Inquirer » et « WornNet-Affect » [107, 137], séparément, et combinés, puis finalement ajoutés à leurs propres identifications syntaxiques, pour performer de la classification émotionnelle en utilisant un réseau bayésien naïf et un SVM avec des résultats assez concluants. Ils ont remarqué que l'ajout de leur propre base de données d'indices syntaxiques émotionnels aux bases de données existantes améliorerait la performance des deux classificateurs. Strapparava et Mihalcea [91] décrivent leurs expériences sur l'analyse émotionnelle en comparant différents classificateurs, un classificateur à mots-clés, un LSA et un réseau bayésien. Il se sont servis d'une base de données émotive développée antérieurement [107, 138] pour effectuer leurs expériences. Cette base de données consiste en des titres de nouvelles annotés émotivement. Suivant les travaux d'Ortony et coll. [139], ils ont considéré non seulement les mots à connotation émotive directe, mais aussi les mots qui pouvaient exprimer de l'émotion indirectement. Pour avoir une base comparative, le premier modèle à base de détection de mots-clés fut développé. Ensuite ils ont développé un classificateur LSA qui permet d'identifier et de classifier différentes données dans un espace vectoriel homogène. Pour représenter les documents sous forme vectorielle, les auteurs de sont servis de la technique « TF-IDF ». Un dernier modèle considéré fut un réseau bayésien entraîné sur des blogues annotés pour détecter les émotions. Les résultats n'étaient pas unanimement en faveur d'un modèle par rapport à un autre lorsqu'on regarde la

classification uni-émotionnelle par modèle. Les auteurs ont donc avancé qu'il fallait récolter d'autres données issues des connaissances sémantiques grammaticales. Pour continuer, Ghazi et coll. [126] présentent un modèle innovateur arrangeant hiérarchiquement la neutralité, la polarité et l'émotion contenues dans un texte, continuant ainsi les travaux avancés antérieurement par les auteurs [140]. Ghazi et coll. commencent par catégoriser les phrases parmi les six émotions de base d'Ekman [48] en considérant également la classe « non émotionnelle ». La hiérarchie des annotations consiste en deux ou trois niveaux, la hiérarchie à deux niveaux sert uniquement à déterminer la polarité textuelle. Les auteurs se sont servis de la base de données annotée de [76], présentée antérieurement et de celle de Alm et coll. [141]. Un classificateur SVM a été utilisé pour faire la classification émotionnelle et sentimentale selon les relations hiérarchiques avancées. La classification se fait donc du premier niveau au dernier en gardant en mémoire l'information recueillie à chaque passage. Ils se sont rendu compte qu'en tenant en compte des résultats des classifications hiérarchiques précédentes pour effectuer celle du niveau actuel pouvait considérablement améliorer la classification émotionnelle, car les niveaux hiérarchiques suivent ce calque : Émotif/Non Émotif -> Polarité -> Émotion. Donc en détectant en premier la présence ou l'absence d'émotion, pour ensuite classifier la polarité (positif, négatif, neutre) et finalement classifier l'émotion détectée permet d'avoir de meilleurs résultats pour la classification finale.

Les techniques utilisées pour effectuer de l'analyse sentimentale sont sensiblement similaires à celles employées pour effectuer de la reconnaissance émotionnelle ; nous ne mentionnons donc que les articles les plus influents. Turney [142] introduit une méthode statistique pour évaluer l'orientation sentimentale d'une phrase en utilisant l'algorithme PMI-IR (Pointwise Mutual Information - Information Retrieval), basé sur l'algorithme PMI de Church [143], qui détermine la force de la relation sémantique entre deux mots dans une phrase. En résumé, cet algorithme se sert d'un moteur de recherche pour déterminer la probabilité qu'une phrase fasse ressortir des commentaires qualifiés de négatifs sur la probabilité qu'une phrase négative fasse ressortir des résultats positifs. En calculant le score moyen de toutes les phrases d'un commentaire, les auteurs peuvent qualifier un extrait textuel comme étant positif ou négatif. Les auteurs de [130], Pang et coll., sont parmi les pionniers de la classification sentimentale moderne. Ils proposent de faire de la classification sentimentale sur des commentaires laissés sur des films du site IMDB³. Une vectorisation par « Bag-of-Words » fut implémentée pour les unigrams et les bi-grams à caractère positifs, puis ces données ont servi à entraîner séparément un classificateur bayésien, un modèle par entropie maximale, et un SVM. Les auteurs se sont rendus compte que pour détecter les sentiments de la manière la plus précise possible, n'utiliser que la présence de mots émotifs et non leur fréquence était plus bénéfique. L'article

3. <https://www.imdb.com/>

de Tan et Zhang [125] introduit une étude empirique pour catégoriser les sentiments dans des documents chinois. Ils ont utilisé quatre techniques différentes de sélection des caractéristiques textuelles pertinentes : le DF (Document Frequency), la mesure statistique CHI, le MI (Mutual Information) et le IG (Information Gain) pour déterminer quels mots seront utilisés afin de construire leurs représentations vectorielles des données selon la technique « TF-IDF ». Par la suite, les auteurs ont entraîné cinq différents classificateurs sentimentaux : un classificateur centroïde (effectivement un KNN avec un seul voisin), un KNN, un classificateur « winnow », un réseau bayésien et un SVM. Selon leurs résultats, ils ont découvert que faire une sélection des mots d'importance à l'aide de l'IG permettait d'obtenir les meilleures performances chez tous les classificateurs, le SVM étant le plus performant, peu importe la technique de présélection des variables utilisée. Ye et coll. [144] font de l'analyse sentimentale sur les commentaires laissés par différents voyageurs sur leurs destinations. Ils effectuent une étude comparative entre les performances de trois différents classificateurs : un réseau bayésien, un SVM et un modèle n-gram à caractères. La vectorisation des commentaires est faite à l'aide de la technique « Bag-of-Words » simple sans termes pénalisants. Le réseau bayésien est alors un simple réseau où on veut maximiser l'espérance sachant la probabilité qu'une suite de mots retenus dans une phrase fassent partie d'une classe sentimentale. Un SVM fut aussi entraîné, essayant de trouver un hyperplan séparant les phrases annotées comme positives des phrases annotées comme négatives avec une marge maximale. Les auteurs ont noté que le SVM performait en général mieux que le réseau bayésien pour cette tâche de classification. Finalement, le n-gram à caractère, contrairement au n-gram conventionnel qui étudie la distribution des mots sur un vocabulaire fixe, étudie la distribution des caractères (lettres, espaces, symboles) sur un vocabulaire de caractères fixe. Pour l'entraînement de ce modèle, ils ont utilisé le « LingPipe DynamicMLClassifier »⁴. Leurs résultats démontrent une précision de 80% pour la tâche de classification sentimentale avec le SVM et le n-gram à caractère performant le mieux. Go et coll. [128] ont écrit un des articles les plus importants en classification sentimentale. Les auteurs étudient eux aussi les microblogs qui circulent sur Twitter pour les classifier sentimentalement. Ils montrent qu'avec des techniques comme les réseaux bayésiens, les classificateurs d'entropie maximale ou bien les SVM, on peut atteindre plus de 80% de précision pour prédire le sentiment d'extraits textuels. L'avantage de Go et coll. est le fait qu'ils utilisent aussi les émoticônes présents dans les textes pour en classifier la polarité. Les caractéristiques textuelles employées sont les unigrams, bi-grams, la combinaison des deux et les unigrams combinés à des pointeurs syntaxiques. Pour les émoticônes, les auteurs les ont retirés pendant l'entraînement, car ils introduisaient trop de bruit et ne faisaient que nuire aux résultats. Les émoticônes ne sont utilisés que pour collecter les microblogs à

4. Alias-I (2006). LingPipe natural language toolkit. : <http://www.alias-i.com/lingpipe/>

tendances négatives ou positives selon les émoticônes utilisés ; les microblogues contenant des émoticônes contradictoires sont rejetés. Les auteurs ont découvert que lorsque l'extraction des caractéristiques textuelles est faite en combinant des unigrams (mots seuls) avec des bi-grams (groupe de deux mots), les résultats de la classification étaient meilleurs et que d'utiliser des pointeurs syntaxiques n'aidait pas du tout, ce qui est compatible avec les conclusions de Pang et coll. [130]. Xia et coll. [73] proposent une étude comparative pour déterminer l'efficacité des modèles combinatoires pour la classification sentimentale. Un modèle combinatoire est une combinaison des techniques d'extraction des caractéristiques textuelles utilisées avec différents classificateurs en vue de combiner leurs forces pour obtenir de meilleures précisions. Les auteurs combinent deux techniques d'extraction d'indices textuels, une basée sur le POS (part-of-speech) et l'autre sur la relation intra vocabulaire (pluri-grams) ; ces deux techniques d'extraction sont ensuite utilisées pour entraîner séparément un réseau bayésien, un modèle d'entropie maximale et un SVM qui serviront de base. Les caractéristiques textuelles sont encodées par la méthode « Bag-of-Words ». Une fois entraînés, ces trois classificateurs sont comparés à différentes techniques d'apprentissage combinatoire : la combinaison des variables d'apprentissage, la combinaison des classificateurs et un mélange des deux. Les classificateurs sont combinés selon trois techniques combinatoires : la combinaison fixe, la combinaison pondérée et la combinaison à l'aide d'un méta classificateur. La combinaison fixe consiste souvent en une somme, un produit, moyenne ou autres règles fixes sur les résultats normalisés des classificateurs à l'étude pour effectuer un vote déterministe afin de calculer un nouveau résultat (classification). La combinaison pondérée est, comme le nom l'indique, une combinaison pondérée des résultats obtenus des classificateurs à l'étude pour obtenir un résultat. Cette combinaison est souvent représentée par une somme pondérée uniforme entre les différents classificateurs, mais n'est pas tenue d'être implémentée ainsi. La méta classification dans ce cas-ci est une simple régression linéaire qui prend comme variables d'entrée les résultats des trois classificateurs et est entraînée afin de minimiser l'erreur de classification entre la classe réelle est celle qui est prédite sur l'ensemble des données. Les meilleurs résultats ont été obtenus en combinant l'extraction lexicale intra vocabulaires combinés au méta classificateur linéaire, sans toutefois être fracassants, chose que les auteurs ont bien notée. Le bémol est que le temps d'inférence est nettement supérieur et que plus la complexité des modèles augmente, ce qui dépendamment du contexte peut être un facteur beaucoup plus pénalisant que d'essayer d'avoir la meilleure précision sur la classification sentimentale.

Les modèles en analyse sentimentale et émotionnelle basés sur les techniques en apprentissage machine sont nombreux et à la fois très similaires les uns des autres. La force de ce genre de techniques, c'est qu'elles peuvent déterminer statistiquement l'influence de mots non affectueux afin de classifier correctement un extrait de texte. Cependant, ces modèles, pour

la majorité d’entre eux, ne prennent pas en compte les structures syntaxiques des textes et l’interaction hiérarchique des mots à l’intérieur d’une phrase, ce qui selon nous est une erreur ; c’est ce qui motive notre recherche à vouloir combiner les forces des modèles en apprentissage machine et ceux issus de la modélisation par règles. De plus, pour avoir une idée générale sur comment classer correctement un extrait, il faut énormément de données pour s’assurer de bien représenter la distribution des caractères textuels utilisés pour effectuer la classification.

2.2.3 Analyse émotionnelle et sentimentale par apprentissage machine profond

Les implémentations en apprentissage profond pour faire de l’analyse sentimentale et émotionnelle sont nombreuses. Les types de réseaux les plus importants et les plus répandus sont : les réseaux neuronaux classiques [71], les réseaux récurrents comme les LSTM, GRU et les RNN [79, 80, 145–151], CNN [79, 80, 151–154], les auto-encodeurs [155] ou bien les réseaux avec un mécanisme d’attention [156, 157].

Pour débiter, explorons un peu ce qui s’est fait du côté de l’analyse émotionnelle. Wang et coll. [158] introduisent un modèle d’attention bilingue pour prédire les émotions dans le texte. Ils utilisent un LSTM pour obtenir la représentation séquentielle de chaque publication dans un document. Cette représentation vectorielle [119] décrit une annotation pour chaque mot du document. Comme les mots ne sont pas tous importants pour prédire les différentes émotions, les auteurs implémentent un mécanisme d’attention, inspiré par les auteurs de [156], qui représente en quelque sorte une loupe entraînée pour n’extraire que les mots qui contribuent de manière statistiquement significative, par rapport aux mots précédents, à reconnaître l’émotion transmise. Le défi ici est que les documents sont écrits en chinois et en anglais. Les auteurs ont commencé par vectoriser les mots chargés émotionnellement pour les deux langues séparément, puis ont continué par la suite pour l’ensemble des documents dans les deux langues, obtenant ainsi trois différentes représentations vectorielles des mots affectueux. Se servant de ces trois représentations vectorielles les auteurs ont ensuite entraîné un classificateur capable de reconnaître les émotions d’Ekman [48]. Plusieurs autres auteurs utilisent des modèles récurrents pour prédire les émotions dans des extraits textuels tels que [149, 159–161]. Pour continuer, Majumder et coll. [150] implémentent un modèle très intéressant, un GRU dialogal aidée d’un mécanisme d’attention pour détecter les émotions à travers une conversation, pour chaque réplique. Le système global est composé de deux GRU pour encoder les répliques des différentes parties et les contextes qui apparaissent tout au long de la conversation, puis aussi l’état général du dialogue. En procédant de la sorte, les auteurs sont capables de tenir compte de tout l’historique de la conversation pour déterminer l’émotion transmise à travers la réplique courante. Pour l’extraction

des caractéristiques textuelles, les auteurs ont utilisé un CNN suivant la technique avancée par [162] pour former un vecteur n-gram contenant les énonciations syntaxiques. Ces vecteurs ont servi à entraîner le modèle récurrent GRU afin de déterminer l'état émotionnel élicité par chaque personne tout au long de la conversation. Le GRU global est entraîné en ayant comme entrée les différentes élicitations syntaxiques et les états des deux personnes durant la conversation. Cette combinaison des états et des élicitations syntaxiques à chaque réplique et celles directement avant représente selon les auteurs le contexte courant de la conversation et c'est ce qui forme le bloc attentif du système. Au final, un GRU émotionnel est utilisé pour classifier l'émotion catégorique [48] de la réplique courante selon l'historique immédiat des émotions, des élicitations syntaxiques détectées précédemment et du contexte dialogal immédiat. D'autres articles implémentent cette idée de reconnaissance émotionnelle dialogale tels que [46, 154, 163–166]. Le dernier modèle que nous aimerions présenter plus en détail est celui de Demszky et coll. [56]. Ce n'est pas le modèle en soi qui soit le plus impressionnant, car ils ont utilisé l'encodeur bidirectionnel attentif entraîné par l'équipe de Devlin et coll. [167], qui s'intitule BERT, pour faire la classification, mais c'est l'effort que cette équipe a mis en place pour produire une base de données de qualité pour effectuer de la classification émotive qui retient l'attention. Cette catégorisation émotive, qui couvre 27 émotions en plus de la neutralité, est tirée de [50]. Les auteurs ont attentivement sélectionné et annotés manuellement 58 000 commentaires en anglais récoltés du site internet Reddit. La motivation des auteurs était que les bases de données émotives disponibles étaient souvent basiques au niveau des émotions identifiées, trop petites, mal annotées ou trop spécifiques pour être appliqués à des domaines plus généraux. Les auteurs ont donc construit la plus grande base de données émotive au monde qu'ils appellent GoEmotions. Les entrées sont regroupées hiérarchiquement par similitude émotionnelle de sorte qu'il soit possible à qui-conque de diminuer la granularité des émotions à reconnaître pour les adapter à leurs besoins. Les commentaires ont été soigneusement choisis pour que les différentes émotions élicitées ne soient pas disproportionnellement représentées pour ainsi améliorer les chances de récolter les meilleurs indices syntaxiques, mots et groupes de mots, qui sont à l'origine de l'élicitation émotionnelle dans un texte. Cette minutie leur a permis d'entraîner un modèle basé sur BERT capable de performer aussi bien dans des tâches spécifiques de reconnaissance émotive que des tâches générales.

Regardons maintenant ce qui s'est fait pour la détection des sentiments en apprentissage profond. Les techniques sont similaires à ce que l'on trouve dans le domaine de l'analyse émotionnelle, mais plus simples vu la nature des deux tâches. Commençons par présenter l'article de Glorot et coll. [71]. Les auteurs ont compris que, dépendamment du domaine, les mots affectueux ne sont pas nécessairement les mêmes. Les auteurs ont alors récupéré 340 000

critiques laissées sur le site d'Amazon, réparties sur 22 types de produits différents, pour en extraire automatiquement les caractéristiques textuelles afin de s'en servir pour entraîner un classificateur sentimental. Pour effectuer l'extraction, les auteurs se sont servis d'un SDA (Stacked Denoising Auto-encoder) introduit par [168]. Dans ce cas-ci, comme les encodeurs et décodeurs sont linéaires et la fonction d'erreur quadratique, cette méthode d'extraction de caractéristiques lexicales équivaut à récupérer l'espace des composantes principales du texte, donc de connaître quels mots affectent la phrase de manière à maximiser la variance sur la polarité perçue tout en diminuant l'impact des mots moins polarisateurs (bruit). Les auteurs se sont servis de cette information pour alors l'injecter en entrée à un classificateur SVM afin d'effectuer de l'analyse sentimentale. Les prochains auteurs à l'étude sont Johnson et Zhang [152], ces auteurs présentent une manière de catégoriser sentimentalement un texte en se servant d'un CNN. Les auteurs sont bien conscients que les techniques de représentation vectorielle des mots, tels que le « Bag-of-Words », ne sont pas parfaites, car on ne regarde pas les mots dans leur contexte (mots voisins) et on perd beaucoup d'information de la sorte. Ils se sont donc inspirés de l'utilisation des CNN en imagerie qui peut tirer avantage d'une structure informationnelle en 2D. Ici, la détection d'influence contextuelle est plutôt faite sur une dimension, le texte en soi. Ce qui est différent de la méthode employée par [150], c'est que le CNN ne sert pas à seulement extraire les caractéristiques textuelles significatives pour les vectoriser afin de servir à l'entraînement. Le CNN ici est le modèle général qui détecte l'influence des mots et de leurs voisins sur un sentiment perçu afin de catégoriser la polarité du texte. On laisse donc le réseau lui-même construire les représentations vectorielles des mots affectueux à travers une série de couches convolutives, puis, effectue la catégorisation à travers une couche linéaire entièrement connectée. Le problème avec l'approche utilisée est la vectorisation des mots. Contrairement à la technique de « Bag-of-Words » qui compte le nombre d'apparitions d'un mot dans un extrait relativement à sa position dans un vocabulaire de taille N , ce qui implique que la longueur totale du vecteur est N , les auteurs emploient un type de représentation « one hot » pour chaque mot relativement à sa position à un vocabulaire. La taille du vecteur est donc de $N \times$ (nombre de mots). Donc si le vocabulaire est immense, le problème devient extrêmement complexe à résoudre. D'autres auteurs, comme Wang et coll. [151], exploitent cette manière de classifier sentimentalement du texte, mais ne font pas de la classification à l'aide du CNN lui-même. Wang et coll. intègrent une couche séquentielle LSTM à la sortie des couches convolutives pour extraire aussi la relation séquentielle entre les phrases ajoutant ainsi une dimension « temporelle » au modèle. Similairement à Wang et coll. [158], Tang et coll. [145] utilisent un modèle GRU pour effectuer de la classification sentimentale de documents. Tout comme dans [152], les auteurs utilisent une architecture convolutive pour déterminer la relation entre les mots d'une phrase et ainsi

extraire l'information significative sous forme vectorielle. Ils se servent donc du CNN pour représenter de manière vectorielle la moyenne des influences des mots significatifs les uns par rapport aux autres. Ces vecteurs (un par phrase) servent alors d'entrée à un modèle GRU qui permet de faire la classification sentimentale sur tout le document en prenant en compte la relation séquentielle des phrases de la même manière que Wang et coll. [151] utilisent un LSTM pour effectuer le même travail. Yang et coll. [156] introduisent une architecture à double mécanisme d'attention pour détecter les sentiments perçus dans des textes. Ils ont implémenté un mécanisme d'attention au niveau des mots afin d'identifier dans une phrase les mots statistiquement significatifs sur le plan sentimental pour ensuite les envoyer dans un deuxième mécanisme d'attention qui détecte comment les phrases encodées contribuent au sentiment général du texte. De cette manière, les auteurs peuvent, d'un coup, prendre en compte la relation hiérarchique des mots dans une phrase et la relation des phrases dans un texte pour classifier la polarité du document. Pour encoder les mots, les auteurs utilisent un GRU bidirectionnel, c'est-à-dire que la représentation vectorielle des mots est faite dans les deux sens de la séquence. Les mots ont chacun une représentation vectorielle différente dépendamment des mots qui suivent et des mots qui précèdent ; la concaténation de ces deux vecteurs donne la représentation d'un mot dans une phrase particulière. Ensuite, on passe chaque mot par le premier mécanisme d'attention qui regarde de quelle manière la représentation vectorielle du mot, face à son contexte, est significative par rapport aux autres mots présents dans la phrase et à leur contexte respectif. C'est une méthode qui attribue un poids de signification à chacun des mots ; en bref, c'est les classer en fonction d'importance du plus important au moins important. Pour former les représentations des phrases, on ne fait qu'additionner la représentation vectorielle des mots ajustée aux poids de signification. Les auteurs effectuent alors les mêmes étapes avec les nouvelles représentations des phrases pour ensuite classifier sentimentalement le document dans son entier.

Les modèles en analyse sentimentale et émotionnelle basés sur les techniques en apprentissage profond sont nombreux et ne se ressemblent pas tout à fait. Les éléments de base se retrouvent partout, mais comme il a été possible de le voir, il y a quasi une infinité de manières possibles de combiner tous ces mécanismes pour créer de nouveaux réseaux. Cependant, comme pour les modèles en apprentissage machine classique, ces implémentations nécessitent énormément de données qui ne sont pas toujours les plus fiables, même s'il est possible de créer soi-même les données à partir de textes de manière non supervisée. Une autre grande faiblesse, c'est encore le manque d'intégration de connaissances générales sur les structures syntaxiques des phrases, des relations intra vocabulaires, ou même des connaissances générales. Oui, certaines techniques d'extraction textuelles par un réseau, telles que les CNN, peuvent prendre en compte le *contexte* entourant les mots ou les phrases, mais ce n'est que de manière sta-

tistique ; on reste donc vulnérable aux données qu'on utilise. Nous pensons que combiner les forces indéniables de l'apprentissage profond avec des connaissances générales modélisées mathématiquement pourrait grandement améliorer les résultats en analyse sentimentale et émotionnelle.

2.2.4 Analyse émotionnelle et sentimentale hybride

Les implémentations hybrides en analyse émotionnelle et sentimentale regroupent plusieurs techniques des différentes sections présentées pour essayer de profiter des forces de chacune. Dans cette section, nous présenterons tout simplement ce qui se fait en analyse sentimentale et émotionnelle sans nécessairement les séparer comme dans les sections précédentes. Ho et Cao [169] présentent un modèle markovien caché combiné à des méthodes d'extraction lexicales pour déterminer l'émotion dans un texte. Les auteurs se servent d'un parseur syntaxique pour identifier les composantes grammaticales dans la phrase. Chaque composante est alors identifiée comme étant une sous-composante d'une élicitation émotive ou bien comme étant émotive en soi. Ces groupes de composantes servent d'états cachés dans le réseau markovien. Les auteurs se servent d'une base de données annotée pour entraîner le réseau et ainsi déterminer les paramètres du réseau. L'idée est qu'en identifiant de quelle manière une séquence d'états contribue à une émotion, il sera possible de reconnaître l'émotion dans n'importe quel texte contenant ces éléments. Shaheen et coll. [87] quant à eux présentent une manière de reconnaître les émotions textuellement en générant automatiquement des règles d'identification d'indices affectueux pour classifier l'extrait. Par exemple, la phrase « J'ai reçu beaucoup de cadeaux à Noël » identifiée comme joyeuse sera analysée par le système et celui-ci identifiera le duo « recevoir cadeaux » comme étant le déclencheur émotive. Pour créer leurs règles, les auteurs se servent d'abord d'outils en NLP pour extraire les mots significatifs du texte et identifient le type de mot. Par la suite ils créent un arbre de branchements qui relie les mots entre eux selon la relation qu'ils ont les uns par rapport aux autres. Par exemple, dans la phrase « c'était le meilleur été que je n'ai jamais eu », le verbe « avoir » est en relation avec le mot « meilleur » et « été ». En analysant bon nombre de phrases de la sorte, les auteurs peuvent construire un catalogue de relations syntaxiques associées à une émotion pour identifier ces mêmes indices syntaxiques dans des phrases complètement différentes afin de les catégoriser émotionnellement. Huynh et Le [88] présentent un modèle intéressant qui intègre des notions grammaticales à un CNN pour la reconnaissance émotive. Similairement à ce que [150, 152] ont fait, les auteurs utilisent un CNN pour à la fois vectoriser les mots et faire la prédiction, mais ici ajoutent une couche en plus. Ils identifient préalablement les mots qui ont de fortes connotations émotives et, lors de l'encodage, doublent les vecteurs associés à ces mots pour augmenter le poids qu'ils ont sur la catégorisation émotionnelle

finale en les concaténant et pour le reste on greffe un vecteur nul. Les auteurs ont remarqué qu'intensifier de la sorte les mots à connotations émotionnelles améliorerait la classification comparée à un même modèle qui ne prend pas avantage de ces connaissances grammaticales. Ghosal et coll. [164] présentent un modèle qui prend en considération des connaissances générales que nous avons sur le monde pour inférer l'émotion dans une conversation. Les auteurs utilisent RoBERTa [170] pour encoder les réponses de chacun des interlocuteurs, puis utilisent un transformateur, appelé COMET [171], pour extraire les caractéristiques lexicales représentant les connaissances générales. Ils gardent aussi en mémoire l'évolution du contexte conversationnel, car, comme vu plus tôt, le contexte conversationnel continu permet d'avoir une meilleure idée des intentions émotives des interlocuteurs à chaque interaction. En passant les interactions encodées avec RoBERTa, COMET et le contexte actuel à un GRU séquentiel, les auteurs arrivent à catégoriser les émotions élicitées à chaque interaction. Khan et coll. [84] introduisent une heuristique de nettoyage combiné, à une série de règles décisionnelles pour classifier adéquatement les sentiments perçus dans les microblogues sur Twitter. Les auteurs nettoient d'abord les microblogues en retirant les hyperliens, les caractères spéciaux, noms propres, mots informels et mots sans information, puis réduisent ce qui reste à leurs racines. Ils se servent par la suite d'une séquence de trois classificateurs différents pour essayer de déterminer la polarité du texte : un classificateur qui prend en compte les émoticônes [172], un classificateur de polarité textuelle [173] et un classificateur émotionnel à base détection de mots-clés [174]. Le but d'utiliser ces trois classificateurs séquentiellement est de diminuer les chances de détecter un microblogue comme étant neutre. Tout à tour, dans l'ordre, l'extrait nettoyé est envoyé à chacun des classificateurs, si l'extrait est détecté comme neutre, il passe au suivant, sinon on garde la classification sentimentale obtenue. Si au troisième classificateur le texte est encore perçu comme neutre, il est alors classifié comme neutre. Ici, les auteurs font une combinaison de plusieurs approches pour espérer avoir de meilleurs résultats et éviter de classifier les microblogues comme neutres s'ils ne le sont pas réellement. Pandey et coll. [83] proposent un modèle qui détecte les mots à connotation sentimentale dans les microblogues de Twitter pour les regrouper dans des clusters de similitude. Par la suite, un algorithme de recherche cuckoo est utilisé pour déterminer si les mots dans un nouveau tweet se retrouvent dans un des clusters associés à un sentiment pour enfin classifier le texte.

2.3 Génération de textes émotifs

La génération de réponses émotives est un sous-domaine de l'analyse émotionnelle et sentimentale qui vise à créer ou transformer des réponses ou textes pour qu'ils soient chargés émotivement selon une émotion et un contexte visé. Dans la littérature, plusieurs implémen-

tations de ce type existent et selon Ma et coll. [175], il existe deux types de modèles qui cherchent à combler ces fonctions : les encodeurs qui sont capables d’interpréter l’émotion dans un dialogue, et les décodeurs capables de générer des réponses émotives. Les encodeurs émotifs sont des mécanismes qui encodent implicitement ou explicitement à l’aide de classificateurs émotifs l’aspect émotionnel d’un extrait textuel et se servent de cet encodage pour essayer de générer du texte émotif. Les décodeurs émotifs quant à eux essaient de générer des réponses émotives selon une émotion cible et/ou un contexte cible en ayant un contrôle supplémentaire lors du décodage pour générer une réponse affective. Ces modèles sont souvent combinés pour avoir un système capable de générer du contenu émotif en tenant compte de l’interaction textuelle.

2.3.1 Encodeurs de réponses émotives

Dans cette section, nous allons couvrir les implémentations qui ne couvrent que l’encodage émotionnel sans apport direct sur la génération de réponses émotives, c’est-à-dire que l’unique contrôle sur l’émotion générée est effectué à travers l’encodage des données en amont du générateur. Huang et coll. [176] présentent une méthode d’encoder l’émotion d’un texte à l’aide d’un LSTM avec un mécanisme d’attention. Ils encodent chaque émotion voulue de manière individuelle sous forme de « mot », ou jeton, qu’ils concatènent à la phrase qu’ils veulent encoder. Pour détecter l’émotion dans la phrase, ils se servent d’un classificateur externe. Les auteurs expérimentent avec le fait de concaténer le jeton au début ou à la fin de la phrase, ce qui selon les auteurs constitue deux modèles différents. La troisième approche consiste à concaténer à la phrase encodée vectoriellement un vecteur émotion qu’ils passent ensuite à un décodeur pour générer une nouvelle phrase émotive. Rashkin et coll. [177] encodent l’émotion dans son contexte en se servant d’une nouvelle base de données qu’ils nomment *EMPATHETICDIALOGUES* et qui consiste en des petites conversations autour d’un sujet où les candidats sont emmenés à converser de manière à faire sortir l’émotion cible. Les auteurs se servent alors de la partie encodeur d’un transformateur [178] pour encoder les réponses émotives qui sont indexées selon l’émotion qu’elles élicitent. La génération de réponses émotives est alors contrôlée par les vecteurs de sortie de l’encodeur et est effectuée à l’aide d’un décodeur standard qui minimise le logarithme de la vraisemblance négative conditionnée sur le vecteur en entrée et le contexte identifié. Kong et coll. [179] implémentent un réseau antagoniste génératif conditionnel (CGAN) [180] pour contrôler la génération de répliques émotives. Les auteurs encodent l’historique conversationnel à l’aide d’un RNN, puis se servent d’un encodeur-décodeur pour encoder la réponse à laquelle il faut générer une réplique et finalement concatènent ces deux vecteurs avec la caractéristique émotionnelle recherchée, encodée elle aussi. Ce nouveau vecteur sert alors d’entrée à un générateur-discriminateur

qui reproduit une réponse émotionnelle grammaticalement adéquate. Niu et Bansal [181] présentent trois modèles différents pour générer des réponses polies ou impolies. Le premier modèle présenté est un encodeur-décodeur séquence à séquence LSTM de base qui génère une réponse polarisée, c'est-à-dire que les réponses sont encodées par un LSTM et décodées par un autre LSTM multicouche. Le deuxième modèle, appelé Fusion, combine le modèle de base avec un encodeur LSTM uniquement entraîné sur des données identifiées comme polies (polite-LM). Une combinaison linéaire des phrases encodées par ces deux modèles est alors effectuée pour l'envoyer à un décodeur LSTM afin de générer une phrase polarisée. En gardant un contrôle sur l'apport de polite-LM dans la combinaison linéaire, les auteurs peuvent contrôler le sentiment projeté dans la phrase générée. Le problème est que l'apport du polite-LM ne tient pas compte du contexte historique conversationnel, donc les phrases générées peuvent être polies sans être forcément pertinentes. Pour contrer ces limitations, les auteurs ont implémenté un troisième modèle qui combine le deuxième modèle avec un modèle par apprentissage par renforcement (polite-RL). Polite-RL est entraîné pour encourager la génération de réponses polies et ces répliques servent comme vecteurs « d'encouragement » au LSTM générateur. D'autres auteurs se servent de mécanismes similaires pour encoder les réponses dialogales combinées à des vecteurs émotionnels tels que [182, 183].

2.3.2 Décodeurs de réponses émotives

Dans cette section, nous allons présenter les modèles qui entraînent explicitement un décodeur-générateur pour influencer l'émotion élicitée dans une réponse générée. Le contrôle ne se fait pas uniquement dans les couches à l'amont du décodeur comme vu dans les modèles précédents. Zhou et Wang [184] présentent un modèle par apprentissage par renforcement combiné à un autoencodeur variationnel conditionné (CVAE) [185] pour contrôler la manière de générer des réponses émotivement chargées à l'aide d'émoticônes. Dans la version sans apprentissage par renforcement, le CVAE encode les microblogues puis concatène le résultat à un vecteur représentant les émoticônes présentes encodées séparément. Un vecteur représentant la réponse réelle au microblogue en question est encodé séparément, puis combiné au vecteur du microblogue et des émoticônes originales pour les envoyer à un décodeur qui générera une réponse affective par l'utilisation de nouvelles émoticônes. Pour construire le modèle par apprentissage par renforcement, les auteurs entraînent un classificateur d'émoticônes qui servira à construire la fonction de valeur du système. Le système utilise alors le CVAE original pour générer une réponse et utilise le classificateur d'émoticônes pour forcer le système à construire une réponse maximisant la valeur du modèle en utilisant les émoticônes appropriées aux endroits appropriés. Ghosh et coll. [186] présentent un modèle générateur LSTM qui peut compléter un début de phrase, le contexte, créant ainsi une phrase émotive

en ajustant directement les paramètres internes du modèle à la volée pour cibler une émotion et intensité émotionnelle recherchée. Le modèle génère mot à mot chaque composante de la phrase en conditionnant le LSTM sur l'émotion recherchée pondérée par l'intensité recherchée. Le modèle prend alors en compte chaque mot à être généré puis, lorsqu'un mot clé cible est détecté, celui-ci est remplacé par un mot dans le même espace vectoriel, construit au préalable, émotivement chargé conditionnellement à l'émotion que l'on veut éliciter. En changeant la condition sur l'émotion recherchée et le facteur de pondération de l'intensité, les auteurs arrivent à directement contrôler quels types de mots seront utilisés pour obtenir une réplique émotionnelle. Zhou et coll. [148] présentent un modèle encodeur-décodeur séquentiel GRU pour générer des réponses émotives. Les auteurs conditionnent ce modèle pour une émotion donnée encodée et injectent cette donnée en entrée, concaténée à la phrase encodée par le modèle, comme entrée au décodeur. Le décodeur GRU garde trois états internes, l'état émotionnel, la mémoire interne et la mémoire externe qui travaillent ensemble pour générer une réponse émotive. L'état émotionnel et la mémoire interne travaillent de manière dynamique pour s'assurer que le modèle garde une constance grammaticale tout en considérant les émotions que les auteurs veulent éliciter textuellement. La mémoire externe, quant à elle, est un module qui sert de discriminateur de vocabulaire donnant la possibilité au décodeur de choisir de générer des mots à partir de l'espace vocabulaire vectoriel interne ou de choisir des mots qui font partie d'un lexique émotionnel externe en attribuant explicitement une distribution de probabilités aux mots contextuellement et syntaxiquement appropriés. De cette manière, les auteurs peuvent biaiser le décodeur à choisir des mots affectueux dont l'élicitation émotionnelle est bien connue dans un contexte donné. Xu et coll. [187] présentent un modèle de transformation sentimentale d'une phrase donnée ; ici, il n'est pas question de générer une nouvelle réponse, mais de transformer la phrase initiale pour qu'elle soit sentimentalement chargée selon la polarité ciblée sans perdre le contenu de celle-ci. Les auteurs se servent d'un modèle par apprentissage par renforcement pour effectuer la transformation et cet ensemble est constitué de deux composantes, un mécanisme neutralisateur et un mécanisme d'injection émotionnelle. Pour le module neutralisateur, les auteurs utilisent un LSTM de sorte à pouvoir générer une probabilité d'appartenance sentimentale à chaque mot dans une séquence, le LSTM reconnaît les mots non polarisateurs, puis le module les garde en mémoire et élimine le reste qui doit être régénéré. Pour le mécanisme d'injection émotionnelle, les auteurs utilisent un encodeur-décodeur à doubles décodeurs, un pour chaque axe polarisateur. Le décodeur est alors en charge de remplacer les mots émotionnels manquants par des mots ou séquence de mots dont la polarité est inversée compte tenu du contexte syntaxique, car l'intention est d'inverser la polarité d'une phrase donnée en entrée. Le modèle est alors entraîné de sorte à maximiser la différence sentimentale entre la phrase donnée en entrée et celle

générée ; de cette manière ils entraînent simultanément le neutralisateur à bien identifier les mots non affectueux et le décodeur à générer des réponses sentimentalement opposées. D'autres implémentations moins connues qui utilisent sensiblement les mêmes concepts que ceux présentés existent et œuvrent aussi de manière à influencer directement la manière dont le décodeur-générateur exhibe une émotion dans un texte généré [188, 189].

Les modèles présentés, qu'ils soient du type à injecter l'émotion dans l'encodeur ou de contrôler directement le décodeur pour la génération de réponses affectives, sont tous issus de l'apprentissage machine profonde et comportent plusieurs lacunes. Par exemple, les modèles qui ne font que combiner l'information émotionnelle pendant l'encodage ou qui concatènent une représentation émotionnelle au vecteur texte n'ont absolument aucun contrôle sur les phrases générées par le décodeur. Ces modèles dépendent de la distribution probabiliste de l'usage des mots à connotation émotive dans leurs contextes utilisés lors de l'entraînement. Ce problème est soulevé par les auteurs de [176] qui remarquent que les réponses générées ont tendance à souvent utiliser les mêmes mots pour faire éliciter une émotion, car ces mots sont disproportionnellement présents dans les textes utilisés pour entraîner le modèle. Les auteurs de [179] dénotent aussi des problèmes au niveau de la génération de réponses trop génériques par manque de contrôle au niveau du générateur, chose à laquelle ils ont essayé de remédier en employant une technique dite de « teacher forcing », donnant ainsi accès au générateur à des répliques dites vraies afin de rediriger l'apprentissage dans la bonne direction. Les modèles qui influencent directement le générateur aussi ont leurs lacunes. Zhou et coll. [184] mentionnent que le classificateur émotif entraîné sur des émoticônes et qui sert de mécanisme de renforcement n'est pas parfait et ainsi impacte négativement la capacité du générateur de formuler des microblogs/répliques émotionnellement chargés. Les auteurs de [148] notent que l'utilisation de données pas assez uniformément distribuées au niveau des répliques émotionnelles impactait négativement la capacité du modèle à pouvoir générer des réponses émotives adéquates pour certaines émotions cibles.

2.4 Les émotions dans l'apprentissage par renforcement

Les algorithmes par renforcement sont des implémentations algorithmiques permettant à un agent d'apprendre par l'interaction de celui-ci avec son environnement [190, 191]. Pour permettre au lecteur de mieux se situer, introduisons brièvement les méthodes générales existantes. En apprentissage par renforcement, il existe deux familles d'implémentations : les implémentations basées sur des modèles et les implémentations dites libres de modèles. Ce qui distingue ces deux familles, c'est la manière dont on construit l'algorithme. Si on connaît le processus de décision markovien [192] pouvant encadrer l'apprentissage du modèle et que

nous le fournissons comme entrée, cette implémentation est qualifiée comme étant basée sur un modèle ou « model-based » en anglais [193]. Si on ne connaît pas les processus décisionnels environnementaux pouvant encadrer l'apprentissage de l'implémentation algorithmique et que celle-ci est forcée à déterminer à travers des essais/erreurs ces différents chemins empruntables ainsi que leurs gains/pertes associés, alors on qualifie cette implémentation comme étant libre de modèles ou « model-free » en anglais [193]. Plusieurs types d'implémentations computationnelles ont été proposés à travers le temps, tels que les méthodes d'itération de valeur [194,195] et les méthodes d'itération de politique [196] qui sont basées sur des modèles, ainsi que le « Q-learning » [197,198], une méthode spécifique de différence temporelle, et les méthodes de Monte-Carlo [199,200] qui sont généralement libres de modèles.

Dans le domaine de l'apprentissage par renforcement, lorsque l'on parle d'intégrer les émotions, on parle surtout d'effectuer du biomimétisme au niveau de l'élicitation des émotions chez l'homme et de ses réactions face à celle-ci. Les mécanismes d'élicitation émotive à l'étude sont les élicitations homéostatiques ou extrinsèques [201–204], les mécanismes d'élicitations basés sur la théorie des appraisals des émotions ou intrinsèques [39,41,205,206], ceux basés sur des mécanismes de valeurs et récompenses [207–210], puis ceux qui sont codés de manière rigide [12,211].

2.4.1 Les implémentations par apprentissage par renforcement suivant les élicitations émotive homéostatiques et extrinsèques

Les élicitations émotionnelles dites homéostatiques ou extrinsèques sont des provocateurs émotifs dont les variables déclencheuses se trouvent à l'extérieur du corps. Cette notion de déclencheurs externes fut introduite par Hull en 1943 [212]. Elle dit que ce mécanisme d'élicitation émotionnelle est le principal moyen pour un agent/personne d'apprendre. Hull décrit ces mécanismes comme étant codé génétiquement et que leur présence fût bénéfique tout au long de notre évolution. Pour donner un exemple de ce que constituent des facteurs homéostatiques d'élicitation émotive, on peut nommer la soif, la faim ou le manque de sommeil qui provoquent chez l'homme une motivation pour chercher de l'eau, de la nourriture ou se reposer. L'analogie computationnelle peut être, par exemple, le niveau d'un réservoir, la capacité restante de la batterie interne, ou l'attente d'une mise à jour. Avoir un agent sensible à ce type de stimuli, comme un être vivant le serait, peut nous permettre de programmer des actions appropriées pour ce suivant les situations. Cela peut aussi aider l'agent de changer dynamiquement d'objectif sans même nécessiter d'intervention, comme le mentionnent certains auteurs [204,213–216]. Pour implémenter ce type de réactions face aux élicitations homéostatiques dans un modèle d'apprentissage par renforcement, on peut associer au modèle un

ensemble de variables homéostatiques dont on fait le suivi tout au long de l'apprentissage, puis associer à chaque variable un ensemble de références à une action à entreprendre ou un état physique observable. Chaque variable homéostatique peut être associée à un ensemble de valeurs spécifiques, qui une fois atteintes, permettent de déclencher des actions pour stabiliser le système. Par exemple, on peut associer le niveau de la batterie interne de l'agent ou robot à l'action d'aller se recharger. Selon Moerland et coll. [12], cette relation dynamique dans un système d'apprentissage par renforcement peut s'écrire sous la forme

$$h_{i,t+1} = \begin{cases} h_{i,t} + \psi(s_t, a_t) & \text{if } a_t \in \bar{a}, s_t \in \bar{s}, \\ h_{i,t} - \epsilon & \text{sinon.} \end{cases}$$

où i indique l'indice de la variable homéostatique, t indique l'indice associé au temps, a est une variable de consommation de ressource permise dans \bar{a} , s une variable indiquant de la présence de la ressource parmi les ressources \bar{s} , $H_t = h_{1,t}, h_{2,t}, \dots, h_{N,t}$ un ensemble de variables extrinsèques, et un effet $\psi(s_t, a_t)$ dépendant de la taille, et disponibilité de la ressource.

Notons aussi qu'il y a une forte corrélation entre l'état homéostatique d'une personne et la réponse émotionnelle qu'elle peut ressentir. Cette corrélation peut être exploitée de manière intéressante pour l'implémentation d'un agent en apprentissage par renforcement mimant les mécanismes déclencheurs d'émotions chez l'humain. Par exemple, la théorie des marqueurs somatiques de Damasio [17] dit que les émotions sont le résultat de sensations physiologiques, ce que confirme les théories de James [5] et celle de Schachter [18]. Gadanho et Hallam [201], pionniers de l'implémentation émotionnelle d'un agent en apprentissage par renforcement, ont par exemple implémenté un ensemble de variables homéostatiques représentant la douleur qui augmente avec le nombre de collisions enregistrées ou la température qui est associée à la durée d'utilisation de leur robot. Les émotions sont alors directement reliées à la fluctuation de ces variables internes. Par exemple la faim, associée au nombre d'unités de temps avec un niveau de ressource restreint, est associée à de la tristesse, tandis que la diminution de la faim couplée à un usage des moteurs internes est associée à de la joie. Ce qui est intéressant, c'est que les auteurs ne se contentent pas d'uniquement associer les états des différentes variables homéostatiques à des émotions, mais considèrent également leurs interactions avec les émotions en tant que telles. Par exemple, les effets que la peur, associée aux collisions fréquentes, aura sur le robot seront moins importants si le robot *ressent* de la faim ou de la fatigue. Le modèle émotif utilisé par Gadanho et Hallam est catégorique dans ce cas-ci, car ils font une distinction claire entre les émotions pouvant être élicitées. De plus, ces auteurs étendent même l'influence des émotions associées aux variables internes à la fonction de récompense du robot, c'est-à-dire que dépendamment de l'état homéostatique, la fonction de valeur peut être dynamiquement modifiée. Par exemple, ils ont utilisé la faim du robot pour modifier

la valeur de la récompense associée à la collecte d'énergie. Si le robot n'est pas en manque d'énergie, la collecte additionnelle de celle-ci est pénalisée, et donc, ce comportement sera moins exécuté. De manière similaire, Coutinho et coll. [57] adoptent des variables homéostatiques semblables à la méthode de Gadanho et Hallam. Cette fois-ci, les auteurs ont adopté une approche plus organique. Par exemple, ils comparent le taux de sucre dans le sang avec le niveau d'énergie du robot qui, lorsque trop bas, ne produit pas une émotion en soi, mais affecte négativement un concept de *bien-être*. Ce concept de *bien-être* remplace les émotions catégoriques vues précédemment. Cette idée représente plutôt une émotion sous une forme dimensionnelle non catégorique. Les auteurs accordent aussi une autre dimension émotionnelle au robot : l'excitation physique qui, par exemple, baisse lorsque le robot a besoin de repos. Cette représentation émotive est directement adoptée de la théorie des émotions de Russell [52]. Une différence importante par rapport à l'implémentation de Gadanho et Hallam est l'interaction des variables extrinsèques. Ces dernières n'ont aucune influence sur le dynamisme de la fonction de valeur. D'autres modèles similaires exploitent le concept d'élicitation émotive homéostatique [202, 203, 217]. Un problème avec ce genre d'approches, c'est la perte de granularité au niveau des réactions.

2.4.2 Les implémentations par apprentissage par renforcement suivant les élicitations basées sur la théorie des appraisals des émotions

Les élicitations émotive de la théorie des appraisals ont été introduites par Lazarus et Folkman [21]. Cette théorie présente l'apparition d'une émotion chez l'homme comme une combinaison de plusieurs facteurs internes⁵ tels que l'expérience du sujet face aux stimuli et de sa capacité d'y répondre de manière appropriée. C'est d'ailleurs pourquoi on associe aussi ces mécanismes d'élicitation émotionnelle à des motivations intrinsèques en apprentissage par renforcement ; car dans les deux cas l'émotion est générée par des variables internes. En apprentissage par renforcement, on représente ces variables comme étant une fonction de l'historique expérimental de l'environnement de l'agent et de sa politique. Moerland et coll. [12] représentent cette relation la manière suivante :

$$\zeta_j(s, a, s') = f_j(g, M)$$

Avec ζ , la fonction d'état qui dépend de l'état initial s , l'action a et l'état suivant associé s' . Cette fonction est égale à la fonction f , qui dépend de son expérience historique environnementale en relation au modèle g et la politique de l'agent M .

5. Dimensions de la théorie des appraisals : certitude, amabilité, activité attentionnelle, contrôle, effort anticipé et responsabilité

L'intégration de la théorie des appraisals est un sujet de recherche actif [205]. Cette approche permet d'avoir des agents qui réagissent mieux face à différents stimuli environnementaux et qui s'adaptent dynamiquement face aux différents états futurs possibles. Comme dans le cas de l'intégration de variables homéostatiques, les variables intrinsèques peuvent directement ou indirectement affecter le comportement du robot, soit en modifiant dynamiquement l'objectif ou soit en modifiant la fonction de valeur/récompense. Marinier et Laird [39] implémentent un agent incorporant les dimensionnalités de la théorie des appraisals avancés par Scherer [20]. Ces variables sont du type sensoriel, comme la pertinence ou la surprise face à un stimulus, ainsi que du domaine du raisonnement, c'est-à-dire le contrôle que l'agent peut exercer sur un stimulus. Ils interprètent la surprise en considérant le laps de temps passé entre l'état présent et la dernière fois que cet état fût atteint. Plus l'écart est grand, plus l'agent ressentira de la surprise, ce qui l'amènera à modifier la fonction de valeur associée à l'état présent. Par exemple, si le robot est en état de surprise, plus il sera plus prudent, ce qui diminuera la récompense associée à l'état peu fréquenté. Les variables intrinsèques sont aussi représentées sous forme d'ensemble. Comme pour les implémentations d'élicitation homéostatiques, leurs influences sur la fonction de valeur sont souvent gouvernées par un facteur de pondération qui agit sur la fonction d'évènements historiques de l'agent. En d'autres termes, si un agent fait face à un certain évènement de manière répétitive, dans des intervalles de temps réduits, les facteurs pondérant la récompense associée à l'état suivant seront plus forts, car l'intervalle de confiance associé à la probabilité d'atteindre un état suivant bénéfique sera réduit et par conséquent l'issue sera plus certaine. Cette manière d'adapter la fonction de valeur dynamiquement fait directement référence à l'adaptation de l'homme face aux situations répétitives. Plus nous faisons face à un évènement, plus nous maîtrisons les différentes issues possibles et plus nous serons confiants face aux actions à entreprendre [20, 21, 218]. Il est aussi intéressant de noter que Marinier et Laird n'associent pas une émotion particulière, catégorique ou dimensionnelle, relative aux variables intrinsèques, ce qui n'empêche toutefois pas l'agent de l'adapter dynamiquement à son environnement. Sequeira et coll. [38], implémentent un agent qui, en plus d'utiliser un sous-ensemble des différentes dimensions de la théorie des appraisals, utilise une variable représentant une *pertinence motivationnelle*. Cette nouvelle dimension représente une relation directe entre l'agent et son but ultime. Plus l'agent s'éloigne de son but, plus cette variable intrinsèque tend vers zéro ce qui indique qu'il entreprend des actions contraires à son objectif. Une autre variable intrinsèque importante est celle du nombre de visites des différents états. Elle ressemble à la variable fréquentielle de Marinier et Laird [39], qui garde en mémoire la durée entre les visites d'un même état, sauf qu'ici on ne regarde que le nombre de fois qu'un état est visité. Beaucoup de visites se traduisent par une dimension de contrôle élevée chez l'agent et un nombre peu élevé représente de la

nouveauté. D'autres mécanismes similaires permettent à l'agent d'ajuster dynamiquement sa politique, en tenant compte de son expérience ainsi que de son environnement immédiat. De manière générale, les implémentations basées sur la théorie des appraisals utilisent des approches similaires avec des différences relativement subtiles. Citons aussi Yu et coll. [206] ou William et coll. [219] qui utilisent des analogies des dimensionnalités de la théorie des appraisals pour un robot de planification qui est capable d'adapter aux différents obstacles qu'il peut rencontrer.

2.4.3 Les implémentations par apprentissage par renforcement suivant des mécanismes de valeurs et récompenses

Les implémentations en apprentissage par renforcement qui se basent sur les élicitations émotionnelles fondées sur les concepts de valeurs et de récompenses ne sont pas considérées directement des mécanismes d'élicitation émotionnelle [12]. En effet, il est ici question d'émotions *d'agents* programmées. Moerland et coll. [12], catégorisent les implémentations en quatre types : celles basées sur la valeur, celles basées sur les différences temporelles, celles qui s'appuient sur la récompense et finalement celles qui utilisent une récompense moyenne.

Bozinovski et coll. [207] ont été des pionniers à avoir implémenté un agent décisionnel en faisant l'analogie entre la valeur du système et les émotions. Les auteurs présentent une implémentation où l'agent adapte dynamiquement son apprentissage en fonction de la situation actuelle, l'action à entreprendre et la conséquence de celle-ci interprétée comme la récompense combinée à la valeur de l'état suivant. Ils associent alors directement la valeur du système à son émotion, sans faire d'interprétations supplémentaires comme le font les systèmes basés sur des mécanismes d'élicitation émotionnelle biologiques. L'agent optimise sa politique en s'assurant d'atteindre toujours un état émotionnel positif. Il est important de noter que les auteurs ne se servent pas de « l'émotion » de l'agent pour influencer dynamiquement la fonction de valeur qui le gouverne, mais en font seulement l'analogie. Matsuda et coll. [208] ont développé une implémentation similaire, mais gardent une variable séparée associée directement à la valeur du système. Cette variable qui suit directement l'évolution de la valeur du système sert à modéliser la peur chez l'agent. Les auteurs ont implémenté une série de conditions inconnues du modèle qui servent à discipliner l'agent en augmentant la variable de la peur si ces conditions sont satisfaites. Par exemple, si le but de l'agent est d'atteindre une cible, dans un quadrillé délimité, dans un temps limite, et si cette limite est dépassée, toutes les valeurs des états diminuent avec chaque état suivant, pénalisant ainsi la fonction de valeur. Cette variable, associée à la peur, permet à l'agent de diminuer les actions qui induisent une pénalité de la fonction de valeur. Les chercheurs ont démontré que l'uti-

lisations de ces concepts pénalisants permettait de diminuer les erreurs et conséquemment d'améliorer le comportement de l'agent.

Un autre type d'implémentation émotive que l'on retrouve utilise la différence temporelle comme mécanisme d'élicitation. En particulier le signe de la différenciation temporelle et la portée de celle-ci sont principalement considérés. Jacobs et coll. [209] présentent un agent dynamique qui associe la joie et la détresse à la différence temporelle. Les auteurs présentent une cartographie émotive qui dépend des valeurs actuelles des variables utilisées. Pour ne pas répéter les mêmes actions, les auteurs introduisent aussi un mécanisme de fréquence de visites d'états qui a pour but de diminuer l'importance d'une différence temporelle positive si l'agent visite un même état à répétition. Ce même mécanisme est employé pour la détresse, mais pour des valeurs de différence temporelle négatives : répéter ces mêmes actions augmente l'influence de la différence temporelle. Si l'action à entreprendre peut augmenter la valeur totale du système, alors l'agent ressentira de l'espoir, et il sera encouragé à prendre ce chemin. La peur se développe aussi directement avec la valeur du système, elle augmente si les états suivants font diminuer la valeur du système. D'autres auteurs ont implémenté des fonctionnalités similaires telles que celle de Moerland et coll. [210] qui suit sensiblement la même approche.

2.4.4 Les implémentations par apprentissage par renforcement émotionnels codées de manière rigide

Les implémentations codées de manière rigide sont basées sur des comportements codés directement à partir d'un apport informationnel, c'est-à-dire que les comportements de l'agent sont directement reliés à un évènement sensoriel externe, contrairement aux autres méthodes présentées qui décrivent des élicitations émotive directement issues de l'interaction de l'agent avec son environnement. Ces agents peuvent par exemple prendre en entrée une vidéo de personnes avec lesquelles ils interagissent pour reconnaître l'émotion chez l'humain et ensuite adapter leur comportement. Par exemple, Hasson et coll. [220] présentent un agent qui ajuste les valeurs internes de ses états en fonction des interactions qu'il a avec des participants via une caméra. Un autre exemple est proposé par Kubota et Wakisaka [211]. Ils présentent un agent qui utilise à la fois l'image et le son de son environnement pour interagir avec ses utilisateurs. Ils effectuent de l'analyse sentimentale sur la trame sonore afin de détecter la polarité de l'interaction dans l'optique de diriger les actions futures du robot. Ces entrées affectent alors directement la fonction de valeur du robot et les émotions du robot sont alors cartographiées en utilisant un mécanisme de différenciation temporelle des états. Dans Ficocelli et coll. [221], les auteurs présentent un agent capable d'interagir socialement avec des humains en adoptant un comportement émotif. L'agent est alors entraîné pour afficher

l'émotion adéquate selon le bien-être détecté chez la personne. C'est un système complexe, car dans ce cas-ci le robot doit aussi prendre en charge des tâches à effectuer pour assister l'utilisateur tout en s'assurant de continuellement être en concordance émotive avec celui-ci. La priorité des tâches à effectuer est alors directement influencée par l'état émotif de la personne prise en charge et les variables intrinsèques du robot qui le poussent à atteindre son but qui est à la fois de terminer les tâches et de plaire à l'utilisateur.

2.5 Métriques d'évaluation automatique

Dans le cadre de projets qui consistent à générer des réponses émotives ou simplement générer du contenu textuel, il est très difficile de mesurer de manière adéquate la qualité du dialogue généré. En effet, l'évaluation de ce type de modèles s'avère être presque aussi complexe que l'implémentation de celui-ci [222]. Plusieurs types de métriques supervisées et non supervisées ont été présentées pour essayer de capturer de manière satisfaisante la qualité des réponses générées par des modèles en intelligence artificielle. On identifie trois types de métriques pour mesurer la qualité de génération de texte : les métriques basées sur une comparaison par superposition de mots, les métriques basées sur la distance des mots similaires et les métriques libres de références [223].

2.5.1 Les métriques basées sur la superposition des mots

Les métriques basées sur la superposition des mots consistent à comparer une réponse générée à une référence connue pour en mesurer la qualité. Les métriques les plus courantes de ce type sont le score BLEU [224], le score ROUGE [225] et le score METEOR [226].

La métrique BLEU fut introduite par Papineni et coll. [224] afin de pouvoir évaluer la qualité des réponses générées automatiquement pour des tâches de traduction. Elle consiste à faire une comparaison par n-grams d'apparition unique dans la réponse générée versus la réponse de référence. Cette métrique a été introduite pour palier à un problème sévère de comparaison par n-grams classiques qui ne font que calculer la proportion des mots, ou groupes de mots, contenus dans la réponse générée versus une référence. Prenons un exemple de comparaison de n-grams classiques. Supposons que le modèle génère la phrase suivante : « je je je je je » et qu'une des références est « je ne mange pas de pommes ». Une comparaison par n-grams classique ne fera que regarder si chacun des mots dans la génération apparaît dans une référence et produit un score. Ici le score unigram serait de 5/5, car le mot « je » apparaît dans la référence. On voit donc assez vite les limites de cette métrique : l'ordre n'est pas pris en compte et on peut facilement construire une réponse qui ne fait aucun sens, mais qui aura

un score parfait. Ce que la métrique BLEU propose est de limiter le nombre de fois que nous comptons un mot, ou groupe de mots, à son nombre d'apparitions maximales dans l'une des références. En reprenant le même exemple, le mot « je » apparaît un maximum d'une fois dans la référence, on peut donc le compter qu'une fois pour calculer le score BLEU qui sera de $1/5$ cette fois-ci. Le score BLEU peut être calculé pour des n-grams avec des n supérieurs à 1, ce qui a comme effet de contraindre davantage les réponses générées à ressembler aux références. Cette métrique concorde difficilement avec le jugement humain [227, 228], car elle ne prend pas en compte le contexte syntaxique lorsque n est petit et est trop rigide si n est grand. En effet, en utilisant des n-grams avec $n=4$, communément utilisé comme n maximal avec BLEU, si les mêmes mots se retrouvent dans la référence et suivent la même structure, mais qu'un autre mot se trouve à « casser » la séquence, BLEU ne le détectera pas. De plus, elle ne prend pas en compte les différentes manières valides d'écrire une phrase, car si le mot n'est pas présent dans une des références et qu'un synonyme est employé à la place, il ne sera pas détecté.

La métrique ROUGE fut introduite par Lin [225] pour automatiquement évaluer des résumés et traductions générés par machine. Cette métrique regroupe en fait cinq différentes versions : ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S et ROUGE-SU. Leurs scores sont calculés en utilisant la mesure-F. ROUGE-N est une superposition par n-grams classique. ROUGE-L fonctionne comme une comparaison par n-grams classique, mais de manière dynamique. ROUGE-L identifie la longueur maximale de mots séquentiels dans une réponse partagée avec les références pour produire un score. ROUGE-W est une amélioration de ROUGE-L qui cherche une séquence maximale dans la réponse générée qui se retrouve dans les candidats de référence, mais qui n'a pas besoin d'être séquentielle. Par contre, on applique un facteur de pénalité plus la séquence de la référence est clairsemée. ROUGE-S est une comparaison par n-grams, généralement avec $n=2$, avec la possibilité de sauter un mot dans une séquence. Cette modification permet une certaine flexibilité en permettant que des mots autres que ceux retrouvés dans la réponse générée soient présents entre les mots d'une séquence comparative. Prenons par exemple l'extrait généré suivant : « Je mange une pomme en ce moment ». Supposons que l'une des références est la suivante : « Je voudrais une pomme ». Les bi-grams sont alors « Je mange », « Je une », « Je pomme », « Je en », etc. On compile ensuite le score avec ces nouvelles paires de mots. ROUGE-SU est une extension de ROUGE-S, mais on lui ajoute l'obligation de compter des unigrams aussi afin de couvrir les séquences qui ne sont pas dans le même ordre dans la référence et la réponse générée. Tout comme la métrique BLEU, ROUGE ne détecte pas les phrases ayant des synonymes comme étant valides et dépendamment de la version que nous utilisons, la métrique peut être très sensible à l'ordre des mots.

La métrique METEOR fut introduite par Denkowski et Lavie [226]. Cette métrique pallie aux problèmes de BLEU, notamment en pouvant détecter des mots qui ne sont pas dans le même ordre dans la réponse générée et les références ou même détecter des synonymes. METEOR effectue une cartographie un à un des unigrams, réduits à leurs racines, de la réponse générée et des références puis calcule une moyenne harmonique en privilégiant le rappel versus la précision d'un facteur de neuf. Si la racine exacte n'est pas détectée, METEOR se sert alors de la base de données de synonymes de « WordNet » [229] pour essayer de détecter un mot de la réponse dans une des références qui pourrait avoir employé un synonyme à la place. L'un des désavantages et avantages de cette méthode est que l'ordre n'est pas important lors de la comparaison d'unigrams et qu'un unigram ne peut être cartographié qu'à un seul unigram dans la référence. On ne peut donc pas compter un unigram plus d'une fois lors du calcul du score. Le désavantage est qu'une réponse générée qui ne fait aucun sens grammaticalement, mais qui contient tous les mots d'une référence aura un pointage parfait vu que l'ordre n'est pas pris en compte au sens strict. Un autre avantage est le fait que cette métrique puisse détecter les synonymes, ce qui ne pénalise pas les réponses générées qui les utilisent.

2.5.2 Les métriques basées sur la distance des mots similaires

Les métriques basées sur la distance des mots similaires font une comparaison des phrases ou mots vectorisés d'une phrase générée avec une ou des références pour en mesurer la ressemblance. Les techniques les plus connues sont le Greedy Matching [230], la moyenne des vectorisations [231, 232], l'extrême vectoriel [233], le Skip-Thought [234] et le BERTScore [235].

Le Greedy Matching est introduit par Rus et Lintean [230] afin de comparer les mots vectorisés d'une réponse générée à ceux d'une référence. Elle consiste à faire une somme des similarités cosinus maximales des mots de la réponse et de la référence, un mot à la fois, et puis de diviser par la cardinalité de l'extrait généré. C'est donc une moyenne des mots vectoriellement similaires. Cette approche ne regarde pas l'ordre des mots, ce qui est un inconvénient. L'avantage par contre est que les mots encodés similaires vont être pris en compte dans le score sans avoir à explicitement mentionner de quel mot il s'agit. Donc les mots qui ont une fonction syntaxique analogue dans la référence et la réponse générée auront une similarité cosinus forte ce qui équivaut à avoir des mots homologues dans la référence et le texte généré.

La moyenne des vectorisations (Embedding Average en anglais) est une métrique qui fût introduite par Wieting et coll. [231] pour des tâches de comparaison de textes générés. Cette méthode consiste à calculer une moyenne des vectorisations de chaque mot dans la réponse

générée et d'en calculer la similarité cosinus avec la moyenne des vectorisations de chaque mot dans la référence. On fait donc une somme des vecteurs de chaque mot divisé par le nombre de vecteurs contenus dans la phrase, et ce pour chaque mot. Plus les réponses générées sont similaires en moyenne aux références, plus le score se rapprochera de 1, dans le cas contraire le score se rapprochera de -1. Cette métrique ne regarde que si la moyenne vectorielle des mots dans une phrase ressemble à une autre, on ne tient donc pas compte des mots individuels ni de la structure de la phrase. L'avantage est que si l'ensemble des constituants d'une phrase occupent en moyenne la même fonction que les mots d'une phrase de comparaison, ces deux extraits seront reconnus comme équivalents.

L'extrême vectoriel [233] est une approche qui compare la réponse générée à des références au niveau de la phrase encodée. L'idée ici est d'encoder les mots de chaque phrase et puis de prendre pour chaque dimension vectorielle la valeur maximale associée à chaque mot pour construire un vecteur d'extrêmes qui représentent la phrase. Une fois cet encodage effectué pour la réponse et la référence, il s'agit d'effectuer une similarité cosinus pour en déterminer la similitude. L'idée derrière cette approche est que prendre les maximums de chaque dimension du vecteur mot priorise les mots contenant beaucoup d'information, qui sont donc loin de l'origine. On effectue donc effectivement qu'une comparaison vectorielle entre les mots les plus influents de la réponse générée et la référence.

Le Skip-Thought introduit par Kiros et coll. [234] est une autre métrique qui compare une réponse générée à une référence. Les auteurs présentent une nouvelle manière de créer un modèle capable de mieux encoder des phrases. En effet, au lieu d'uniquement se servir des mots précédents pour essayer de trouver une représentation vectorielle adéquate aux mots à la « Word2Vec » [119], les auteurs se servent de la phrase précédente et suivante pour entraîner un modèle capable d'encoder une phrase centrée entre les deux. Plus précisément, le modèle essaie de prédire la phrase précédente et suivante à l'aide d'un décodeur qui prend en entrée la représentation vectorielle de la phrase du milieu encodée. L'idée est de pouvoir généraliser au mieux possible les représentations vectorielles des phrases dans leurs contextes. La métrique consiste alors à utiliser l'encodeur, une fois entraîné, pour faire calculer une similarité cosinus entre la phrase générée et la référence.

Une dernière métrique intéressante est celle de Zhang et coll. intitulé BERTScore [235]. Cette métrique se sert du modèle BERT introduit par Devlin et coll. [167] pour faire une cartographie des mots vectorisés de la phrase générée avec ceux de la référence. Ici on ne cherche pas à créer un couple d'un mot avec un autre mot similaire dans les deux extraits, mais on fait une sorte de matrice de maximums de similarités un-versus-tous, c'est-à-dire que chaque mot vectorisé de la réponse est comparé à tous les autres de la référence pour découvrir

ceux qui ont une similarité maximale. Une fois les paires identifiés, le facteur de similitude est pondéré par la fréquence d'utilisation de chaque mot puis on somme pour tous les mots et on normalise par la somme des facteurs de pondération pour donner un score en -1 et 1 . Comme BERT a été entraîné sur plusieurs millions d'extraits textuels et que le modèle en soit comporte plusieurs millions de paramètres, il représente assez bien statistiquement l'utilisation contextuelle des mots et comme nous maximisons automatiquement la similarité vectorielle entre les phrases, il produit des scores qui sont du même ordre que des scores donnés par des humains [167]. L'avantage est que par sa construction, BERT est capable d'adéquatement représenter les mots de manière vectorielle dans leur contexte, on tient donc en compte l'ordre même si ce n'est pas parfait, et par sa construction, des phrases similairement construites qui utilisent des mots différents, pas nécessairement dans le même ordre, seront reconnues comme étant équivalentes.

2.5.3 Les métriques libres de références

Toutes les métriques présentées jusqu'à maintenant ont une chose en commun, l'utilisation de références est nécessaire afin de pouvoir évaluer les modèles générateurs. Dans des tâches de génération de dialogue, il est difficile d'obtenir des références adéquates auxquelles comparer les réponses générées dues à la nature très ouverte de la tâche [236]. De plus, il n'y a pas de critères fixes pour décrire un bon dialogue [237], ces critères dépendent généralement du contexte spécifique d'implémentation des différents modèles générateurs, les qualités recherchées peuvent alors varier grandement [238]. Pour essayer de pallier à ces problèmes, des métriques dites libres de références ont été introduites. Ces métriques ne nécessitent pas de référence pour quantifier la qualité d'un dialogue ou d'un extrait généré. Elles sont généralement construites afin d'identifier et évaluer plusieurs sous-aspects d'un dialogue [223] ce qui en fait des métriques versatiles. Parmi les métriques d'intérêt on retrouve *USR* et *FED* de Mehri et Eskenazi [223, 239] et *RUBER* de Tao et coll. [240].

La métrique *USR* introduite par Mehri et Eskenazi [223] est composée de cinq sous-métriques qui évaluent différentes caractéristiques d'une conversation. Les cinq aspects évalués sont la clarté, le naturel de la réplique, le maintien du contexte, l'originalité et l'utilisation de connaissances du dialogue. La combinaison de ces cinq sous-métriques attribue un score *USR* global à la réplique générée. Les auteurs se sont basés sur le modèle préentraîné *RoBERTa* [170] pour évaluer la vraisemblance d'une réponse générée sachant son contexte conversationnel. Suite à l'affinement du modèle *RoBERTa* sur le corpus « Topical-Chat » [241] et/ou sur le corpus « PersonaChat » [242] les auteurs réussissent à évaluer automatiquement la clarté et le naturel d'une réplique générée dans un contexte conversationnel. Les auteurs ont par la suite

affiné le modèle RoBERTa pour qu'il soit capable de capturer des éléments clés d'une conversation, tel que les noms propres, certaines informations spécifiques afin de pouvoir détecter si une réponse générée incorpore des éléments soulevés durant l'historique du dialogue. Cet affinement permet d'évaluer le maintien du contexte dialogal, l'utilisation des connaissances et l'originalité.

La métrique FED introduite aussi par Mehri et Eskenazi [239] garde la trace de 18 différentes caractéristiques dialogales afin d'évaluer automatiquement une conversation et ses échanges. Les auteurs utilisent le modèle préentraîné DialoGPT [243] pour essayer de capturer les connaissances implicites acquises par ce modèle afin de construire leur métrique. Les auteurs ont d'abord construit pour chaque caractéristique dialogale un ensemble de réponses dites positives et un set de réponses négatives qui concordent avec la caractéristique du dialogue à détecter. Le modèle DialoGPT est alors affiné pour pouvoir détecter les caractéristiques syntaxiques attribuées à chaque caractéristique conversationnelle. Par la suite les auteurs ont construit une fonction qui compare la réponse générée du système à être évalué, pour un contexte donné, à une réponse générée par DialoGPT, pour ce même contexte, et notent la concordance positive et négative des deux réponses pour chaque caractéristique recherchée. Afin d'obtenir un score global, les auteurs calculent une moyenne des scores associés à chaque caractéristique.

La métrique RUBER introduite par Tao et coll. [240] est une métrique qui utilise à la fois un mécanisme d'évaluation des phrases encodées basée sur des références et un mécanisme d'évaluation libre de références. Le premier mécanisme mesure la similarité entre une réplique générée et une référence associée par la méthode de l'extrême vectoriel [233]. C'est-à-dire qu'on calcule une similarité cosinus entre les valeurs extrêmes des vecteurs mots de la réplique générée et la réplique de référence afin d'en calculer la similitude et par le fait même évaluer la qualité de la réplique générée. Le deuxième mécanisme, libre de référence, mesure la similarité entre la réplique générée encodée et la requête (réplique) conversationnelle précédente. Un score combiné est alors calculé en combinant les résultats normalisés des deux mécanismes à l'aide d'une heuristique.

CHAPITRE 3 L'INJECTION D'ÉMOTION : PHILIA

Dans ce chapitre, nous présentons les mécanismes du système que nous avons baptisé *Philia* ainsi que certains détails pratiques pour une véritable implémentation de celui-ci. *Philia* est un système qui vient se greffer sur un ou plusieurs agents conversationnels et qui permet d'injecter de l'émotion en filtrant et modifiant la ou les répliques données par les agents. Dans un premier temps et dans le cadre de ce mémoire, l'injection d'émotion se traduit par la substitution de certains mots par des synonymes ou des antonymes dépendamment de l'effet émotif recherché.

Nous combinons des approches et algorithmes de la recherche opérationnelle et de l'apprentissage automatique. L'algorithme principal est un apprentissage renforcé qui repose sur la combinaison de cinq blocs qui seront chacun présentés de manière générique et puis de manière plus détaillée avec des considérations d'ordre pratique pour une première implémentation. Si les idées de base sont relativement simples, l'implémentation est parfois délicate et des raffinements pratiques sont nécessaires pour développer un système cohérent et probant.

Le fonctionnement global du système consiste à récolter différentes répliques générées par un ou plusieurs agents conversationnels distincts suite à une requête, puis à les analyser syntaxiquement, sentimentalement (analyse des sentiments) et émotionnellement (analyse des émotions) pour construire un modèle d'optimisation qui permet de retourner une réplique émotionnelle optimisée suivant une émotion cible dans un contexte donné.

Cette nouvelle réplique est alors analysée par deux boucles rétroactives - une automatique et une autre humaine - qui permettent à *Philia* d'apprendre comment affiner ses paramètres internes pour mieux injecter de l'émotion. La première boucle est essentiellement une analyse émotionnelle automatique qui consiste en un classificateur affectif utilisant de l'apprentissage profond et qui détermine un score associé à la consonance ou dissonance émotive entre l'utilisateur et *Philia*. La deuxième boucle consiste à directement demander à l'utilisateur humain de retourner un score selon sa perception émotive de la réplique et de la pertinence des synonymes ou antonymes employés.

Dans ce chapitre, nous détaillons les cinq blocs constitutifs de *Philia* et décrivons l'architecture générale de *Philia*. Nous commençons par le *Bloc 1 : Les agents conversationnels* à

la section 3.1 où nous décrivons comment Philia collecte les répliques des agents conversationnels et quels agents conversationnels ont été utilisés, puis le *Bloc 2 : Compréhension du langage naturel* à la section 3.2 qui décrit comment Philia analyse syntaxiquement et émotionnellement les répliques collectés précédemment. Le *Bloc 3 : Vocabulaire* à la section 3.3 décrit les deux bases de données utilisées dans ce projet : une base de données de synonymes et antonymes et une base de données des scores émotifs utilisés. Ces scores sont utilisés dans la construction de notre modèle d’optimisation décrit dans le *Bloc 4 : Optimisation* à la section 3.4. Une fois la réplique de Philia transmise à l’utilisateur, le *Bloc 5 : Centre de commande* à la section 3.5 analyse la concordance ou dissonance émotionnelle entre Philia et l’utilisateur. Finalement, nous présentons l’architecture globale de Philia et comment chacun des blocs s’imbrique dans le système dans le *Fonctionnement de Philia* à la section 3.6.

3.1 Bloc 1 : Les agents conversationnels

Philia est un système qui se greffe à des agents conversationnels existants. Le *Bloc 1 : Les agents conversationnels* est considéré comme une boîte noire qui génère des répliques pour chaque réplique de l’utilisateur. Dans le cadre de ce projet, nous avons décidé d’implémenter des agents conversationnels basés sur des transformateurs pré-entraînés sur une vaste quantité de données ouvertes et d’utiliser un agent en ligne. Trois différents agents conversationnels sont utilisés : un agent conversationnel basé sur Blenderbot de Facebook introduit par Ng et coll. [244], un agent conversationnel basé sur DialoGPT de Microsoft introduit par Zhang et coll. [243] et un agent conversationnel en ligne appelé Mitsuku¹. Pour la combinaison de ces trois agents conversationnels, nous avons procédé de la manière suivante. Lorsque le système récupère la réplique de l’usager, celle-ci est envoyée à tous les agents conversationnels séparément et on récupère dans une liste l’ensemble des répliques générées par ceux-ci. Cette liste est alors importée dans un premier temps dans le *Bloc 2 : Compréhension du langage naturel* afin d’être pré-traitée et elle est utilisée par la suite par le *Bloc 4 : Optimisation* qui retournera une réplique dont l’émotion sera optimisée. Une fois l’optimisation et la reconstruction de la réplique effectuées, la nouvelle réplique est envoyée à l’usager pour continuer la conversation et ainsi de suite jusqu’à la terminaison de cette dernière. Tout au long de la conversation, des sauvegardes partielles du dialogue sont effectuées afin de nous permettre d’analyser les performances du système et de faire une étude comparative.

1. Mitsuku : <https://chat.kuki.ai/>

3.2 Bloc 2 : Compréhension du langage naturel

Le *Bloc 2 : Compréhension du langage naturel* occupe une place prépondérante dans ce projet, car il permet à la fois de décomposer chaque réplique en composantes syntaxiques, de déterminer les mots candidats susceptibles d'être remplacés, d'effectuer des analyses émotionnelle et sentimentale dont Philia se sert pour déterminer la concordance ou dissonance affective avec l'utilisateur, et finalement de reconstruire une réplique cohérente une fois l'optimisation de la réplique originale effectuée.

3.2.1 Décomposition des répliques

Pour pouvoir décomposer automatiquement les répliques obtenues des agents conversationnels, nous utilisons la librairie spaCy [245] introduite par Hannibal et Montani. SpaCy est un outil puissant qui est capable entre autres de détecter automatiquement la racine des mots d'une phrase donnée en entrée, la classe syntaxique à laquelle appartient chaque mot, la dépendance entre les mots ainsi que de déterminer les mots d'arrêt (stopword). Dans le cadre de ce projet, nous n'utilisons spaCy que pour la décomposition syntaxique des répliques. Nous avons choisi de travailler avec la racine des mots pour réduire le nombre de tuples possibles dans un premier temps. Nous tenons cependant compte du rôle des mots dans les phrases. Ainsi nous distinguons les noms communs des verbes, etc. Chacune racine est donc accompagnée d'une classe syntaxique et nous parlerons dorénavant de *racine-classe* ou *tuple racine-classe*². Le but de cette étape est de créer une liste de tuples racines-classes afin d'identifier les mots que le système peut modifier pour maximiser le score affectif interne de la réplique générée par Philia. C'est aussi à cette étape que nous vérifions l'existence des racines-classes dans notre base de données vocabulaire. Si l'entrée est inexistante, nous créons une nouvelle entrée initialisée par défaut. Pour mieux comprendre l'étape de décomposition, nous référons le lecteur à la figure 3.1. Dans un premier temps et dans le cadre de ce mémoire, nous nous concentrons sur le remplacement de mots uniques par des synonymes. Dans un deuxième temps, nous pourrions envisager de remplacer des sous-phrases ou bien des combinaisons de mots par d'autres sous-phrases ou combinaisons de mots. Nous détaillons la reconstruction de la réplique dans la section 3.4.4.

2. Dans le cadre de ce projet, nous ne considérons que des couples (racines, POS), c'est-à-dire un couple formé de la racine du mot ainsi que son étiquette syntaxique (Part-of-speech tag). Dans un développement futur, nous pourrions utiliser des tuples plus détaillés.

	text	lemma (racine)	POS (classe)	explain	stopword
0	This	this	DET	determiner	True
1	is	be	AUX	auxiliary	True
2	a	a	DET	determiner	True
3	test	test	NOUN	noun	False
4	sentence	sentence	NOUN	noun	False
5	to	to	PART	particle	True
6	show	show	VERB	verb	True
7	the	the	DET	determiner	True
8	spaCy	spacy	NOUN	noun	False
9	decomposition	decomposition	NOUN	noun	False
10	we	-PRON-	PRON	pronoun	True
11	perform	perform	VERB	verb	False
12	for	for	ADP	adposition	True
13	Philia	Philia	PROPN	proper noun	False
14	.	.	PUNCT	punctuation	False

FIGURE 3.1 Exemple de décomposition que le système effectue avec spaCy. La phrase donnée en entrée est "This is a test sentence to show the spaCy decomposition we perform for Philia."

3.2.2 Analyses sentimentale et émotionnelle

Une composante importante de ce bloc est l'analyse sentimentale et émotionnelle des différentes répliques, que ce soit pour les phrases originales générées par les agents conversationnels, la nouvelle réplique optimisée ou la réplique de l'utilisateur. La nuance entre l'analyse sentimentale et émotionnelle est que l'analyse sentimentale sert à détecter un état mental (positif, négatif, neutre) intimement lié aux émotions, mais sans les catégoriser explicitement, tandis que l'analyse émotionnelle sert à détecter les émotions en soi. Le système utilise les analyses sentimentale et émotionnelle de plusieurs façons. Une première utilisation des analyses sentimentale et émotionnelle est de permettre à la boucle rétroactive d'ajuster les scores émotifs de chaque tuple racine-classe dans son contexte dialogal. C'est un des mécanismes d'apprentissage par renforcement du système. Nous détaillons le fonctionnement de

la boucle de rétroaction automatique dans la section 3.5.1. Une autre utilisation est de permettre de mesurer la performance de Philia en mesurant la différence émotive - concordance on dissonance émotive - perçue pour chaque interaction agent-utilisateur. Cette évaluation est détaillée dans la section 3.5.1.

Le classificateur émotionnel est basé sur l'implémentation de Demszky et coll. [56] à l'aide de la librairie Pytorch. Les auteurs ont créé un classificateur émotionnel/sentimental en choisissant et annotant à la main plus de 58 000 commentaires collectés sur Reddit. Ils ont identifié initialement 27 émotions différentes, plus la neutralité. De plus, en regroupant les émotions selon des clusters émotionnels suivant le modèle d'Ekman [48], ils ont aussi créé un classificateur pour ces mêmes émotions (la peur, la colère, la joie, la tristesse, le dégoût et la surprise). C'est cette dernière version que nous utilisons pour classifier émotionnellement nos différents extraits textuels. Les auteurs ont aussi annoté leurs données pour effectuer une analyse sentimentale que nous exploiterons également dans le cadre de ce projet. Ces classificateurs basés sur le modèle transformateur préentraîné BERT [167] sont affinés par la suite sur les données récoltées par les auteurs.

3.2.3 Reconstruction de la réplique optimale

La dernière composante importante du *Bloc 2 : Compréhension du langage naturel* est la reconstruction des phrases. Notre système fonctionne en interchangeant les tuples racine-classe des mots candidats par des tuples racines-classes qui représentent des synonymes ou des antonymes dépendamment de l'émotion que l'on veut éliciter. En nous assurant de prendre un nouveau synonyme ou antonyme dont la classe syntaxique est identique à celle du mot que nous voulons modifier, nous nous assurons de garder la même structure syntaxique de la phrase initiale, ce qui réduit la possibilité de modifier de manière incorrecte la syntaxe et le sens de la réplique initiale. Pour reconstruire une phrase, Philia identifie les tuples racine-classe retenus par le *Bloc 4 : Optimisation*, note leurs emplacements à l'intérieur de la phrase, puis remplace les racines-classes par les mots associés. Il faut maintenant conjuguer les mots remplacés. Le système utilise un correcteur générateur appelé Gramformer³ pour s'assurer de la bonne conjugaison.

3. Gramformer : <https://github.com/PrithivirajDamodaran/Gramformer>. Nous avons testé plusieurs correcteurs automatiques pour corriger nos répliques et selon nos observations, celui-ci était de loin le plus performant et produit des corrections généralement satisfaisantes.

3.3 Bloc 3 : Vocabulaire

Le *Bloc 3 : Vocabulaire* de ce projet est composé de deux parties : une base de données *vocabulaire* contenant plus de 58 000 mots, leurs synonymes et leurs antonymes, et une base de données de *scores* contenant l'information pertinente sur les tuples racine-classe employés lors de l'optimisation et la reconstruction des phrases.

3.3.1 Base de données des synonymes et antonymes

La première base de données contenant l'information sur les mots, leurs synonymes et leurs antonymes a été construite en se servant d'une liste de plus de 58 000 mots récupérée du site *mieliestronk*⁴. Pour étoffer notre base de données, nous nous sommes servis d'un module de grattage du web qui, pour chaque mot de la liste, récupère ses synonymes et antonymes, sous forme de liste, sur sept différents sites internet qui collectent ces informations⁵. Si aucun synonyme ou antonyme n'a pu être trouvé, la liste est alors laissée vide. Comme les différents sites peuvent proposer des synonymes et antonymes différents, nous avons procédé à un nettoyage de nos données comme suit. Dans un premier temps, nous retirons toutes les entrées des listes qui sont formées de deux mots ou plus, car lors de l'optimisation le système échange certains mots par leurs synonymes ou antonymes. L'utilisation de deux mots (ou plus) comme synonymes ou antonymes n'est pas, dans un premier temps, considérée car elle rend la mise à jour des scores qui portent sur des tuples racines-classes individuels délicate. Nous retirons donc de telles combinaisons. Dans un deuxième temps, nous éliminons tous les synonymes et antonymes qui peuvent être décomposés par *spaCy* en mots d'arrêt. Les mots d'arrêt sont des mots qui, dans les approches traditionnelles en apprentissage automatique, sont exclus des analyses, car ils sont considérés comme n'apportant aucune information pertinente. Sans vouloir entrer dans le débat de la pertinence des mots d'arrêt, nous les excluons aussi, car la plupart des librairies que nous utilisons les excluent. Pour les nouveaux mots qui ne sont pas répertoriés dans la base de données, nous avons développé un mécanisme de recherche sur internet pour récupérer les synonymes et antonymes associés. De cette manière nous pouvons continuellement enrichir notre base de données.

4. liste de plus de 58 000 mots, en anglais, tiré du site *Mieliestronk* : <http://www.mieliestronk.com/wordlist.html>

5. Sites sources pour l'obtention de synonymes et antonymes : classicthesaurus.com, collinsdictionary.com, merriam-webster.com, synonym.com, thesaurus.com, wordhippo.com, wordnet.princeton.edu

3.3.2 Base de données des scores

La deuxième base de données est celle qui garde en mémoire les différentes informations pertinentes sur les tuples racine-classe des mots employés et leurs scores affectifs utilisés lors de l’optimisation. Cette base de données est hiérarchisée en fonction des émotions cibles, dans ce cas-ci, ce sont les six émotions de base d’Ekman [48], et du contexte actuel de la conversation. Pour chaque tuple racine-classe, nous pouvons avoir différents scores dépendamment de l’émotion cible et du contexte actuel.

La grandeur de cette base de données, c’est-à-dire le nombre d’entrées, est majorée par :

$$\#entrées \leq \#émotions \cdot \#contextes \cdot \#classes \cdot \#racines \quad (3.1)$$

où $\#entrées$ est le nombre d’entrées total, $\#émotions$ est le nombre d’émotions considérées, $\#contextes$ le nombre de contextes considérés, $\#classes$ est le nombre de classes syntaxiques auxquelles les racines peuvent appartenir et $\#racines$ les racines des mots du vocabulaire considéré. Les contextes sont des identifiants permettant de se situer dans le développement d’une conversation. Un exemple de contexte serait la situation d’une salutation entre deux personnes. Il est à noter qu’un contexte pourrait aussi correspondre à un état dans la recherche ou correspondre à une variable interne du système. Les classes syntaxiques des mots sont des identifiants quant à leurs fonctions dans la phrase. Des exemples de classes syntaxiques sont les verbes et leurs conjugaisons, les noms, les adjectifs, etc. Comme on le voit, la construction de cette base de données peut être très lourde et l’entraînement adéquat s’avère être un défi de taille sans les ressources nécessaires. C’est pour cette raison que, dans le cadre de ce projet, nous nous sommes limités à une seule émotion cible et un seul contexte test afin de valider nos idées de base. La base de données elle-même a été construite comme une succession de « hash maps » imbriquées (des dictionnaires de dictionnaires en Python) de longueur arbitraire.

Chaque émotion cible contient de l’information sur les émotions dites similaires et opposées. Par exemple, la joie est associée à l’amusement ou la satisfaction, alors que la colère et le dégoût seront considérés comme des émotions opposées. L’idée de considérer des émotions similaires et opposées est de permettre une optimisation plus fine : le solveur doit à la fois maximiser l’émotion cible tout en respectant des bornes aussi bien pour les émotions similaires que les émotions opposées. Nous développons cette idée plus en détail dans la section 3.4.2.

Des considérations pratiques sont nécessaires pour encadrer le concept de score. Ces scores

doivent pouvoir être entraînés, doivent pouvoir être utilisés comme coefficients dans des modèles d’optimisation et donc pouvoir être comparables et contenir une information pertinente. L’implémentation proposée dans ce mémoire n’utilise pas directement les scores de la base de données dans les modèles, mais plutôt des coefficients calculés à partir de ces scores. Nous détaillons le calcul des coefficients utilisés dans la section 3.4.3.

Pour chaque émotion \mathcal{E} et contexte \mathcal{C} donnés, nous avons une sous-base de données correspondantes. Pour éviter qu’un score $c(t, \mathcal{C}, \mathcal{E}) \in \mathbb{N}$ correspondant à un tuple racine-classe t ne soit renforcé indéfiniment, nous le bornons entre deux valeurs θ_{\min} et θ_{\max} : $\theta_{\min} \leq c(t, \mathcal{C}, \mathcal{E}) \leq \theta_{\max}$ avec $\theta_{\min} < 0$ et $0 < \theta_{\max}$. Ces deux bornes peuvent être considérées comme des hyper-paramètres du système. Dans une première implémentation, nous avons considéré $-\theta_{\min} = \theta_{\max}$ et pris la grandeur moyenne des listes de synonymes par mot multipliée par deux comme valeur ce qui donne $\theta_{\max} = 50$. De cette façon, nous jugeons que le domaine des scores est assez large afin de permettre au système de facilement discriminer les mots à caractère émotif. Quand le contexte et l’émotion cible sont connus, nous simplifions la notation du score : $\xi_t = c(t, \mathcal{C}, \mathcal{E})$.

Un autre hyper-paramètre utilisé est un facteur de pondération τ qui permet de contrôler la vitesse à laquelle le système apprend les scores affectifs dans la phase rétroactive du système. Nous reviendrons sur ce facteur dans la section 3.5.1.

Nous gardons également en mémoire le nombre de fois qu’un tuple racine-classe fut utilisé par le système dans une réplique optimisée et le nombre de fois que le score individuel de la réplique optimisée le contenant est reconnu par notre scoreur individuel comme transmettant l’émotion cible. Le *scoreur individuel* permet d’évaluer la pertinence d’une réplique pour une émotion donnée. Il est détaillé à la section 3.5.1. De cette manière, nous pouvons calculer ce que nous appelons la *précision* p_t sur l’utilisation d’une classe-racine t qui correspond au ratio entre le nombre de fois où la réplique impliquant la racine-classe t est jugée pertinente par rapport à l’émotion cible et le nombre de fois où la racine-classe t est utilisée dans une réplique par Philia. Ce facteur est utilisé lors de la construction des coefficients dans le modèle d’optimisation.

3.3.3 Initialisation de la base de données des scores

La base de données des scores peut être initialisée de plusieurs manières. Citons, par exemple, l’initialisation préliminaire des scores à l’aide d’une base de données externe qui comporte

certaines informations sur la force émotive des mots, telle que la base de données *EmoWordNet* [246], l’initialisation par l’entraînement intensif de Philia sur une base de données dialogale telle que *Alexa Topical-Chat* [241] ou simplement en commençant avec des scores neutres. Dans le cadre de ce projet, nous avons choisi de prioriser les deux premières approches, car elles nous permettaient d’avoir un système fonctionnel rapidement et de pouvoir évaluer les améliorations apportées lors des transformations effectuées par Philia. La base de données *EmoWordNet* est constituée d’un peu plus de 60 000 racines de mots et leurs classes syntaxiques en anglais pour lesquels une probabilité d’appartenance aux émotions suivantes, la peur, l’amusement, la haine, l’agacement, l’indifférence, la joie, l’inspiration et la tristesse, a été compilée. Comme ces émotions ne concordent pas exactement avec les émotions que nous identifions à l’aide de notre classificateur émotif et dans le cadre restreint de ce mémoire, nous avons uniquement choisi de récolter les mots et leurs poids d’appartenance associés à l’émotion de la joie. Pour calculer le score des tuples (ξ_t) , nous avons choisi de pondérer les scores à 50 % de l’amplitude de la borne supérieure θ_{\max} allouée multipliée par la probabilité d’appartenance récoltée de la base de données *EmoWordNet* (p_{EWN}^t) :

$$\xi_t = p_{EWN}^t \cdot 0.5 \cdot \theta_{\max} \quad (3.2)$$

Nous avons choisi de transformer les probabilités d’appartenance de la base de données *EmoWordNet* en scores de la sorte, car en bornant les scores transformés à 50% de la borne supérieure θ_{\max} , nous laissons du jeu au système afin d’affiner l’information émotive pendant son utilisation.

L’initialisation de Philia à partir de la base de données dialogale *Alexa Topical-Chat* a été faite comme suit. *Alexa Topical-Chat* consiste en trois ensembles de conversations entre deux usagers, un ensemble d’entraînement, un ensemble de validation et un ensemble test. Pour entraîner le système, il faut passer à travers toutes les conversations de l’ensemble d’entraînement en indiquant qu’un des acteurs de la conversation est l’usager et l’autre est le système, puis mettre à jour les scores à chaque interaction. Une faiblesse de cette méthode est que les répliques de l’acteur considéré comme l’usager ne changent pas en fonction des répliques transformées du système. Nous supposons que les répliques optimisées sont suffisamment similaires aux répliques originales et que les acteurs usagers répondent de manière adéquate au système. Notons que l’initialisation de la base de données des scores par *Alexa Topical-Chat* nécessite une attention toute particulière quant aux paramètres associés à la *variable émotionnelle interne*. Cette variable émotionnelle interne est définie plus loin et correspond globalement à une évaluation de la concordance ou dissonance émotionnelle entre l’utilisateur humain et Philia. Nous développons ce concept et explicitons la démarche à suivre dans

la section 3.5.1. Il est également possible de combiner les deux approches d’initialisation en entraînant Philia sur *Alexa Topical-Chat* une fois l’initialisation primaire effectuée à l’aide des données de *EmoWordNet*. Dans le cadre de ce projet, nous avons décidé d’initialiser le système en combinant les deux méthodes.

3.4 Bloc 4 : Optimisation

Pour illustrer l’optimisation des émotions, prenons un exemple de base où Philia remplace simplement certains mots par des mots plus emphatiques et enjoués. Pour effectuer une telle transformation, Philia puise dans une liste de synonymes. Pour rester simple, nous ne permettons à Philia que de changer les verbes, adverbes et adjectifs un par un. Des possibilités plus sophistiquées sont souhaitables, mais les principes sont similaires.

Le *Bloc 4 : Optimisation* est la partie du système qui s’occupe d’optimiser les choix des tuples racine-classe pour toutes les répliques données en entrée par les agents conversationnels. Le but est de choisir la réplique et les tuples racines-classes qui maximisent un objectif émotif. Dans cette section, nous présentons trois modèles : un premier modèle minimaliste suivi par un modèle plus sophistiqué et enfin le modèle implémenté dans notre première approche. Nous modélisons le problème en nombres entiers à l’aide du solveur CP-SAT de la librairie OR-Tools [247].

3.4.1 Un premier modèle minimaliste

Pour améliorer le processus d’optimisation, nous rassemblons plusieurs répliques possibles $i \in \{1, \dots, m\}$ retournées par le *Bloc 1 : Les agents conversationnels*. Pour chaque réplique $i \in \{1, \dots, m\}$, nous construisons une liste de p synonymes que nous indiquons par $k \in \{1, \dots, p\}$ (où $k = 1$ correspond au mot original donné dans la réplique du chatbot) pour chaque mot modifiable j dans la réplique i . Soit la variable

$$x_{ij}^k = \begin{cases} 1 & \text{si pour le } j^{\text{ième}} \text{ mot de la } i^{\text{ième}} \text{ réplique, on choisit le } k^{\text{ième}} \text{ synonyme.} \\ 0 & \text{sinon.} \end{cases} \quad (3.3)$$

Dans un premier temps, nous ne permettons pas de mélanger différentes répliques et n’acceptons donc qu’une seule réplique i . On définit

$$y_i = \begin{cases} 1 & \text{si la réplique } i \text{ est choisie.} \\ 0 & \text{sinon.} \end{cases} \quad (3.4)$$

Nous pouvons maintenant modéliser le problème avec les contraintes suivantes :

$$x_{ij}^k \leq y_i \quad \forall i, j, k \quad (3.5a)$$

$$\sum_i y_i \leq 1 \quad \forall i \quad (3.5b)$$

$$\sum_k x_{ij}^k \leq 1 \quad \forall i, j \quad (3.5c)$$

$$y_i \leq \sum_k x_{ij}^k \quad \forall i, j \quad (3.5d)$$

(3.5a) implique que si un tuple racine-classe x_{ij}^k est choisi, la réplique correspondante y_i doit aussi être choisie et vice-versa si une réplique y_i n'est pas choisie, aucun des mots correspondants ne peut être choisi. (3.5b) interdit plus d'une réplique, (3.5c) implique que seul un synonyme peut être utilisé à la fois alors que (3.5d) complémente (3.5a) dans l'autre sens.

Nous utilisons des scores émotifs pour pouvoir optimiser une émotion cible dans un contexte donné. Comme nous le verrons dans la section 3.4.3, le calcul de ces coefficients à partir des scores ξ_t demande une certaine précaution. Ici, nous utilisons un concept générique de score émotif noté $c(\mathcal{C}, \mathcal{E})$ suivant une émotion cible \mathcal{E} et un contexte \mathcal{C} donnés.

Notre premier modèle minimaliste est le suivant :

$$\begin{aligned} \max_x \quad & \sum_{i,j,k} c_{ij}^k(\mathcal{C}, \mathcal{E}) x_{ij}^k \\ \text{s.t.} \quad & x_{ij}^k \leq y_i \quad \forall i, j, k \\ & \sum_i y_i \leq 1 \quad \forall i \\ & \sum_k x_{ij}^k \leq 1 \quad \forall i, j \\ & y_i \leq \sum_k x_{ij}^k \quad \forall i, j \\ & x_{ij}^k \in \{0, 1\} \quad \forall i, j, k \\ & y_i \in \{0, 1\} \quad \forall i \end{aligned} \quad (3.6)$$

Notons que ce modèle est en fait une famille de modèles paramétrisés par des coefficients/scores.

3.4.2 Un deuxième modèle plus sophistiqué

Un problème récurrent dans les approches basées uniquement sur l'apprentissage automatique est que les répliques des agents conversationnels ne peuvent pas être facilement contrôlées [13]. La recherche opérationnelle utilise quant à elle des techniques qui permettent un contrôle strict ou partiel par exemple avec l'utilisation de contraintes. Nous pouvons raffiner notre modèle en exigeant que certaines bornes - aussi bien pour les émotions similaires que pour les émotions contraires - soient respectées.

Soient $\text{sim}(\mathcal{E}) = \{\mathcal{E}_1 = \mathcal{E}, \mathcal{E}_2, \dots\}$ un ensemble d'émotions similaires à l'émotion cible et $\text{con}(\mathcal{E}) = \{\mathcal{F}_1, \mathcal{F}_2, \dots\}$ un ensemble d'émotions opposées, l'émotion cible étant \mathcal{E} . Si $\mathcal{E} = \text{joie}$, alors $\text{sim}(\mathcal{E}) = \{\text{joie}, \text{surprise}, \dots\}$ and $\text{con}(\mathcal{E}) = \{\text{haine}, \text{peur}, \dots\}$. Pour renforcer l'empathie, nous pouvons exiger que certaines bornes soient respectées pour les émotions similaires :

$$c(\mathcal{C}, \mathcal{E}_k)^T x \geq \theta(\mathcal{C}, \mathcal{E}_i) \quad \forall \mathcal{E}_i \in \text{sim}(\mathcal{E}) \quad (3.7)$$

avec θ , un hyper-paramètre réel, et où nous avons remplacé la somme sur les i, j, k par le produit scalaire sur des vecteurs. Quand $\mathcal{E}_i = \mathcal{E}$, Philia ne remplacera une réplique que si la solution optimisée respecte un certain seuil.

De manière similaire, nous pouvons borner les émotions opposées :

$$c(\mathcal{C}, \mathcal{E}_k)^T x \leq \theta(\mathcal{C}, \mathcal{F}_i) \quad \forall \mathcal{F}_i \in \text{con}(\mathcal{E}) \quad (3.8)$$

Les contraintes (3.7) et (3.8) permettent de jouer avec les affinités entre émotions similaires et opposées en ajustant les paramètres $\theta(\mathcal{C}, \mathcal{E}_i)$ et $\theta(\mathcal{C}, \mathcal{F}_i)$ en conséquence.

Le modèle (3.6) combiné avec les contraintes (3.7) et (3.8) constitue notre modèle plus sophistiqué. Remarquons comment l'ajout de ces deux contraintes supplémentaires nous permet de trouver des solutions réalisables sans nécessairement rechercher des solutions optimales.

Dans le cadre de ce mémoire, nous n'avons pas implémenté le modèle plus sophistiqué, mais plutôt une version concrète du modèle minimaliste que nous présentons dans la prochaine section.

3.4.3 Le modèle implémenté

Nous présentons dans cette section le modèle que nous avons implémenté. Il s'agit du modèle minimaliste défini à la section 3.4.1. Pour avoir un système cohérent qui fonctionne, nous avons apporté quelques modifications ou défini certaines opérations. Les trois ajouts les plus importants sont l'utilisation d'une *variable émotionnelle interne* V qui permet de guider le type de recherche (phases d'exploration ou de renforcement) pendant les dialogues, le calcul des coefficients utilisés dans le modèle d'optimisation et qui s'adaptent aux différents types de recherches de Philia, et finalement la mise à jour des scores qui demande une certaine prudence. Cette implémentation est un essai parmi d'autres implémentations possibles. Nous justifions la plupart de nos choix dans ce chapitre et plus encore dans le chapitre suivant concernant les résultats obtenus.

La variable émotionnelle interne V reflète la dissonance ou consonance émotionnelle entre l'utilisateur humain et Phillia. Dans un contexte et avec une émotion cible qui privilégie la concordance, plus la concordance est élevée, plus Philia peut se permettre d'être en mode exploratoire et d'essayer de nouvelles combinaisons. Par contre, si la concordance est faible, Philia essaiera d'utiliser son apprentissage pour ne pas sortir des sentiers battus et utilisera une phase de renforcement dans son apprentissage. Nous détaillons cette variable interne ainsi que sa relation avec un coefficient λ utilisé lors du calcul des coefficients du modèle dans la section 3.5.1. Pour le moment, retenons que V et λ sont directement proportionnels : $\lambda \propto V$.

Nous conservons exactement les mêmes variables et mêmes contraintes que celles utilisées dans le modèle minimaliste. Nous introduisons cependant deux différences significatives. La première est le calcul des coefficients et la deuxième est l'introduction d'un mécanisme pour filtrer les candidats tuples t_{ij}^k qui peuvent être utilisés pour remplacer les tuples t_{ij} . Présentons d'abord le calcul des coefficients avant de décrire le mécanisme de filtrage des candidats.

Calcul des coefficients

Les coefficients ν_{ij}^k correspondants aux variables x_{ij}^k sont calculés comme suit :

$$\nu_{ij}^k = \left\lfloor \frac{(\xi_k + \epsilon_{ijk}) \cdot p_k}{\frac{m}{\hat{m}}} \right\rfloor \quad (3.9)$$

La matrice ν_{ij}^k est composée des scores associés à l'émotion cible et le contexte donné ξ_k pour chaque tuple de remplacement k auxquels nous ajoutons un terme d'erreur aléatoire ϵ_{ijk}

utilisé pendant l’initialisation par entraînement intensif de Philia. Ce terme est inversement proportionnel à la durée de la phase d’initialisation. Le terme d’erreur est ajouté pour faciliter une sélection aléatoire des tuples de remplacement en début d’entraînement intensif pendant la phase d’initialisation de la base de données des scores à l’aide d’une base de données dialogale. En début d’entraînement, nous ne connaissons pas la réelle pertinence émotive des mots, nous voulons donc que le terme d’erreur soit non négligeable afin de permettre une sélection forcée des tuples. Le terme d’erreur suit une loi normale de moyenne nulle et de variance égale à 2,5, mais qui diminue proportionnellement avec la durée de l’entraînement. En rendant ce terme inversement proportionnel à la durée de l’entraînement intensif, nous permettons progressivement au système de décider par lui-même quels tuples choisir.

Le terme $(\xi_k + \epsilon_{ijk})$ est ensuite pondéré par la précision p_k ⁶ d’utilisation du tuple. La raison pour laquelle nous pondérons les scores par leur précision est due aux bornes. En effet, il n’est pas rare — après une longue période d’entraînement — que plusieurs tuples positivement renforcés atteignent les bornes supérieures. Dans ces cas-là, le système est incapable de discriminer entre ces tuples puisqu’ils ont tous des scores maximaux. En pondérant les scores des tuples par leur précision, nous introduisons un mécanisme de distinction supplémentaire.

Le numérateur est ensuite normalisé par la longueur relative de l’ensemble de tuples $\frac{m}{\hat{m}}$ où m correspond au nombre de candidats pour la réplique donnée et \hat{m} au nombre de candidats maximum pour toutes les répliques. Si on ne normalise pas, le système est biaisé à choisir les répliques les plus longues, surtout quand les scores sont similaires en début d’apprentissage. La raison pour laquelle nous normalisons par la longueur relative de chaque ensemble de tuples d’une réplique i par rapport à la réplique ayant la cardinalité de tuples de remplacement maximale est dû au fait que nous sommes obligés d’arrondir les scores afin de satisfaire la modélisation en nombres entiers. Si nous normalisons les scores pondérés par m uniquement, nous nous exposons à une disparition d’information importante sur les scores.

Finalement, nous prenons l’entier le plus proche en appliquant l’opérateur $\lfloor \cdot \rfloor$.

Mécanisme de filtrage des candidats tuples racines-classes

Nous nous tournons maintenant vers un mécanisme de filtrage des tuples racines-classes, c’est-à-dire un mécanisme qui permet de déterminer quels sont les candidats possibles ou non dépendamment de la phase de recherche dans laquelle se retrouve Philia. Rappelons que Philia possède une *variable interne émotive* qui permet de jauger la concordance ou dissonance émotive entre Philia et un utilisateur humain pendant leur conversation. Sachant

6. Rappelons que nous avons défini la précision p_t à la page 50.

que λ est directement proportionnel à V , l'idée de base du mécanisme de filtrage est la suivante. Soit

$$\Upsilon_{ij}^k = \frac{e^{\frac{\nu_{ij}^k}{\lambda}}}{\sum_l e^{\frac{\nu_{ij}^l}{\lambda}}} \quad (3.10)$$

la probabilité de choisir le tuple t_{ij}^k parmi les K candidats qui peuvent remplacer la racine-classe $t_{ij}^1 = t_{ij}$. La matrice Υ_{ij}^k est une matrice de probabilités compilée à partir des scores de l'ensemble des K tuples k pouvant être choisis pour le remplacement du j ième mot dans la i ème réplique. Nous nous sommes inspirés du modèle logit de la théorie des choix discrets afin d'instaurer un mécanisme d'exploration aléatoire qui dépend du paramètre λ , c'est-à-dire de la variable interne émotionnelle. L'idée est qu'en début d'entraînement ou de conversation, le système débute avec une valeur émotionnelle basse, ce qui correspond à une faible valeur pour λ , et permet au système de commencer dans une phase de renforcement. Lors de cette phase de renforcement, le système choisit les tuples ayant les scores les plus élevés parmi les tuples possibles. Quand λ tend vers 0 on observe que :

$$\lim_{\lambda \rightarrow 0} \Upsilon_{ij}^k = \begin{cases} 1 & \text{si } \nu_{ij}^k > \nu_{ij}^l \quad \forall l \neq k, \\ 0 & \text{si } \nu_{ij}^k < \max_l \{\nu_{ij}^l\}. \end{cases} \quad (3.11)$$

D'un autre côté, plus la valeur émotionnelle interne augmente, plus le système performe bien. Nous voulons donc encourager le système à prendre des risques en choisissant des tuples qui ont un score émotionnel faible ou non initialisé. Il s'agit donc de lancer une phase exploratoire qui permet d'introduire de nouveaux tuples. Remarquons que lorsque λ augmente, le système a la possibilité de choisir parmi un éventail plus large de tuples. À la limite, quand λ tend vers des valeurs très grandes, tous les tuples ont une probabilité équiprobable :

$$\lim_{\lambda \rightarrow \infty} \Upsilon_{ij}^k = \frac{1}{K} \quad (3.12)$$

Pour déterminer quels sont les candidats admissibles, nous construisons un ensemble d'indices admissibles R_{ij} . Cet ensemble est construit en pigeant avec remise K fois un indice de position $k \in \{1, \dots, K\}$ de probabilité Υ_{ij}^k . Cet ensemble nous donne une liste d'indices admissibles. Nous intégrons cette admissibilité directement dans le modèle en rajoutant la contrainte suivante :

$$x_{ij}^k \leq \hat{x}_{ij}^k \quad \forall i, j, k \quad (3.13)$$

avec \hat{x}_{ij}^k une matrice indicatrice de l'admissibilité des indices :

$$\hat{x}_{ij}^k = \begin{cases} 1 & \text{si l'indice } k \in R_{ij} \quad \forall i, j, k, \\ 0 & \text{sinon.} \end{cases} \quad (3.14)$$

Ce mécanisme d'exploration permet aussi de nous assurer que le système utilise des synonymes et antonymes qui n'ont pas encore de scores ou qui n'ont pas été sélectionnés suffisamment pour déterminer un véritable score émotionnel.

Le modèle

Le modèle implémenté est basé sur le modèle minimaliste avec les coefficients calculés ci-dessus et l'introduction de la contrainte supplémentaire 3.13. Réécrivons le modèle.

$$\begin{aligned} \max_x \quad & \sum_{i,j,k} \nu_{ij}^k x_{ij}^k \\ \text{s.t.} \quad & x_{ij}^k \leq \hat{x}_{ij}^k \quad \forall i, j, \\ & x_{ij}^k \leq y_i \quad \forall i, j, k \\ & \sum_i y_i \leq 1 \quad \forall i \\ & \sum_k x_{ij}^k \leq 1 \quad \forall i, j \\ & y_i \leq \sum_k x_{ij}^k \quad \forall i, j \\ & x_{ij}^k \in \{0, 1\} \quad \forall i, j, k \\ & y_i \in \{0, 1\} \quad \forall i \end{aligned} \quad (3.15)$$

3.4.4 Détails sur la reconstruction de la réplique optimale

Une fois obtenue une solution x_{ij}^k réalisable, nous pouvons reconstruire la réplique choisie et l'envoyer à l'utilisateur. Ici, il s'agit d'aller chercher les indices des tuples racine-classe dans la matrice x_{ij}^k auxquelles le solveur aura assigné la valeur vraie et d'aller récupérer les tuples correspondants. Pour reconstruire la phrase, on remplace les mots sélectionnés lors de l'optimisation par les tuples de la solution après les avoir soigneusement adaptés à leur rôle dans les phrases. Finalement, lorsque la réplique originale est optimisée et reconstruite, notre correcteur grammatical corrige celle-ci pour corriger toute erreur introduite lors de la reconstruction de la réplique.

3.5 Bloc 5 : Centre de commande

La dernière partie du système est celle du centre de commande. C'est en quelque sorte le cerveau du système, car son rôle est de diriger l'apprentissage. Il est essentiellement composé de deux boucles rétroactives et a principalement trois fonctions. Premièrement, une boucle rétroactive automatique calcule un score delta de mise à jour des scores à partir d'une évaluation de la différence de l'émotion perçue entre l'utilisateur et le système à chaque interaction. Deuxièmement, une boucle rétroactive usager sollicite directement l'utilisateur lorsque des anomalies au niveau de la consonance et/ou de la dissonance émotives surviennent de manière répétitive. Finalement, le centre de commande met à jour la variable émotionnelle interne qui permet de diriger et contrôler les phases exploratrices du système.

3.5.1 Boucle rétroactive automatique

La boucle de rétroaction automatique est un mécanisme de contrôle de la consonance et dissonance émotives entre l'utilisateur et le système. À chaque interaction avec l'utilisateur, les scores des tuples racines-classes choisis par le *Bloc 4 : Optimisation* sont mis à jour. Cette boucle rétroactive comporte trois éléments : un *scoreur individuel* qui attribue un *score*⁷ *individuel* à chaque réplique de l'utilisateur et du système, un *scoreur combinatoire* qui analyse la différence entre les scores individuels des deux parties pour déterminer un *score global* et mettre à jour les scores des tuples racines-classes de notre base de données et finalement, notre *variable interne émotionnelle* dont nous avons déjà abondamment parlé.

Scoreur individuel

Le scoreur individuel utilise les classificateurs sentimentaux et émotionnels du *Bloc 2 : Compréhension du langage naturel* pour détecter et mesurer la présence de l'émotion cible aussi bien dans la réplique de l'utilisateur que dans celle du système. Il permet aussi de déterminer la polarité de celles-ci. Dans un premier temps et dans le cadre de ce mémoire, nous utilisons exactement les mêmes paramètres du scoreur individuel aussi bien pour l'utilisateur que pour le système.

Nous reprenons les mêmes notations que celles décrites à la section 3.4.2 pour noter les familles d'émotions similaires et opposées. Soient $\text{sim}(\mathcal{E}) = \{\mathcal{E}_1 = \mathcal{E}, \mathcal{E}_2, \dots\}$ un ensemble d'émotions similaires à l'émotion cible \mathcal{E} et $\text{con}(\mathcal{E}) = \{\mathcal{F}_1, \mathcal{F}_2, \dots\}$ un ensemble d'émotions opposées.

7. Bien que nous parlons de score individuel, il faut bien distinguer les scores des tuples racines-classes des scores - individuels ou globaux - des répliques.

Le scoreur individuel calcule une combinaison convexe $N(i)$ d'un score émotionnel $\alpha(i)$ et d'un score sentimental $\beta(i)$ pour chacune des réplique i :

$$N(i) = \lambda_0 \alpha(i) + \lambda_1 \beta(i) \quad (3.16)$$

avec $\lambda_0 + \lambda_1 = 1$. Pour l'implémentation de Philia, nous utilisons des scores entiers et donc nous prenons l'entier le plus proche :

$$N_i = \lfloor N(i) \rfloor. \quad (3.17)$$

Avant d'expliquer $\alpha(i)$ et $\beta(i)$, introduisons deux fonctions indicatrices pour une réplique i donnée. La première, $E(i, \mathcal{E})$, permet d'indiquer si une réplique i peut être considérée comme reflétant l'émotion \mathcal{E} d'après le classificateur émotionnel. Ce dernier retourne une probabilité $P(\mathcal{E}|i)$ et nous définissons

$$E(i, \mathcal{E}) = \begin{cases} 1 & \text{si } P(\mathcal{E}|i) > \Phi_{\mathcal{E}}, \\ 0 & \text{sinon.} \end{cases} \quad (3.18)$$

pour un certain seuil $\Phi_{\mathcal{E}}$ donné. La deuxième, $S(i)$, permet de traduire la polarité — $\{positif, neutre, négatif\}$ — d'une réplique i en l'ensemble $\{1, 0, -1\}$ comme suit :

$$S(i) = \begin{cases} 1 & \text{si } P(positive|i) > \Phi_{pos}, \\ -1 & \text{si } P(negative|i) > \Phi_{neg}, \\ 0 & \text{sinon.} \end{cases} \quad (3.19)$$

avec Φ_{pos} et Φ_{neg} des seuils donnés et $P(\cdot|i)$ la probabilité de polarité donnée par le classificateur sentimental pour une réplique i donnée.

Nous sommes maintenant en mesure d'introduire les hyper-paramètres α_0, α_1 et α_2 ainsi que les définitions de $\alpha(i)$ et $\beta(i)$ pour une réplique i donnée :

$$\begin{aligned} \alpha(i) = & \alpha_0 E(i, \mathcal{E}) + \\ & \alpha_1 \sum_{\mathcal{E}_l \in \text{sim}(\mathcal{E}) \setminus \mathcal{E}} E(i, \mathcal{E}_l) + \\ & \alpha_2 \sum_{\mathcal{F}_l \in \text{con}(\mathcal{E})} E(i, \mathcal{F}_l) \end{aligned} \quad (3.20)$$

et

$$\beta(i) = S(i). \quad (3.21)$$

La somme convexe 3.16 permet de pondérer l'apport au score individuel du classificateur émotionnel et du classificateur sentimental. Si $\lambda_0 = 1$, nous ne considérons que le classificateur émotionnel, alors que si $\lambda_1 = 1$, nous ne considérons que le classificateur sentimental. Les hyper-paramètres α_0, α_1 et α_2 permettent quant à eux de pondérer le poids attribués à l'émotion cible \mathcal{E} , aux émotions similaires (différentes de l'émotion cible) et aux émotions opposées respectivement.

Scoreur combinatoire

Le scoreur combinatoire est un mécanisme de détection de la consonance et dissonance émotive entre l'utilisateur et le système. Essentiellement, nous comparons les scores individuels du système N_0 et de l'utilisateur N_1 et calculons un score global $T(N_0, N_1)$. Rappelons qu'il s'agit de scores entiers et que nous avons trois possibilités :

- $N_i \geq 1$: l'émotion cible est détectée dans la réplique i ;
- $N_i \leq -1$: les émotions opposées sont détectées dans la réplique i ;
- $N_i = 0$: l'analyse n'est pas concluante.

Nous comparons les 9 possibilités pour les valeurs de N_0 et N_1 et agissons en conséquence. Dépendamment du cas dans lequel nous nous trouvons, Philia effectue une mise à jour des scores des tuples racine-classe immédiatement, n'effectue aucune mise à jour ou déclenche la boucle de rétroaction usager. Nous illustrons le cas où l'émotion cible et le contexte donnés requièrent une concordance émotive, Philia essaiera donc d'être sur la même longueur d'onde émotive que l'utilisateur. Nous nous servirons alors de l'algorithme 1. Quand une dissonance émotive est détectée trop souvent, cas où $\text{nombre_erreurs} \geq \text{nombre_erreurs_permis}$ et $\text{nombre_fausses_erreurs} \geq \text{nombre_fausses_erreurs_permis}$ dans l'algorithme, Philia enclenche alors la boucle rétroactive usager pour demander directement à celui-ci de participer à l'évaluation des scores des tuples racines-classes dans les répliques du système.

Algorithm 1 Calculer le score combiné $T(N_0, N_1)$, partie 1/2

Variables d'entrée :

N_0 = score individuel de la réplique du système

N_1 = score individuel de la réplique de l'utilisateur

Initialisation :

$\{N_0, N_1\} \in \mathbb{Z}$

$nombre_erreurs \leftarrow 0$

$nombre_erreurs_permis \leftarrow \alpha$

$nombre_fausses_erreurs \leftarrow 0$

$nombre_fausses_erreurs_permis \leftarrow \beta$

$\{\alpha, \beta\} \in \mathbb{N}$

les différents cas $\{N_0, N_1\}$:

if $N_0 \geq 1$ and $N_1 \geq 1$ **then**

 // Le système et l'utilisateur sont en accord :

$T(N_0, N_1) \leftarrow 1$

else if $N_0 \geq 1$ and $N_1 == 0$ **then**

 // Le système et l'utilisateur sont moyennement en accord :

$T(N_0, N_1) \leftarrow 1$

else if $N_0 \geq 1$ and $N_1 \leq -1$ **then**

 // Le système et l'utilisateur ne sont pas en accord :

if $nombre_erreurs < nombre_erreurs_permis$ **then**

$T(N_0, N_1) \leftarrow 0$

$nombre_erreurs \leftarrow +1$

else if $nombre_erreurs \geq nombre_erreurs_permis$ **then**

$scores_rétroaction_usager \leftarrow rétroaction_active()$

$T(N_0, N_1) \leftarrow scores_rétroaction_usager$

$nombre_erreurs \leftarrow 0$

end if

else if $N_0 == 0$, $N_1 \geq 1$ **then**

 // Le système et l'utilisateur sont moyennement en accord :

$T(N_0, N_1) \leftarrow 0$

else if $N_0 == 0$ and $N_1 == 0$ **then**

 // Le système et l'utilisateur sont neutres :

$T(N_0, N_1) \leftarrow 0$

Algorithm 2 Calculer le score combiné $T(N_0, N_1)$, partie 2/2

```

else if  $N_0 == 0$  and  $N_1 <= -1$  then
  // Le système et l'utilisateur sont moyennement en désaccord :
  if nombre_erreurs < nombre_erreurs_permis then
     $T(N_0, N_1) \leftarrow 0$ 
    nombre_erreurs  $\leftarrow +1$ 
  else if nombre_erreurs  $\geq$  nombre_erreurs_permis then
    scores_rétroaction_usager  $\leftarrow$  rétroaction_active()
     $T(N_0, N_1) \leftarrow$  scores_rétroaction_usager
    nombre_erreurs  $\leftarrow 0$ 
  end if
else if  $N_0 <= -1$  and  $N_1 \geq 1$  then
  // Le système et l'utilisateur sont en désaccord sur la négativité perçue :
  if nombre_fausses_erreurs < nombre_fausses_erreurs_permis then
     $T(N_0, N_1) \leftarrow 0$ 
    nombre_fausses_erreurs  $\leftarrow +1$ 
  else if nombre_fausses_erreurs  $\geq$  nombre_fausses_erreurs_permis then
    scores_rétroaction_usager  $\leftarrow$  rétroaction_active()
     $T(N_0, N_1) \leftarrow$  scores_rétroaction_usager
    nombre_fausses_erreurs_permis  $\leftarrow 0$ 
  end if
else if  $N_0 <= -1$  and  $N_1 == 0$  then
  // Le système et l'utilisateur sont moyennement en désaccord sur la négativité perçue :
  if nombre_fausses_erreurs < nombre_fausses_erreurs_permis then
     $T(N_0, N_1) \leftarrow 0$ 
    nombre_fausses_erreurs  $\leftarrow +1$ 
  else if nombre_fausses_erreurs  $\geq$  nombre_fausses_erreurs_permis then
    scores_rétroaction_usager  $\leftarrow$  rétroaction_active()
     $T(N_0, N_1) \leftarrow$  scores_rétroaction_usager
    nombre_fausses_erreurs_permis  $\leftarrow 0$ 
  end if
else if  $N_0 <= -1$  and  $N_1 <= -1$  then
  // Le système et l'utilisateur sont en accord sur la négativité perçue :
   $T(N_0, N_1) \leftarrow -1$ 
end if

```

Les hyper paramètres α et β permettent de contrôler la fréquence d'apparition de la boucle rétroactive usager.

Le score global $T(N_0, N_1)$ est utilisé pour mettre à jour les scores ξ_i^t des tuples racines-classes t de la base de données comme suit :

$$\xi_i^t = \xi_i^t + \tau \cdot T(N_0, N_1) \quad (3.22)$$

Nous utilisons un facteur de pondération τ pour contrôler la vitesse de la mise à jour des scores. Notons que la mise à jour se fait dans la même direction (positive ou négative) pour tous les tuples racine-classe t de la réplique reconstruite i . Ceci est une faiblesse du système, car nous savons que tous les mots ne contribuent pas de la même force à l'élicitation émotionnelle. Une amélioration intéressante à apporter serait d'implémenter un mécanisme d'attention issu du classificateur émotionnel pour détecter quels mots contribuent statistiquement le plus à l'émotion reconnue et de mettre à jour les scores proportionnellement au poids des tuples contributeurs.

Variable interne émotionnelle

Le dernier rôle de la boucle rétroactive automatique est de mettre à jour la variable émotive interne V de Philia. Nous nous sommes inspirés de certains travaux en apprentissage par renforcement qui ont utilisé un tel mécanisme. Coutinho et al. [57] implémentent un agent qui tire profit d'une variable interne homéostatique qui modifie directement la fonction de valeur (*value function*) afin de modifier le comportement du système. Cette valeur est associée à des facteurs externes au robot tels que la durée de son utilisation ou sa performance et ces facteurs contribuent à modifier la fonction de la valeur interne de l'agent. Les auteurs notent que l'utilisation d'une telle variable émotive permet à l'agent d'être plus stable dans la sélection de ses états futurs, car il prend en compte non seulement l'état actuel, mais aussi l'état des variables homéostatiques qui doivent satisfaire d'autres besoins connexes qui sont aussi nécessaires au bon fonctionnement de l'agent. C'est donc cette idée de vouloir contrôler le comportement de l'agent de manière dynamique qui nous a inspirés à intégrer une telle variable. Notre variable interne émotionnelle est une sorte de métrique sur la performance du système. Elle se réfère à la capacité de Philia de transmettre l'émotion cible à travers les répliques générées, fortement pénalisée par la génération de répliques transmettant une des émotions opposées. La fonction la plus importante de cette variable interne est de modifier la probabilité de sélection des tuples racine-classe candidats considérés pour le remplacement

des tuples lors de l'optimisation en faisant varier le paramètre λ lors de la construction de la matrice de probabilités de sélection Υ_{ijk} que l'on retrouve dans la présentation du modèle à la section 3.4.3. La variable interne émotionnelle et le facteur λ sont reliés selon la formule suivante :

$$\lambda = \log(a \cdot V) \quad (3.23)$$

Dans le cadre de ce mémoire, nous avons utilisé $a = 5$. Nous avons déterminé empiriquement la valeur de a de sorte à obtenir le comportement voulu du système. L'idée de faire varier la variable interne émotionnelle est que lorsque le modèle performe mal, ou est en début de la phase dialogale, nous voulons que les chances de sélection des tuples valides ayant un score et une précision élevés soient quasi assurées. La raison est que lorsque la variable émotionnelle interne est basse, nous ne pouvons nous permettre de courir des risques en choisissant des mots qui potentiellement ne transmettent pas l'émotion voulue de manière adéquate. Nous voulons que notre système soit conservateur dans sa sélection des tuples candidats. Quand la variable interne émotionnelle est élevée, nous pouvons interpréter cela comme une performance cohérente, ce qui nous donne une certaine confiance dans le système. C'est le moment idéal pour pousser le système à essayer de choisir de remplacer les termes originaux par des tuples dont la performance émotive est plus incertaine, ce qui permet au système d'explorer de nouvelles combinaisons et d'éviter un renforcement trop rigide.

La mise à jour de la variable interne se fait pendant la phase rétroactive automatique. Nous comptons le nombre d'interactions entre le système et un usager particulier, puis lorsqu'il est temps d'attribuer un score global à l'interaction, nous mettons à jour la variable interne comme suit :

$$V = \frac{\#Q - \alpha \cdot \#O}{\#U} \quad (3.24)$$

où $\#Q$ représente le nombre de fois que le scoreur combinatoire reconnaît que la réplique générée par le système élicite l'émotion cible tout en s'assurant que l'usager la perçoit comme telle, $\#O$ représente le nombre de fois que le scoreur combinatoire reconnaît que la réplique générée par le système élicite les émotions opposées et $\#U$ représente le nombre d'échanges dans la conversation. L'hyper paramètre α permet de pénaliser différemment Philia. Ce facteur peut être ajusté pour diminuer ou augmenter l'impact de la génération d'une réplique opposée à l'émotion cible du système. Situons-nous de nouveau dans le cas où Philia doit se mettre en concordance émotive avec l'utilisateur. Si nous voulons encourager une phase

exploratoire plus importante, il suffit de diminuer α de sorte à négliger les erreurs commises par le système et par-là même augmenter λ ce qui aura pour conséquence de faire tendre les probabilités Υ_{ij}^k vers des probabilités équiprobables comme nous l'avons déjà mentionné. À l'inverse, si nous voulons avoir un système qui performe bien, ayant déjà une base de données des scores suffisamment étoffée, nous pouvons augmenter α de sorte que le système choisit les tuples remplaçables ayant le score et la précision les plus élevés.

Il faut également noter que dans l'éventualité où la pénalité induite par l'hyper paramètre α sur la variable interne émotionnelle V est trop forte, c'est-à-dire que $a \cdot V \leq 1$, alors nous fixons la valeur de λ égal à 0.01 pour toute la durée où cette condition est rencontrée. La raison est que nous voulons éviter le $\log(0)$ et nous voulons demeurer dans le domaine des réels positifs pour λ donc $a \cdot V > 1$. Nous avons choisi λ égal à 0.01, car nous avons déterminé de manière empirique que cette valeur était suffisamment petite pour garantir la sélection du tuple de remplacement ayant le score pondéré maximal.

3.5.2 Boucle rétroactive usager

La deuxième boucle d'interaction est un mécanisme qui implique activement l'utilisateur durant l'apprentissage continu du modèle. Cette boucle consiste à demander directement à l'utilisateur son opinion sur la valeur émotive des tuples racines-classes utilisés pour la réplique reconstruite ou de donner son avis sur la pertinence des synonymes ou antonymes employés. Cette boucle de rétroaction utilise deux requêtes visuelles envoyées à l'utilisateur où il doit noter, selon une échelle imposée, les tuples dépendamment de l'évaluation demandée. Ces requêtes sont lancées lorsque les erreurs considérées pendant le calcul du score combinatoire dépassent un certain seuil (voir la section 3.5.1).

Les boucles rétroactives usagers ont plusieurs utilités. Premièrement, elles permettent de corriger manuellement des paramètres qui pourraient être mal initialisés. Deuxièmement, l'intervention de l'utilisateur permet de mieux cibler sa compréhension de certaines racines-classes dans certains contextes pour une émotion donnée. Troisièmement, comme ces requêtes ne sont envoyées que lorsque Philia a un doute ou une hésitation, elles permettent de trancher dans le bon sens. Finalement, l'interaction directe avec l'utilisateur permet à celui-ci de se sentir impliqué dans le processus de construction de Philia ce qui permet peut-être un certain rapprochement et une certaine empathie et donc tolérance pour le système dans son ensemble.

Il est à noter que l'utilisation des boucles rétroactives usager se base sur une aide bienveillante des utilisateurs. Si certains n'apportent pas des répliques sensées, on peut espérer que dans l'ensemble la plupart des répliques seront données avec sérieux et bienveillance et donc seront

statistiquement valides. On peut aussi rajouter un mécanisme de contrôle qui permettrait d'écarter les répliques douteuses ou bien rajouter des requêtes plus approfondies pour déjouer les malhonnêtetés.

Requête d'évaluation émotionnelle La requête d'évaluation émotionnelle demande à l'utilisateur de noter de 1 à 5 la corrélation émotionnelle des différents tuples de la phrase reconstruite en fonction de l'émotion cible visée durant la conversation. La figure 3.2 illustre une fenêtre où Philia propose une requête d'évaluation émotionnelle à l'utilisateur.

The screenshot shows a window titled "Requested user feedback window" with standard window controls (minimize, maximize, close). The main text reads: "Please rate each synonym word from 1 to 5, 5 being the strongest, according to the target emotion 'JOY'." Below this, there is a list of four words: "wish", "devour", "adore", and "ok". Each word is in a separate input field, and to the right of each field is a small numeric rating control showing a value (3, 2, 4, 3 respectively) and a small arrow icon. At the bottom right of the window is a "Close" button.

FIGURE 3.2 Un exemple de requête graphique d'évaluation émotionnelle envoyée à l'utilisateur pour noter le score émotionnel des tuples sélectionnés dans la réplique reconstruite optimale.

Requête d'évaluation des synonymes et antonymes La requête d'évaluation des synonymes et antonymes demande à l'utilisateur de noter de 1 à 3 la force des synonymes ou antonymes employés pour remplacer le tuple original. Cette requête permet de faire un nettoyage des synonymes ou antonymes inappropriés associés aux tuples présents dans notre base de données vocabulaire. La figure 3.3 illustre une fenêtre où l'on sert une requête d'évaluation des synonymes ou antonymes à l'utilisateur.

Requested user feedback window

rate each synonym from 1 to 3, 3 being the strongest, according to it's synonym strength to the original word

The original words	The synonym words	Rate the synonyms
pray	wish	1
eat	devour	2
love	adore	3
yes	ok	2

Close

FIGURE 3.3 Un exemple de requête graphique d'évaluation des synonymes et antonymes envoyée à l'utilisateur pour améliorer les synonymes et antonymes de la base de données.

Lors de cette requête, nous demandons à l'utilisateur de noter de 1 à 3 la qualité des synonymes et antonymes dans le contexte présent pour l'émotion cible et la réplique du moment. Ces notes sont conservées par Philia afin d'effectuer un nettoyage progressif de la base de données des synonymes et antonymes. Lorsque l'utilisateur attribue la note de 1 à un synonyme ou antonyme employé, le système garde en mémoire le fait qu'il a été signalé. Lorsque ce même synonyme ou antonyme du mot associé est signalé de nouveau, nous le retirons de la liste.

3.6 Fonctionnement de Philia

Maintenant que toutes les composantes de Philia ont été introduites, nous pouvons décrire le mécanisme d'ensemble⁸. Celui-ci est présenté par un organigramme illustré sur la figure 3.4. Sur cet organigramme, les interactions principales entre les blocs présentés sont identifiées à l'aide de flèches colorées de trois couleurs. Les flèches noires représentent le cycle principal de Philia, les flèches bleues représentent le cycle de la rétroaction automatique et les flèches rouges indiquent les étapes de la rétroaction usager. Ces cycles peuvent se recouper et doivent être vus comme un tout cohérent et non pas comme trois cycles indépendants.

8. Un pseudo-code de la boucle principale de Philia est présenté à l'annexe C.

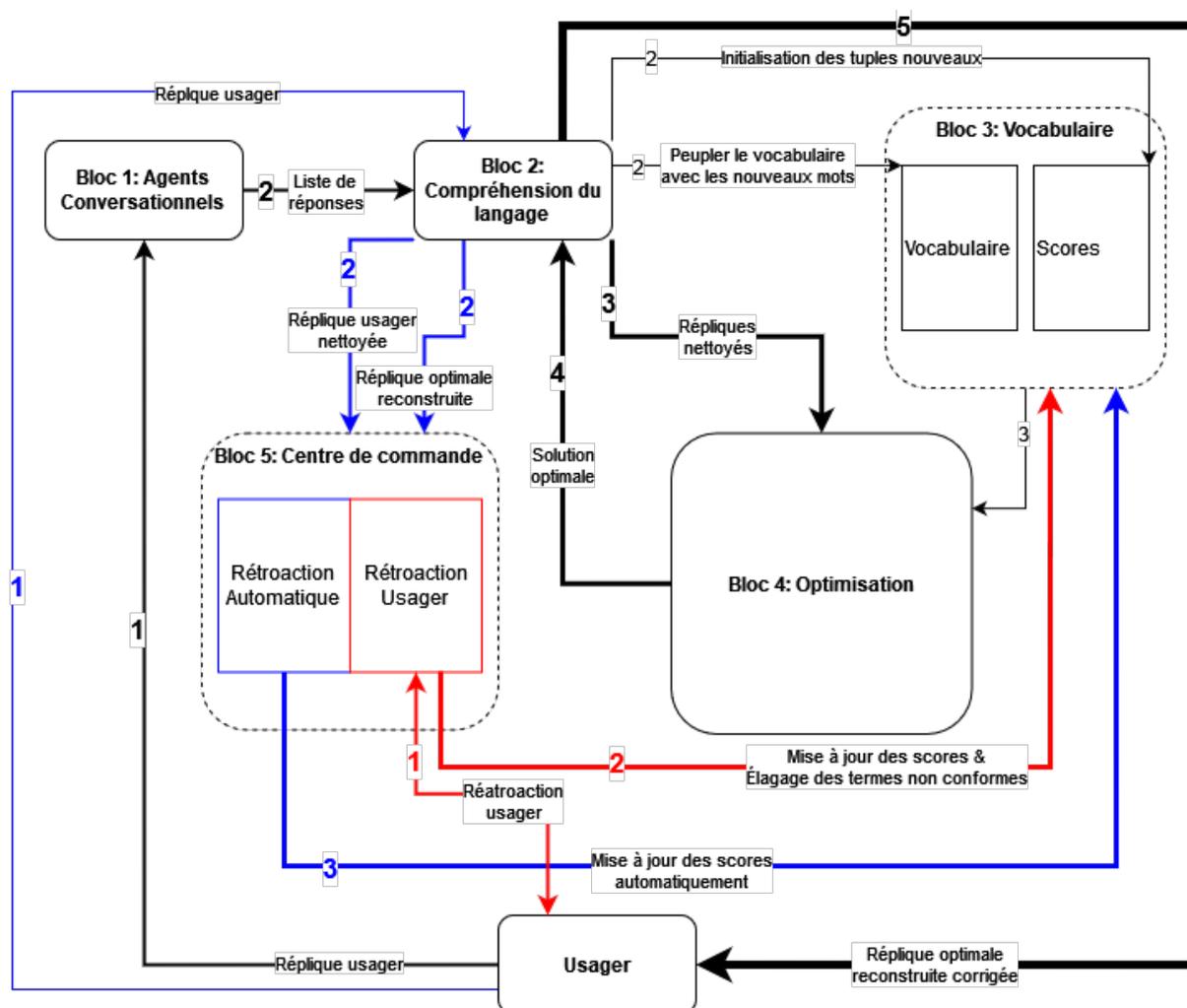


FIGURE 3.4 L'organigramme complet du fonctionnement de Philia, les flèches noires représentent le cycle principal, les flèches bleues représentent la séquence de rétroaction automatique et les flèches rouges la séquence de rétroaction usager.

Cycle principal

Le cycle principal de Philia est le mécanisme de fonctionnement de base du système, c'est-à-dire le cycle de transformation des répliques sans tenir compte de la mise à jour des scores. Suivons le cycle étape par étape :

1. Le cycle débute avec une interaction, une requête, de l'utilisateur envers Philia. La requête est envoyée aux agents conversationnels pour qu'ils génèrent une liste de répliques pertinentes.
2. La liste de répliques est envoyée au *Bloc 2 : Compréhension du langage naturel* afin de décomposer chaque réplique en tuples racine-classe, les nettoyer des mots d'arrêts et

de la ponctuation. Pendant cette décomposition, Philia détermine l'existence ou non de chaque mot original dans la base de données vocabulaire. Si un mot manquant est identifié, une entrée est créée automatiquement dans la base de données vocabulaire qui est complétée par une recherche automatique sur internet pour trouver les synonymes et antonymes associés. Pour chaque tuple manquant, une entrée dans la base de données scores est initialisée pour permettre au système de commencer à collecter de l'information pour ces entrées.

3. La liste des répliques décomposées et nettoyées est ensuite envoyée au *Bloc 4 : Optimisation* pour déterminer les remplacements possibles et collecter les candidats tuples admissibles. Un modèle est construit avec des coefficients adaptés et résolu.
4. La solution optimale (possiblement vide) est envoyée au *Bloc 2 : Compréhension du langage naturel* afin de procéder à la reconstruction et la correction de la réplique optimale choisie.
5. Cette réplique optimale est ensuite renvoyée à l'utilisateur pour continuer le dialogue.

Cycle de rétroaction automatique

Le cycle de rétroaction automatique de Philia est le mécanisme qui met automatiquement à jour les scores de notre base de données ainsi que la variable émotionnelle interne. Suivons le cycle étape par étape :

1. Comme pour le cycle principal, nous récoltons la requête de l'utilisateur, mais, cette fois-ci, elle est envoyée directement au *Bloc 2 : Compréhension du langage naturel*, afin de la décomposer et la nettoyer.
2. La réplique nettoyée est envoyée au *Bloc 5 : Centre de commande*, plus spécifiquement à la composante de la rétroaction automatique qui calcule le score individuel associé. La réplique optimale reconstruite se voit aussi attribué un score individuel.
3. Finalement, un score global est calculé pour les deux répliques et les tuples de la réplique optimale sont mis à jour. C'est aussi à cette étape que le système met à jour la variable interne émotionnelle.

Cycle de rétroaction usager

Le cycle de rétroaction usager de Philia est le mécanisme qui s'occupe de nettoyer progressivement la base de données du vocabulaire et affiner les scores émotifs des tuples grâce à l'intervention continue et bienveillante des utilisateurs. Suivons le cycle étape par étape :

1. Lorsque jugé nécessaire, Philia lance les requêtes usager pour obtenir un feedback humain.
2. Ce feedback est mis a profit pour mettre à jour les scores des tuples de la réplique optimale de Philia et ajuster les entrées de la base de données vocabulaire si nécessaire.

CHAPITRE 4 MÉTRIQUE D'ÉVALUATION ET RÉSULTATS

Nous présentons dans ce chapitre une métrique d'évaluation globale qui nous permet d'évaluer automatiquement la qualité d'un dialogue - aussi bien au niveau de l'émotion que plus globalement - et les résultats obtenus par Philia par rapport à cette métrique. Nous discuterons aussi les choix de certains hyper-paramètres et les résultats obtenus conséquemment. Nous portons une attention toute particulière à la variable interne émotionnelle, car elle est centrale dans notre implémentation proposée de Philia.

4.1 Quelques considérations pratiques sur les métriques

Avant d'aborder la métrique choisie et son adaptation à Philia, nous aimerions faire deux commentaires. Le premier commentaire concerne la difficulté notoire, surtout pour des systèmes de dialogues ouverts, de concevoir des métriques réellement utilisables aussi bien pour la recherche que pour une mise en production d'un agent conversationnel dans l'industrie. La plupart des métriques connues pour les dialogues ouverts — que ce soient les métriques BLUE [224], ROUGE [225] ou même BERTScore [235] — utilisent des références, c'est-à-dire des répliques modèles prédéfinies. Pour mesurer la pertinence y compris émotive des systèmes de dialogues ouverts, ces métriques comparent donc des répliques générées à des répliques cibles. Sans discuter de la pertinence des moyens utilisés pour comparer une réplique à une réplique cible, on peut se demander si l'approche forcément subjective en tant que telle est véritablement pertinente. Cette approche reste néanmoins la plus courante et utilisée faute de mieux.

Le deuxième commentaire concerne Philia. Pour mesurer l'avantage d'utiliser un système tel que Philia, nous devons utiliser une métrique capable de mesurer, par exemple, la concordance ou la dissonance émotive avec ou sans l'ajout de Philia. La comparaison des deux mesures devrait pouvoir nous indiquer si un progrès est fait ou pas et en avoir une certaine estimation. Nous voulons cependant attirer l'attention du lecteur sur le fait que Philia utilise en permanence sa propre métrique interne puisque c'est justement en comparant certains scores que Philia maximise l'injection d'émotion et prend des décisions concernant le choix des mots à utiliser dans les répliques des agents conversationnels. Nous ne pouvons cependant pas utiliser cette métrique telle quelle pour comparer les effets de Philia sur un ou plusieurs agents conversationnels puisque par définition Philia maximise celle-ci. Par contre, et c'est ce que nous faisons dans le cadre de cette thèse, nous pouvons nous servir des classificateurs émotionnel et sentimental internes pour mesurer l'injection d'émotion par Philia.

Ces commentaires sont à la base de notre choix de la métrique utilisée dans le cadre de ce mémoire. Nous avons choisi une métrique reconnue pour estimer l'ensemble des interactions entre le système et l'utilisateur non seulement dans son ensemble, mais aussi réplique par réplique. Notre choix a porté sur une métrique qui ne tient pas forcément compte des émotions et nous l'avons donc complété par l'utilisation des classificateurs émotionnels et sentimentaux déjà utilisés.

4.2 Métrique d'évaluation automatique

Afin de pouvoir mesurer adéquatement la performance de notre système, nous devons concevoir ou adapter une métrique d'évaluation globale de dialogue ouvert. Nous nous sommes inspirés des métriques d'évaluation de dialogue libre de références, plus spécifiquement, nous avons adapté la métrique FED (Fine-grained Evaluation of Dialog) introduite par Mehri et Eskenazi [239]. Nous référons le lecteur à l'article original pour les détails spécifiques concernant cette métrique. Nous décrivons ici la métrique de manière générale.

4.2.1 La métrique FED

La métrique FED permet d'évaluer automatiquement un dialogue ouvert entre deux interlocuteurs en tenant compte de 18 différentes caractéristiques dialogales. La métrique FED se sert du modèle préentraîné DialoGPT de Microsoft sur des dialogues de Reddit (147 millions d'interactions) [243] afin d'extraire l'information pertinente nécessaire à l'évaluation. Les avantages de cette métrique sont qu'elle :

- ne repose pas sur une réponse de référence (ground-truth) ;
- n'a pas besoin d'entraînement et
- mesure finement aussi bien la qualité des répliques individuelles que l'ensemble des dialogues.

Le fonctionnement de la métrique FED est basé sur deux hypothèses. Premièrement, lors de l'entraînement de DialoGPT, le modèle réussit à capturer des connaissances générales sur le langage. Les auteurs émettent donc l'hypothèse que DialoGPT contient implicitement des connaissances générales sur la qualité d'un dialogue. Les auteurs affirment qu'en l'absence de données supplémentaires permettant d'affiner le modèle DialoGPT, ils ne peuvent vérifier leur hypothèse directement, mais qu'il était possible de tout de même exploiter le modèle afin d'extraire l'information quant à la qualité d'un dialogue. Deuxièmement, en se basant sur les travaux d'Eskenazi et coll. [248], les auteurs affirment qu'il est possible de déterminer la qualité d'une interaction conversationnelle en étudiant la réponse de l'utilisateur après avoir

reçu une réplique du système. En effet, Eskenazi et coll. proposent l'idée que lorsqu'un usager répond à une réplique du système, cette réponse peut, explicitement ou implicitement, convoier de l'information quant à la qualité du dialogue en cours. C'est une forme de rétroaction implicite. Par exemple, lorsque l'utilisateur répond par *ce n'est pas très intéressant* à une réplique du système, nous pouvons implicitement collecter de l'information quant à la qualité du dialogue/de la réplique face à une caractéristique spécifique, telle que l'intérêt dans ce cas-ci.

En se basant sur ces hypothèses, les auteurs de la métrique FED ont construit un mécanisme d'évaluation automatique du dialogue sans avoir recours explicitement à des références externes. Les auteurs ont construit à la main, pour chacune des 18 caractéristiques dialogales que nous décrivons dans la section suivante, un ensemble de répliques positives $\{p_i\}$ et négatives $\{n_j\}$ qui servent de références (absolues et internes) pour chacune des 18 caractéristiques. La phrase citée plus haut *ce n'est pas très intéressant* est d'ailleurs un exemple d'une phrase négative pour la caractéristique *intérêt*. La métrique FED utilise alors une différence entre deux scores, D_+ pour les phrases positives et D_- pour les phrases négatives, qui correspondent à une corrélation entre la réplique générée et les répliques positives et négatives dépendamment d'un contexte donné. Ces scores sont calculés pour chacune des caractéristiques séparément. Nous référerons le lecteur à l'article et, plus encore, au code¹ pour une compréhension plus détaillée du calcul exact de ces scores. Ces deux scores sont alors soustraits l'un de l'autre :

$$D = D_+ - D_- \tag{4.1}$$

L'équation 4.1 détermine donc un score de similitude différentiel pour chaque caractéristique dialogale entre les répliques générées et celles contenues dans les ensembles $\{p_i\}$ et $\{n_j\}$. Si la valeur de D est positive pour une caractéristique donnée, cela peut s'interpréter comme une propension du dialogue ou de la réplique à abonder dans le sens des phrases positives $\{p_i\}$ et inversement une valeur négative indique un caractère négatif pour la caractéristique. Nous mettons en garde le lecteur, car les scores dépendent fortement des ensembles $\{p_i\}$ et $\{n_j\}$ et notamment de leur taille respective. Nous utilisons ces valeurs de manière relative pour pouvoir déterminer une différence entre deux systèmes, peu importe l'interprétation exacte de ces scores somme toute relatifs.

1. <https://github.com/Shikib/fed>

4.2.2 Les 18 caractéristiques comparatives

Les 18 caractéristiques considérées sont divisées en deux sous-ensembles. Le premier sous-ensemble d'évaluation comprend les caractéristiques associées à la qualité des échanges directs entre les interlocuteurs, c'est-à-dire un échange de deux répliques. Le deuxième sous-ensemble quant à lui, comprend les caractéristiques associées à la qualité globale d'une conversation, c'est-à-dire toutes les répliques du dialogue.

Les caractéristiques d'évaluation visant les échanges directs entre un usager et le système sont les suivantes : l'*intérêt*, l'*engagement*, l'*originalité*, la *pertinence*, la *compréhension* entre interlocuteurs, savoir si les répliques sont *sémantiquement appropriées*, la *compréhension* associée à la réplique et finalement la *fluidité*. Afin de comprendre ce à quoi chaque caractéristique se réfère, il est utile de se poser certaines questions à leurs égards. Par exemple, pour comprendre et connaître l'*intérêt* d'une réplique, la question qu'il faut se poser est si la réplique générée est intéressante pour l'usager moyen. Pour l'*engagement*, nous voulons détecter si la réplique générée suscite de l'engagement chez l'utilisateur moyen. L'*originalité* essaie d'évaluer la spécificité d'une réplique, est-elle applicable dans un cadre général ou est-elle spécifique au dialogue en cours ? La *pertinence* évalue la pertinence de la réplique face à la conversation en cours, c'est-à-dire vise à détecter si le système est hors sujet ou non. La *compréhension* entre l'usager et le système essaie de déterminer si les interlocuteurs se comprennent, par exemple, de savoir si la réplique générée crée des malentendus ou non chez l'utilisateur moyen. Évaluer si les répliques sont *sémantiquement appropriées* consiste à s'assurer que les répliques générées préservent un sens syntaxique correct. La *compréhension* de la réplique évalue si la réplique générée est claire pour l'usager moyen. Finalement, la *fluidité* de la réplique évalue si une réplique se lit de manière ininterrompue et fluide.

Les caractéristiques associées à l'évaluation du dialogue dans sa globalité sont les suivantes : la *cohérence*, la *recupération d'erreurs*, la *constance*, la *diversité*, la *profondeur*, la *sympathie* du système, la *compréhension*, la *flexibilité*, l'*informativité* du système et finalement la *curiosité*. Pour la *cohérence*, la métrique tente d'évaluer si durant le dialogue le système reste cohérent et fluide, c'est-à-dire que les échanges suivent une trame narrative constante. La *recupération des erreurs* évalue la capacité du système à pouvoir corriger les erreurs commises ayant été soulevées par l'usager. La *constance* évalue si le système est stable relativement aux informations fournies tout au long des interactions. La *diversité* évalue si le système est capable de modifier les réponses générées afin d'éviter les répétitions. La *profondeur* tente d'évaluer si le système creuse les sujets amenés par l'usager ou s'il change constamment de sujet. Évaluer la *sympathie* du système consiste à évaluer si l'utilisateur trouve le système sympathique à travers la génération de ses réponses. La *compréhension* évalue la capacité

du système à comprendre les requêtes de l'utilisateur, à savoir s'il y répond adéquatement. La *flexibilité* évalue la capacité du système à s'adapter aux changements amenés par l'utilisateur et à le suivre au long de ces nouveaux sujets amenés. Évaluer l'*informativité* du système consiste à déterminer si, durant le dialogue, le système est capable d'apporter de l'information nouvelle en rapport avec le sujet traité. Finalement, évaluer la *curiosité* du système consiste à déterminer si durant le dialogue le système cherche à en savoir davantage sur les sujets amenés par l'utilisateur tout en restant cohérent.

Selon les auteurs de la métrique, ce ne sont pas toutes les caractéristiques dialogales qui contribuent de manière équivalente à la perception de sa qualité. En effet, pour les caractéristiques dialogales inter-utilisateurs, ce sont l'intérêt, la pertinence et la fluidité qui contribuent majoritairement à la qualité d'un dialogue. Pour les caractéristiques dialogales globales, ce sont la cohérence, la sympathie et la compréhension du système qui contribuent plus significativement à la qualité perçue d'une conversation. Ce sont donc ces six caractéristiques que nous devons tenter d'améliorer ou de préserver tout en injectant l'émotion cible. Il est important de noter que l'objectif principal de Philia n'est pas d'améliorer la qualité du dialogue, mais d'injecter de l'émotion dans des dialogues existants. Nous ne cherchons donc pas à maximiser les scores d'évaluation associés aux différentes caractéristiques dialogales. Il est tout de même important d'en tenir compte, car nous ne voulons pas transformer les répliques de sorte à altérer radicalement la perception qualitative d'une conversation et la métrique FED nous permettra de déterminer si Philia réussit cette transformation adéquatement. Concernant l'émotion proprement dite, nous proposons une adaptation de la métrique dans la prochaine section.

4.2.3 Adaptation de la métrique FED

Dans le cadre de ce projet, nous avons adapté la métrique FED en nous servant de l'implémentation des auteurs retrouvée sur Github² en y ajoutant nous-mêmes des répliques supplémentaires aux ensembles (p) et (n) de manière balancée, afin d'avoir un éventail plus étendu de répliques de comparaison pour la fonction D . Concernant l'évaluation des émotions, nous avons intégré trois autres caractéristiques calculées différemment. Plus spécifiquement, nous avons d'abord intégré la *précision de la détection de l'émotion cible* des répliques générées. Cette précision est simplement le rapport du nombre de fois que l'émotion cible est détectée à travers les répliques générées, divisé par le nombre de répliques générées pour une conversation. Ensuite, nous utilisons le *score individuel moyen* du système par conversation à l'aide de notre scoreur individuel présenté à la section 3.5.1. Finalement, nous utilisons aussi

2. <https://github.com/Shikib/fed>

le *score combiné moyen* par conversation à l'aide de notre scoreur combinatoire présenté à la section 3.5.1. La combinaison de ces 23 caractéristiques nous permet à la fois de vérifier l'évolution de la qualité du dialogue après transformations tout en évaluant l'injection de l'émotion cible par Philia.

4.3 Évaluation et résultats

Pour évaluer le système, nous considérons la joie comme émotion cible (avec la peur, la colère, la tristesse, ainsi que le dégoût comme émotions opposées) et nous étudions sommairement deux contextes qui se sont révélés essentiellement similaires : un premier contexte (section 4.3.1) où l'analyse sentimentale est utilisée, mais dans une moindre mesure que l'analyse émotionnelle, et un deuxième contexte (section 4.3.2) où Philia se base uniquement sur l'analyse émotionnelle. Dans les deux contextes, nous avons étudié la variation de plusieurs hyper-paramètres. De manière générale, nous déduisons que Philia augmente significativement l'émotion cible sans altérer la qualité du dialogue.

Pour chacun des contextes, nous avons initialisé/entraîné Philia et puis nous avons figé la base de données des scores pour pouvoir évaluer Philia sur des données test. En effet, si nous ne figeons pas la base de données des scores, Philia continuera de mettre à jour les scores et le système sera différent à chaque itération de l'évaluation. Tous les tests ont été effectués sur un ordinateur avec 32GB DDR4-3600 de RAM, un CPU AMD Ryzen 5800x et un GPU GTX 1060 6GB. Mentionnons que l'initialisation/entraînement prend environ 15h alors que le calcul d'une évaluation prend environ 3h.

Pour l'évaluation du système, nous avons procédé de la manière suivante. Tout d'abord, pour l'initialisation du système, nous avons combiné les scores récoltés de la base de données *EmoWordNet* [246] avec un entraînement intensif sur l'ensemble des données d'entraînement (103 247 échanges) de la base de données dialogale *Alexa Topical-Chat* [241]. Puis pour l'évaluation, nous avons pris les données test (6069 échanges) de cette même base de données.

Pour l'initialisation/entraînement et les tests, nous procédons de manière similaire : l'*agent 1* identifié dans chaque conversation est considéré comme l'utilisateur humain et l'*agent 2* est considéré comme un agent conversationnel dont Philia modifie les répliques. Lors d'une itération du système à travers toutes les conversations des données test, nous gardons en mémoire les répliques de l'*agent 1*, les *répliques originales* de l'*agent 2*, les *répliques transformées* de l'*agent 2* par Philia et ces mêmes *répliques transformées corrigées* par le correcteur. Nous faisons d'ailleurs une comparaison entre les répliques transformées et les répliques transformées et corrigées dans la section 4.3.3 afin de mieux comprendre l'apport de ce dernier par

rapport à Philia sans correcteur.

Dans ce mémoire, nous sommes surtout intéressés à comprendre le rôle de la variable interne émotionnelle V (voir la section 3.5.1) et comment celle-ci influence le système dans sa globalité. Pour faciliter la compréhension de notre propos, nous réécrivons l'équation 3.24 ici :

$$V = \frac{\#Q - \alpha \cdot \#O}{\#U} \quad (3.24)$$

et rappelons que $\#Q$ représente le nombre de fois que le scoreur combinatoire reconnaît que la réplique générée par le système élicite l'émotion cible tout en s'assurant que l'usager la perçoit comme telle, $\#O$ représente le nombre de fois que le scoreur combinatoire reconnaît que la réplique générée par le système élicite les émotions opposées et $\#U$ représente le nombre d'échanges dans la conversation. L'hyper-paramètre α permet de pénaliser différemment Philia. Ce facteur peut être ajusté pour diminuer ou augmenter l'impact de la génération d'une réplique opposée à l'émotion cible du système. Rappelons aussi que nous avons émis l'hypothèse qu'une petite valeur pour α encouragerait une phase exploratoire plus importante, c'est-à-dire des réponses plus souples, alors qu'avec de grandes valeurs de α nous devrions avoir un système - si la base de données des scores est assez étoffée - qui performe bien, c'est-à-dire qui est capable d'injecter l'émotion cible.

Pour les deux contextes, nous faisons varier V de deux manières différentes. Tout d'abord nous laissons V évoluer librement, mais en fixant le facteur de pénalité α à 0, 3, 6 et 9. Nous avons choisi de limiter α à 9, car il est rare que les conversations contenues dans l'ensemble test dépassent dix échanges, donc augmenter ce facteur davantage n'aurait que peu d'effets sur la variable interne émotionnelle. Ensuite, nous fixons la variable interne émotionnelle pour des valeurs de 0 à 1, avec un pas de 0,1 pour mieux observer les comportements exploratoire et de renforcement du système. Nous nous arrêtons à $V = 1$, car avec cette valeur nous sommes déjà en mode exploratoire continuellement, c'est-à-dire que les probabilités de choisir un synonyme quelconque sont équiprobables.

Pour chaque combinaison d'hyper-paramètres testée, nous effectuons une itération sur les données test.

Notons que pour faire une évaluation automatique, nous devons retirer la partie interactive de Philia avec l'utilisateur. Comme nous ne pouvons plus compter sur une correction manuelle de ce dernier, nous avons aussi baissé la tolérance par rapport aux erreurs permises. Ainsi le seuil du nombre d'erreurs permises est passé de trois à une et le seuil du nombre de fausses erreurs permises est passé de trois à deux. Même si ces modifications sont mineures, nous

réécrivons l'algorithme 1 de la page 62 dans l'algorithme 3 pour faciliter sa compréhension.

Algorithm 3 Calculer le score combiné $T(N_0, N_1)$ (*modifié*), partie 1/2

Variables d'entrée :

N_0 = score individuel de la réplique du système

N_1 = score individuel de la réplique de l'utilisateur

Initialisation :

$\{N_0, N_1\} \in \mathbb{Z}$

$nombre_erreurs \leftarrow 0$

$nombre_erreurs_permis \leftarrow 1$

$nombre_fausses_erreurs \leftarrow 0$

$nombre_fausses_erreurs_permis \leftarrow 2$

les différents cas $\{N_0, N_1\}$:

if $N_0 \geq 1$ and $N_1 \geq 1$ **then**

 // Le système et l'utilisateur sont en accord :

$T(N_0, N_1) \leftarrow 1$

else if $N_0 \geq 1$ and $N_1 == 0$ **then**

 // Le système et l'utilisateur sont moyennement en accord :

$T(N_0, N_1) \leftarrow 1$

else if $N_0 \geq 1$ and $N_1 \leq -1$ **then**

 // Le système et l'utilisateur ne sont pas en accord :

if $nombre_erreurs < nombre_erreurs_permis$ **then**

$T(N_0, N_1) \leftarrow 0$

$nombre_erreurs \leftarrow +1$

else if $nombre_erreurs \geq nombre_erreurs_permis$ **then**

$T(N_0, N_1) \leftarrow -1$

$nombre_erreurs \leftarrow 0$

end if

else if $N_0 == 0$, $N_1 \geq 1$ **then**

 // Le système et l'utilisateur sont moyennement en accord :

$T(N_0, N_1) \leftarrow 0$

else if $N_0 == 0$ and $N_1 == 0$ **then**

 // Le système et l'utilisateur sont neutres :

$T(N_0, N_1) \leftarrow 0$

Algorithm 4 Calculer le score combiné $T(N_0, N_1)$ (*modifié*), partie 2/2

```

else if  $N_0 == 0$  and  $N_1 <= -1$  then
    // Le système et l'utilisateur sont moyennement en désaccord :
    if nombre_erreurs < nombre_erreurs_permis then
         $T(N_0, N_1) \leftarrow 0$ 
        nombre_erreurs  $\leftarrow +1$ 
    else if nombre_erreurs  $\geq$  nombre_erreurs_permis then
         $T(N_0, N_1) \leftarrow -1$ 
        nombre_erreurs  $\leftarrow 0$ 
    end if
else if  $N_0 <= -1$  and  $N_1 >= 1$  then
    // Le système et l'utilisateur sont en désaccord sur la négativité perçue :
    if nombre_fausses_erreurs < nombre_fausses_erreurs_permis then
         $T(N_0, N_1) \leftarrow 0$ 
        nombre_fausses_erreurs  $\leftarrow +1$ 
    else if nombre_fausses_erreurs  $\geq$  nombre_fausses_erreurs_permis then
         $T(N_0, N_1) \leftarrow 1$ 
        nombre_fausses_erreurs_permis  $\leftarrow 0$ 
    end if
else if  $N_0 <= -1$  and  $N_1 == 0$  then
    // Le système et l'utilisateur sont moyennement en désaccord sur la négativité perçue :
    if nombre_fausses_erreurs < nombre_fausses_erreurs_permis then
         $T(N_0, N_1) \leftarrow 0$ 
        nombre_fausses_erreurs  $\leftarrow +1$ 
    else if nombre_fausses_erreurs  $\geq$  nombre_fausses_erreurs_permis then
         $T(N_0, N_1) \leftarrow -1$ 
        nombre_fausses_erreurs_permis  $\leftarrow 0$ 
    end if
else if  $N_0 <= -1$  and  $N_1 <= -1$  then
    // Le système et l'utilisateur sont en accord sur la négativité perçue :
     $T(N_0, N_1) \leftarrow -1$ 
end if

```

Nous définissons maintenant deux contextes différents suivant l'utilisation ou non de l'analyse sentimentale dans Philia. Rappelons que nous utilisons à la fois une analyse sentimentale et une analyse émotionnelle. La raison principale est que les deux outils ne sont pas très performants en général, mais aussi qu'ils se complètent bien. Dans la section 3.5.1, nous avons défini le score du scoreur individuel comme la combinaison convexe $N(i)$ d'un score émotionnel $\alpha(i)$ et d'un score sentimental $\beta(i)$ pour chacune des répliques i :

$$N(i) = \lambda_0 \alpha(i) + \lambda_1 \beta(i) \quad (3.16)$$

avec $\lambda_0 + \lambda_1 = 1$. Dans le premier contexte, nous définissons $\lambda_1 = 0,25$ et dans le deuxième contexte, $\lambda_1 = 0$.

4.3.1 Premier contexte, le cas $\lambda_1 = 0.25$

Afin d'alléger le texte, les résultats obtenus avec un système initialisé avec l'utilisation d'une analyse sentimentale seront présentés de manière condensée sous forme de tableau récapitulatif. Pour avoir accès à l'entièreté des résultats compilés, nous invitons le lecteur à consulter l'annexe A, page 119.

Le tableau 4.1 présente un résumé des résultats obtenus à l'aide de la métrique adaptée FED pour l'ensemble des évaluations effectuées pour le système initialisé avec un apport sentimental pondéré avec $\lambda_1 = 0,25$.

Définissons par $\text{FED}_c^+(I)$ le score obtenu par la métrique FED augmentée pour la caractéristique c pour un ensemble d'échanges I et notons par O_c les différences, en pourcentages, entre les scores de la métrique FED augmentée du résultat d'une transformation T obtenue avec Philia et du texte initial I pour une caractéristique c :

$$O_c = \frac{\text{FED}_c^+(T) - \text{FED}_c^+(I)}{|\text{FED}_c^+(I)|} \cdot 100 \quad (4.2)$$

Parfois, quand le contexte le permet, nous utiliserons O sans référencer explicitement la caractéristique c . Ce sont ces valeurs qui sont répertoriées dans le tableau 4.1 à partir de la troisième colonne.

TABLEAU 4.1 Résultats obtenus pour l'évaluation du système initialisé avec un apport sentimental égal à 0,25. Les différences, en pourcentage O , représentent la différence des scores obtenus par le système par rapport aux scores des conversations originales.

Caractéristiques:	Scores originaux	Différence (%), pénalité = 0		Différence (%), pénalité = 3		Différence (%), pénalité = 6		Différence (%), pénalité = 9		Différence (%), Variable interne = 0		Différence (%), Variable interne = 0.5		Différence (%), Variable interne = 1	
		Système	Système corrigé	Système	Système corrigé	Système	Système corrigé	Système	Système corrigé						
Intérêt	-0.502	4.84	2.44	4.80	2.63	5.08	2.80	5.04	2.82	4.91	2.42	5.10	2.70	5.18	2.95
Engagement	-0.625	-0.99	-2.33	-1.16	-2.86	-0.96	-2.10	-0.97	-2.03	-0.78	-2.09	-1.10	-2.33	-0.98	-2.18
Originalité	-0.328	9.47	6.35	9.40	6.93	9.68	6.83	9.70	7.08	9.11	6.25	9.60	6.54	9.62	6.91
Pertinence	-0.206	-10.09	-7.32	-10.16	-7.61	-9.73	-7.56	-9.92	-7.29	-9.77	-6.85	-10.25	-7.48	-10.14	-7.37
Compréhension interloc.	0.096	-138.41	-119.05	-141.09	-126.59	-140.13	-121.16	-141.32	-121.26	-138.05	-117.41	-143.60	-122.27	-142.98	-124.36
Sémaniquement approprié	0.374	-4.86	-2.56	-4.88	-2.56	-4.87	-2.76	-5.03	-2.89	-4.91	-2.54	-5.10	-2.97	-5.04	-2.88
Compréhension réplique	0.329	-10.54	-7.35	-10.59	-7.93	-10.99	-8.16	-11.09	-7.92	-10.66	-7.43	-11.25	-8.26	-11.17	-8.11
Fluidité	0.354	-9.06	-5.80	-9.27	-6.24	-9.19	-6.34	-9.44	-6.36	-9.11	-5.93	-9.54	-6.59	-9.44	-6.43
Concurrence	0.142	-2.58	-1.86	-3.11	-1.94	-3.11	-2.45	-2.98	-1.45	-3.30	-2.21	-3.10	-2.04	-3.19	-2.39
Récupération d'erreurs	-0.654	-1.66	-2.46	-1.91	-2.77	-1.44	-2.18	-1.53	-2.31	-1.58	-2.43	-1.65	-2.38	-1.60	-2.24
Constance	0.560	2.77	3.46	2.84	3.65	2.74	3.43	2.84	3.34	2.77	3.54	2.78	3.44	2.72	3.31
Diversité	0.392	-1.23	-0.03	-1.22	-0.22	-1.10	-0.29	-1.32	-0.49	-1.25	0.03	-1.38	-0.32	-1.35	-0.34
Profondeur	0.072	29.33	26.19	28.75	27.28	28.64	25.39	28.42	26.09	29.39	26.86	27.71	24.46	27.61	24.78
Sympathie	-0.382	-1.02	-2.72	-1.35	-2.85	-1.10	-2.53	-1.08	-2.42	-0.80	-2.50	-1.30	-2.63	-1.27	-2.69
Compréhension générale	0.008	-404.78	-325.54	-421.25	-362.96	-413.54	-339.81	-413.97	-334.71	-405.88	-323.79	-421.63	-341.86	-422.31	-352.96
Flexibilité	-0.099	-3.40	-6.32	-4.52	-7.49	-4.71	-6.49	-3.88	-6.33	-1.95	-4.59	-4.31	-7.23	-4.27	-6.98
Informatif	-0.245	9.34	6.11	9.18	7.07	9.19	6.61	9.17	6.96	8.90	5.92	9.46	6.70	9.38	6.80
Cultorisé	-0.574	4.95	2.40	4.84	2.64	4.90	2.69	4.92	2.87	4.98	2.58	4.88	2.67	4.94	2.74
Score individuel moyen	0.571	18.51	17.86	23.05	18.51	19.81	13.96	22.40	17.86	26.95	23.48	19.37	16.45	17.42	16.13
Précision emotion moyenne	0.506	20.15	19.41	24.18	22.34	19.41	16.48	24.91	24.18	30.53	28.45	20.15	19.78	18.07	17.24
Score combiné moyen	0.571	18.51	17.86	23.05	18.51	19.81	13.96	22.40	17.86	26.95	23.48	19.37	16.45	17.42	16.13

Tableau récapitulatif apport sentimental 0,25

Nous décrivons maintenant les éléments du tableau 4.1. Tout d’abord, nous retrouvons dans la toute première colonne l’ensemble des 18 caractéristiques dialogales de la métrique FED augmentées par nos trois caractéristiques supplémentaires concernant l’injection de l’émotion. Dans la colonne scores originaux, nous avons regroupé les scores reliés aux différentes caractéristiques dialogales évaluées par la métrique FED augmentée pour les conversations originales issues de l’ensemble test. Le domaine des scores d’évaluation pour les 18 premières caractéristiques se situe entre -1 et 1 inclusivement, 1 étant un score parfait. Seuls les scores associés au score individuel moyen, la précision sur la détection de l’émotion cible moyenne et le score combiné moyen se situent entre 0 et 1 inclusivement. Pour le reste des colonnes, nous affichons le pourcentage de différence O_c (4.2) pour chaque caractéristique dialogale, et ce pour toutes les itérations d’évaluation où nous faisons varier le facteur pénalisant de la variable interne émotionnelle de Philia. Les trois dernières colonnes présentent les résultats moyens des itérations d’évaluation associés à des valeurs fixes de la variable interne émotionnelle afin d’exagérer le comportement du système. Une variable interne émotionnelle de 0 indique au système d’être perpétuellement en mode renforcement et donc d’uniquement choisir les tuples remplaçables ayant les scores pondérés maximaux. Une variable interne émotionnelle de 1 indique au système d’être en mode exploratoire continuellement, c’est-à-dire que les probabilités de sélection des tuples remplaçables sont équiprobables. Nous avons choisi d’uniquement présenter les résultats associés aux valeurs statiques 0, 0,5 et 1 de la variable interne émotionnelle dans les tableaux principaux pour alléger la présentation des données, mais nous présentons aussi d’autres tableaux avec plus de détails par la suite. Finalement, chaque colonne est subdivisée en deux sous-colonnes : la colonne *système* qui reprend les résultats de Philia produits avant l’utilisation du correcteur et la colonne *système corrigé* qui reprend les résultats de Philia produits avec la correction de ces résultats par le correcteur.

Une graduation automatique des couleurs allant du vert vers le rouge permet d’identifier plus aisément, pour chacune des caractéristiques, les améliorations, en vert, et les détériorations, en rouge. Plus la couleur est vive, plus les changements apportés sont importants. De plus, pour chaque caractéristique évaluée, les meilleures améliorations obtenues par le système, toutes configurations confondues sont identifiées en **gras**, alors que les plus grandes détériorations sont identifiées en ***gras italique doublement souligné***.

Nous avons émis l’hypothèse qu’en remplaçant uniquement les tuples originaux d’une phrase par des tuples ayant des scores émotifs plus élevés, nous pouvions injecter l’émotion cible dans une phrase tout en minimisant l’impact d’une telle transformation sur le sens d’une phrase. En observant les résultats du tableau 4.1, nous observons exactement ce comportement. Commençons par l’injection d’émotion. Les changements positifs les plus marquants

par rapport aux scores originaux de la deuxième colonne se situent - comme on pouvait s'y attendre - dans les trois caractéristiques ajoutées, soient le score individuel moyen (amélioration de 13,96% à 26,95%), la précision quant à la détection de l'émotion cible (amélioration de 16,48% à 30,53%) ainsi que le score combiné moyen par conversation (amélioration de 13,96% à 26,95%). Les résultats pour le score individuel moyen ainsi que pour le score combiné moyen sont quasi identiques. En effet, il n'y a presque aucune dissonance émotionnelle entre les interlocuteurs dans l'ensemble test de la base de données dialogale *Alexa Topical-Chat*. Ensuite, nous remarquons que l'ensemble des caractéristiques dialogales de la métrique FED sont relativement stables avec approximativement des changements absolus de maximum 10%. Trois caractéristiques originales de la métrique FED sortent du lot : la compréhension des interlocuteurs (détérioration de -323,79% à -422,31%), la compréhension générale (détérioration de -117,41% à -143,60%) et la profondeur (amélioration de 24,46% à 29,39%). Ces changements conséquents doivent sans doute être pris avec une certaine suspicion, car ce sont aussi les caractéristiques qui ont les scores originaux les plus bas et donc toute légère variation se traduit par de plus grands pourcentages que pour les autres caractéristiques.

Dans la première partie du tableau, quand nous laissons au système le soin de mettre la variable émotionnelle à jour automatiquement, nous n'observons pas de corrélation entre le facteur pénalisant α et la qualité de l'injection des émotions. Comme mentionné auparavant, il y a très peu de dissonances dans le jeu de données test et donc nous ne possédons pas assez de données pour analyser une éventuelle dépendance par rapport à ce facteur. Notons toutefois que, dans la deuxième partie du tableau, celle où la variable interne émotionnelle est fixée, nous pouvons observer une corrélation directe entre la valeur de cette variable et la qualité de l'injection d'émotion : plus la variable interne émotionnelle est basse, meilleure est la qualité de l'injection d'émotions. En effet, plus la variable interne émotionnelle est basse, plus le système est forcé à choisir des synonymes qui ont un score émotionnel maximal et plus elle est haute, plus le choix de ces synonymes est fait de manière aléatoire. Nous illustrons plus en détails ce phénomène pour un ensemble plus fourni de valeurs de V sur la figure 4.1 avec la droite correspondante à la régression linéaire.

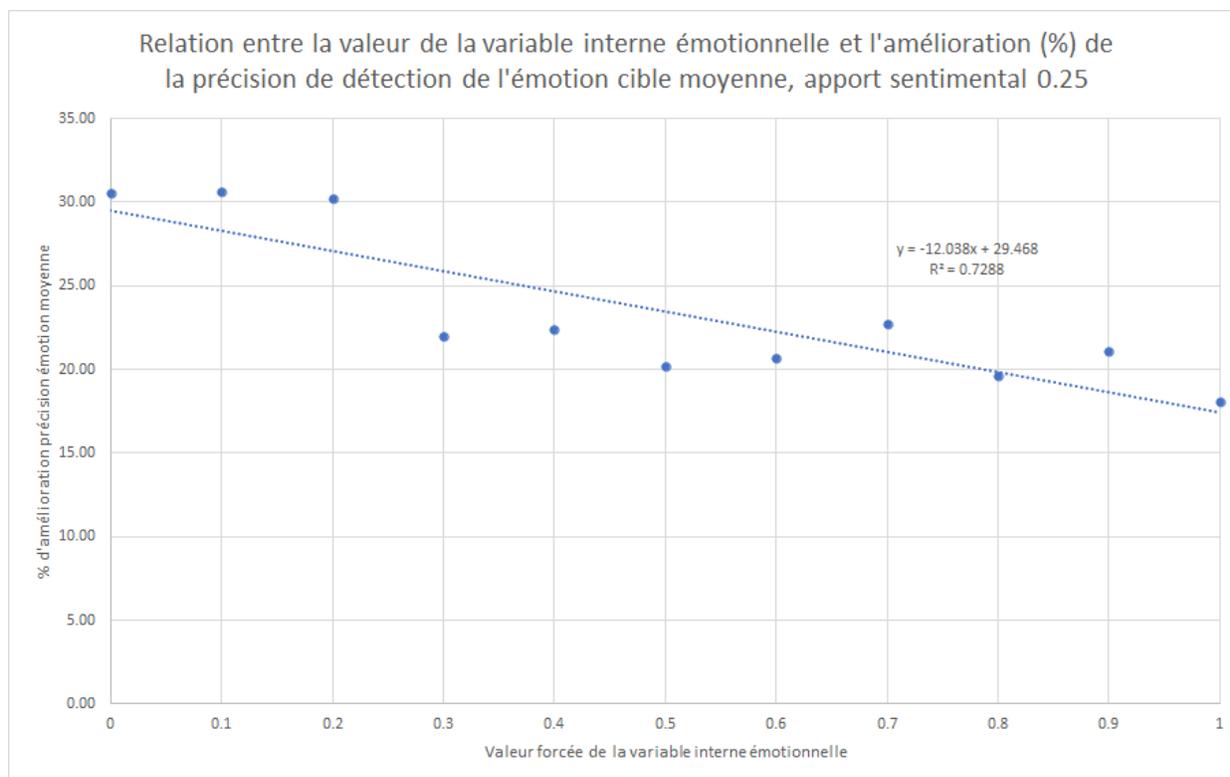


FIGURE 4.1 Relation entre la valeur de la variable interne émotionnelle et la précision de détection de l'émotion cible moyenne. Système initialisé en pondérant l'apport sentimental au score individuel des répliques à 0,25.

Remarquons sur la figure 4.1 un plateau pour les valeurs de la variable interne émotionnelle fixée à 0, 0,1 et 0,2. Ce plateau est dû à la manière dont la variable interne émotionnelle affecte λ dans l'équation 3.23 que nous reproduisons ici :

$$\lambda = \log(a \cdot V) \quad (3.23)$$

Rappelons que nous avons choisi de manière empirique $a = 5$ ce qui fait que pour les valeurs de $V \leq 0,2$, nous fixons empiriquement $\lambda = 0,01$. Cette simplification au niveau du calcul de λ n'affecte pas la performance du système, car nous avons observé que les échanges entre interlocuteurs sont généralement en consonance émotive et, de ce fait, la valeur de la variable interne émotionnelle se trouve rapidement hors de ce domaine particulier.

Finalement, remarquons l'effet du correcteur sur l'injection des émotions dans le tableau 4.1. Systématiquement, ce dernier détériore la qualité de l'injection. Nous reprendrons plus en détails le correcteur et son éventuelle influence dans la section 4.3.3.

Dans le cadre de ce premier contexte, nous concluons que dans l'ensemble Philia injecte bel et bien l'émotion cible dans les répliques originales tout en minimisant l'impact des transformations sur la qualité générale du dialogue.

4.3.2 Deuxième contexte, le cas $\lambda_1 = 0$

Comme pour la sous-section 4.3.1, les résultats obtenus avec un système initialisé sans analyse sentimentale seront présentés de manière condensée sous forme de tableau récapitulatif. Pour avoir accès à l'entièreté des résultats compilés, nous invitons le lecteur à consulter l'annexe B.

Intuitivement, on pourrait penser que, si nous reprenons le premier contexte (section 4.3.1) à l'identique, sauf pour la suppression de l'analyse sentimentale, nous obtiendrions de meilleurs résultats. Hors il n'en est rien, comme on peut le constater dans le tableau 4.2 qui ne reprend que les valeurs fixées pour la variable interne émotionnelle. Une explication pourrait être que les outils que nous utilisons sont sensiblement les mêmes pour les analyses émotionnelles et sentimentales dans le cas de la joie³.

TABLEAU 4.2 Résultats obtenus pour O_c en forçant la valeur de la variable interne émotionnelle égale à 0, 0,5, et 1. Système initialisé en pondérant l'apport sentimental au score individuel des répliques à 0.

Tableau récapitulatif: apport sentimental nul, seuils de détection ajustés pour le scoreur individuel							
		Différence (%), Variable Interne = 0		Différence (%), Variable Interne = 0.5		Différence (%), Variable Interne = 1	
Caractéristiques:	Scores originaux	Système	Système corrigé	Système	Système corrigé	Système	Système corrigé
Intérêt	<u>-0.503</u>	3.96	<u>1.63</u>	4.57	2.78	5.04	3.01
Engagement	<u>-0.625</u>	-2.13	<u>-3.03</u>	-2.07	-2.52	-1.76	-2.47
Originalité	<u>-0.328</u>	8.39	<u>5.55</u>	9.06	6.14	9.78	6.54
Pertinence	<u>0.206</u>	-9.65	<u>-6.56</u>	-10.76	-8.38	-11.00	-8.51
Compréhension interloc.	<u>0.036</u>	-147.77	-123.10	-152.53	-131.57	<u>-157.87</u>	-135.37
Sémantiquement approprié	<u>0.374</u>	-3.72	<u>-1.70</u>	-4.17	-2.62	-4.55	-2.81
Compréhension réplique	<u>0.329</u>	-9.80	<u>-6.50</u>	-10.53	-8.17	<u>-11.23</u>	-8.31
Fluidité	<u>0.354</u>	-8.22	<u>-5.03</u>	-8.65	-6.35	-9.31	-6.61
Cohérence	<u>0.142</u>	-1.87	<u>-1.47</u>	-2.44	-2.91	-3.12	-2.90
Récupération d'erreurs	<u>-0.654</u>	-2.41	<u>-2.90</u>	-2.20	-2.56	-2.05	-2.37
Constance	<u>0.560</u>	3.35	3.66	3.27	3.51	2.93	3.06
Diversité	<u>0.392</u>	-0.86	0.20	-1.14	-0.17	-1.29	-0.59
Profondeur	<u>0.077</u>	30.25	25.28	26.39	22.23	26.59	22.26
Sympathie	<u>-0.382</u>	-2.47	<u>-3.54</u>	-2.47	-3.39	-2.16	-3.12
Compréhension générale	<u>0.008</u>	-433.47	-338.21	-443.72	-377.99	<u>-456.07</u>	-379.63
Flexibilité	<u>-0.099</u>	-4.57	<u>-6.14</u>	-6.75	<u>-8.96</u>	-6.30	-7.82
Informatif	<u>-0.245</u>	9.33	6.30	10.01	7.28	10.32	7.51
Curiosité	<u>-0.574</u>	3.98	<u>1.97</u>	4.11	2.29	4.55	2.70
Score individuel moyen	<u>0.579</u>	28.46	31.62	19.37	20.55	18.97	15.42
Précision émotion moyen	<u>0.506</u>	30.04	28.94	20.88	23.44	17.58	17.22
Score combiné moyen	<u>0.579</u>	28.46	31.62	19.37	20.55	18.97	15.42

3. En effet, la polarité d'une réplique et l'émotion qu'elle élicite sont intimement liées dans ce cas-ci, car non seulement les mêmes données ont servi à entraîner les classificateurs sentimentale et émotionnel du système, mais une polarité positive et l'émotion cible de la joie sont directement liées.

Vu la similarité des résultats entre les deux contextes, nous avons introduit un changement supplémentaire dans ce deuxième contexte. Comme il est question - dans un développement ultérieur - de laisser le soin à Philia de déterminer par lui-même ce que représente une émotion et comment la caractériser, c'est-à-dire laisser à Philia l'optimisation de certains hyper-paramètres par lui-même, nous nous sommes intéressés à la variation des paramètres définissant une émotion. Dans ce deuxième contexte, nous nous sommes intéressés à ce qui se passerait si on changeait la caractérisation d'une émotion entre la phase d'initialisation/entraînement et la phase test/détection. En particulier, que se passerait-il si on changeait le seuil de reconnaissance d'une émotion entre les deux phases ?

Reprenons la définition de la fonction indicatrice d'une émotion \mathcal{E} proposée à la page 60 :

$$E(i, \mathcal{E}) = \begin{cases} 1 & \text{si } P(\mathcal{E}|i) > \Phi_{\mathcal{E}}, \\ 0 & \text{sinon.} \end{cases} \quad (3.18)$$

pour un certain seuil $\Phi_{\mathcal{E}}$ donné avec i une réplique donnée et \mathcal{E} représentant la joie. Dans ce deuxième contexte, nous fixerons ce seuil à 30% pour l'entraînement, mais à 50% pour les tests. Dans le premier contexte, les seuils étaient les mêmes (50%) pour les deux phases. Philia étant donc plus permissive à l'entraînement qu'à la détection, nous nous attendons en conséquence à une dégradation des résultats pour l'injection d'émotion, et c'est exactement ce que nous observons.

Le tableau 4.3 rassemble les données obtenues dans ce deuxième contexte. Afin de mieux saisir la présentation des résultats, nous invitons le lecteur à se référer au paragraphe décrivant le tableau 4.1, sous-section 4.3.1.

TABLEAU 4.3 Résultats obtenus pour l'évaluation du système initialisé avec un apport sentimental nul mais avec des seuils de reconnaissance Φ_e différents suivant les phases d'initialisation et de détection. Les différences, en pourcentage, représentent la différence O des scores obtenus par le système par rapport aux scores des conversations originaux.

Caractéristiques:	Scores originaux	Différence (%), pénalité = 0		Différence (%), pénalité = 3		Différence (%), pénalité = 6		Différence (%), pénalité = 9		Différence (%), Variable Interne = 0		Différence (%), Variable Interne = 0.5		Différence (%), Variable Interne = 1	
		Système	Système corrigé	Système	Système corrigé	Système	Système corrigé	Système	Système corrigé						
Intérêt	<u>-0.503</u>	6.18	3.76	5.92	3.52	6.13	3.76	6.13	3.38	5.86	3.64	5.81	3.71	5.82	<u>3.31</u>
Engagement	<u>-0.625</u>	-0.40	-1.64	-0.53	-1.50	-0.32	-1.60	-0.19	-1.65	-0.29	-1.47	-0.44	-1.29	-0.17	<u>-1.25</u>
Originalité	<u>-0.328</u>	10.43	7.53	10.29	7.27	10.17	7.21	10.27	<u>6.44</u>	9.79	7.27	9.95	6.78	9.82	6.89
Profondeur	<u>0.206</u>	-10.54	-7.47	-10.61	-7.79	-10.77	-7.82	-10.76	-7.79	-10.04	<u>-6.84</u>	-10.40	-7.50	-10.08	-7.27
Compréhension Interloc.	<u>0.036</u>	-141.28	-118.89	-142.34	-117.77	-143.02	-124.74	-139.30	-118.33	-134.09	-112.97	-139.80	-119.35	-138.06	-124.58
Sémantiquement approprié	<u>0.374</u>	-4.95	-2.70	-4.96	-2.81	-5.23	-3.08	-5.28	-3.09	-4.96	-2.67	-5.20	-3.13	-4.84	-2.58
Compréhension réplique	<u>0.329</u>	-10.82	-7.69	-10.91	-7.74	-11.16	-8.01	-10.92	-7.87	-10.40	-7.16	-10.84	-7.86	-10.46	-7.51
Fluidité	<u>0.354</u>	-9.07	-6.00	-9.08	-6.04	-9.35	-6.36	-9.29	-6.36	-8.77	-5.62	-9.09	-6.24	-8.93	-6.02
Conférence	<u>0.142</u>	-4.54	-3.21	-4.02	-3.21	-4.34	-3.68	-3.95	-2.51	-4.36	-3.42	-4.25	-3.61	-3.00	-2.77
Récupération d'erreurs	<u>-0.654</u>	-0.78	-1.57	-1.00	-1.57	-0.82	-1.69	-0.74	-1.71	-0.53	-1.15	-1.02	-1.58	-0.79	<u>-1.12</u>
Constance	<u>0.580</u>	2.26	2.72	2.34	2.72	2.33	2.80	2.28	2.80	2.45	2.60	2.36	2.59	2.34	2.34
Diversité	<u>0.392</u>	-1.67	-0.76	-1.52	-0.52	-1.77	-0.65	-1.43	-0.55	-1.55	-0.46	-1.41	-0.51	-1.82	-0.58
Profondeur	<u>0.077</u>	23.86	21.92	23.34	<u>19.47</u>	24.15	21.14	24.13	20.83	23.41	21.65	23.91	21.36	24.95	23.79
Sympathie	<u>-0.382</u>	-0.90	-2.23	-0.95	-2.01	-0.80	-2.01	-0.69	-2.33	-0.77	-1.79	-1.09	-2.16	-0.67	<u>-2.59</u>
Compréhension générale	<u>0.008</u>	-406.35	-327.91	-412.41	-329.61	-419.63	-349.97	-399.61	-335.22	-382.65	-299.48	-405.55	-333.86	-393.03	-334.80
Flexibilité	<u>-0.099</u>	-4.38	-6.37	-5.07	-7.53	-5.30	-7.39	-3.37	-5.49	-2.50	-4.06	-4.79	-6.46	-3.69	-7.45
Informatif	<u>-0.245</u>	9.28	<u>5.87</u>	9.44	5.94	9.83	6.51	9.99	6.33	9.20	5.97	9.59	6.17	9.25	6.48
Curiosité	<u>-0.574</u>	4.88	2.79	4.78	2.76	5.02	2.75	5.13	2.95	4.98	2.82	4.82	2.68	4.95	2.58
Score Individuel moyen	<u>0.579</u>	16.35	15.38	17.31	14.10	15.06	12.18	17.31	14.10	20.83	19.23	13.78	13.46	16.03	13.74
Précision emotion moyen	<u>0.506</u>	19.05	20.15	18.68	16.12	19.05	16.48	22.34	18.32	24.54	24.91	16.85	15.02	19.78	19.78
Score combiné moyen	<u>0.579</u>	16.35	15.38	17.31	14.10	15.06	12.18	17.31	14.10	20.83	19.23	13.78	13.46	16.03	13.74

Tableau récapitulatif: apport sentimental nul

Les résultats du tableau 4.3 sont relativement similaires à ceux obtenus dans le tableau correspondant 4.1 du premier contexte. L'injection de l'émotion - donnée dans les trois caractéristiques rajoutées à la métrique FED - est améliorée, et non seulement la plupart des caractéristiques de la métrique FED sont relativement inchangées (fluctuation d'environ 10% au maximum), mais ce sont aussi exactement les mêmes caractéristiques dialogales qui ressortent du lot. Il est toutefois intéressant de noter que le changement de seuil entre la phase d'initialisation et la phase de détection a amoindri l'injection de l'émotion. Ainsi on passe d'une amélioration de 13,96% à 26,95% pour les scores individuel et combiné moyens dans le premier contexte à seulement une amélioration de 12,18% à 20,83%, et d'une amélioration de 16,48% à 30,53% pour la détection de l'émotion cible dans le premier contexte à seulement une amélioration de 15,02% à 24,91% dans ce deuxième contexte. Ces résultats sont attendus et peuvent se comprendre aisément. En définissant le seuil $\Phi_{\mathcal{E}} = 0,3$ pendant la phase d'initialisation/entraînement, nous avons dans un sens affaibli la sévérité avec laquelle notre système considère une réplique comme transmettant l'émotion cible, et du même fait, nous comptabilisons des scores émotionnels plus faibles. Cet affaiblissement est directement détecté dans les scores plus faibles dans le tableau 4.3. Ce qui est intéressant, c'est que, même en considérant une valeur relativement faible de $\Phi_{\mathcal{E}}$ pendant l'initialisation/entraînement, Philia est quand même capable d'injecter de l'émotion de manière probante. C'est peut-être une première indication de la possibilité pour Philia d'optimiser par lui-même non seulement les seuils, mais aussi de définir par lui-même ce qu'il entend par une émotion en tâtonnant avec des hyper-paramètres, tout en conservant sa capacité à injecter de l'émotion.

La figure 4.2 est la figure équivalente pour la figure 4.1 du contexte 1. Remarquons ici aussi une corrélation directe entre V et la qualité de l'injection des émotions. On peut aussi y apercevoir le plateau pour les premières valeurs de V que nous avons décrit plus haut pour le contexte 1.

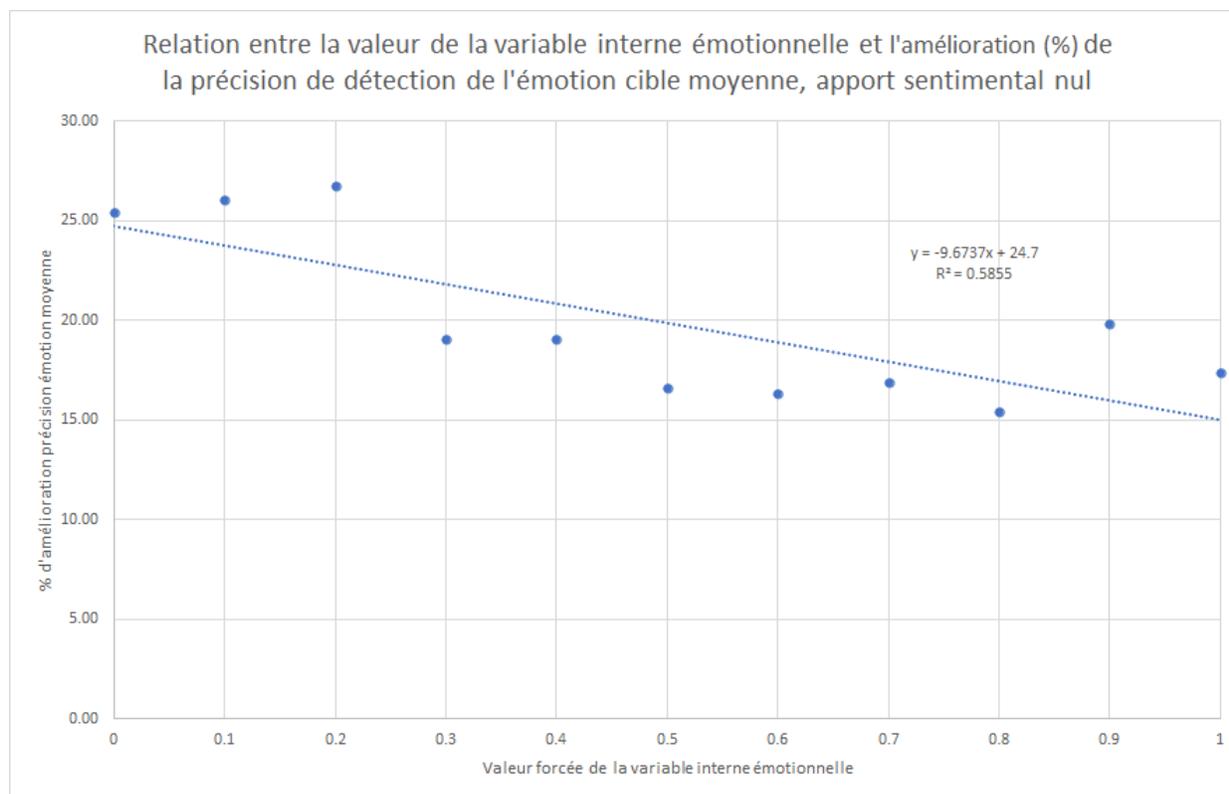


FIGURE 4.2 Relation entre la valeur de la variable interne émotionnelle et la précision de détection de l'émotion cible moyenne. Système initialisé en pondérant l'apport sentimental au score individuel des répliques à 0.

4.3.3 Impact du correcteur et incohérence sémantique

Dans cette section, nous analysons à l'aide d'exemples concrets de quelle manière l'utilisation du correcteur affecte les résultats obtenus ainsi que les cohérences syntaxique et sémantique des résultats.

Commençons par traiter le cas du correcteur. Tout d'abord, rappelons que le correcteur n'est utilisé que pour transmettre un message corrigé à l'utilisateur et n'affecte d'aucune manière directe la mise à jour des scores de la base de données de Philia qui est effectuée avant la correction.

Notons qu'en général, l'impact négatif que les transformations ont sur les scores associés à la qualité conversationnelle est amoindri après l'utilisation du correcteur grammatical du système et, du même coup, nous observons que les scores associés à l'injection émotive se trouvent à être généralement diminués après correction. La correction semble impacter généralement positivement les scores associés à la pertinence, la compréhension interlocuteurs,

le fait d’être sémantiquement approprié, la compréhension de la réplique, la fluidité, la cohérence, la constance et la compréhension générale, tandis qu’elle impacte négativement en général l’intérêt, l’engagement, l’originalité, la récupération d’erreurs, la profondeur, la sympathie, la flexibilité, le fait d’être informatif, la curiosité et - ce qui nous intéresse le plus dans ce mémoire - tous les scores associés à la détection de l’émotion cible. Il est difficile d’expliquer précisément la raison pour laquelle nous observons de tels changements au niveau des scores entre les répliques transformées par le système et celles qui sont corrigées pour les 18 premières caractéristiques. Premièrement, à cause de la nature subjective des caractéristiques évaluées et, deuxièmement, à cause de la nature des outils que nous utilisons pour effectuer les corrections et l’évaluation. Nous illustrons l’effet du correcteur dans le tableau 4.4 par quatre exemples issus des données test. Nous avons choisi ces exemples - relativement complexes - non seulement pour illustrer l’effet du correcteur, mais aussi pour mettre en évidence la difficulté d’améliorer l’injection d’une émotion tout en conservant une sémantique satisfaisante. De fait, Philia n’est pour l’instant programmé que pour améliorer un score émotif sans tenir compte d’une quelconque cohérence syntaxique ou sémantique. Nous reprenons les dénominations *original* pour la réplique originale dans les données test, *système* pour les répliques transformées par Philia et finalement *système corrigée* pour les répliques transformées par Philia et puis corrigées par le correcteur.

Les quatre extraits de conversations présentés dans le tableau 4.4 sont issus de l’évaluation du système avec une variable interne émotionnelle forcée à 0, initialisé avec un apport sentimental à 0,25 (contexte 1). Les mots considérés lors de l’optimisation par le système sont écrits en **gras**, les mots choisis après optimisation sont écrits en ***gras italique***, les mots corrigés par le correcteur grammatical sont en ***gras italique souligné*** et finalement lorsque le correcteur altère la structure de la phrase, ces changements sont écrits en ***gras italique rouge souligné***.

TABLEAU 4.4 Philia en action sur quatre exemples extraits des conversations utilisées pendant son évaluation.

<i>agent 1</i>	"Did you know that the University of Iowa's locker room is painted pink? I wonder why?"
<i>agent 2 : original</i>	"I think I did hear something about that. I imagine it is an attempt to psych the other team out."
<i>agent 2 : système</i>	"I believe I did read mind about that. I envision it is an tryout to strengthen the other set out."
<i>agent 2 : système corrigé</i>	"I believe I did read your mind on that as I envision it to be a tryout to strengthen the other set ."
<i>agent 1</i>	"No. I don't have regular cable. I use a ROKU and I think you would have to pay extra for ESPN. I didn't watch it when I did have cable though."
<i>agent 2 : original</i>	"Oh, I was going to tell you about them winning an emmy but that's okay . Well , do you watch or have you watched pokemon?"
<i>agent 2 : système</i>	"Oh, I was whirling to confess you about them winning an honor but that 's good . Right , do you watch or have you seen pokemon?"
<i>agent 2 : système corrigé</i>	"Oh, I was whirling to confess you about them winning an honor , but that s good . Right , do you watch or have you seen pokemon?"
<i>agent 1</i>	"Do you happen to know why he won the emmy?"
<i>agent 2 : original</i>	" apparently for the work he did calling the games . He was inducted into the sportscasters hall of fame too."
<i>agent 2 : système</i>	" Professedly for the aid he did addressing the sports . He was seat into the writers pass of renown too."
<i>agent 2 : système corrigé</i>	" Professedly for the aid he did addressing the sports and sat into the writers pass of renown too."
<i>agent 1</i>	"I imagine this number in the millions taking into account he had a 54 year broadcasting career, so sad, but he was old."
<i>agent 2 : original</i>	"I like that he used this phrase alot Whoa, Nellie!"
<i>agent 2 : système</i>	"I like that he worked this catchphrase wholesale Whoa, Nellie!"
<i>agent 2 : système corrigé</i>	"I like that he worked this catchphrase wholesale Whoa, Nellie!"

Les exemples du tableau 4.4 nous permettent de mieux comprendre la détérioration obtenue dans l'injection d'émotions. En effet, le correcteur grammatical, construit à base d'un transformateur générateur, ne se contente pas d'uniquement corriger les mots mal écrits, mais altère également légèrement la structure de la réplique. Cette altération est probablement à l'origine des disparités entre les scores obtenus lors de l'évaluation du système. Remarquons également que ces altérations tentent de rendre les répliques transformées plus lisibles et fluides, ce qui concorde avec la légère amélioration au niveau des scores associés à la compréhension interlocuteurs, le fait d'être sémantiquement approprié, la fluidité et la compréhension globale du dialogue.

En observant les extraits du tableau 4.4, nous pouvons voir comment Philia transforme maladroitement les répliques. Les transformations successives donnent l'impression d'assister au jeu du *passe-parole* pendant lequel une phrase est répétée de bouche à oreille et se retrouve successivement transformée au point de ne plus ressembler à la phrase originale. Ce résultat ou plutôt ce manque de résultats et cette incohérence sémantique, mais aussi syntaxique est attendue : Philia optimise globalement l'injection d'émotion, mais en changeant ponctuellement des mots sans tenir compte aucunement d'une sémantique globale ou d'un contexte plus général.

Dans le premier exemple, les mots *hear* et *something* sont remplacés par les mots *read* et *mind* respectivement modifiant carrément la compréhension syntaxique et - de manière tout aussi sinon plus importante - sémantique. Dans le deuxième exemple, le verbe *going* est remplacé par le verbe *whirling* et *enemy* devient *honor*, *okay right* et *Well Right*. Le troisième exemple est encore plus incompréhensible. On peut se demander si la complexité des répliques ne viendrait pas embrouiller Philia, mais le quatrième exemple, relativement simple, montre que la complexité n'est pas la seule en cause pour justifier les résultats obtenus.

Plusieurs explications permettent de mieux comprendre les résultats actuels. Tout d'abord, le système a été entraîné à partir de dialogues figés et non pas adaptés par Philia pendant les discussions. Ensuite, Philia récolte automatiquement les synonymes (et antonymes) sur internet sans aucune intervention humaine. En particulier, les synonymes ne sont pas validés. Dans la pratique, nous nous retrouvons avec des listes assez longues de synonymes éparses ayant parfois des liens bien distendus. Nous n'avons pas non plus pu utiliser le mécanisme d'intervention humaine pour corriger les erreurs (*human in the loop*). Mentionnons aussi le manque de contexte plus général. Il n'y a pas de contexte sémantique et le contexte syntaxique est bien trop basique pour diriger convenablement la recherche de synonymes et de leur remplacement.

Dans la section suivante, nous nous attardons à la construction des listes de synonymes.

4.3.4 Impact de la liste des synonymes et cohérence sémantique

Nous l'avons vu dans la section précédente, les résultats obtenus après transformations ne sont pas sémantiquement satisfaisants, voire même valides. Parmi les raisons évoquées, nous avons mentionné les longues listes de synonymes qui sont collectées automatiquement sans tri préalable ni validation. Sans doute faudrait-il développer un mécanisme capable de tirer profit de l'information associée au contexte syntaxique et/ou sémantique afin de remplacer les mots originaux par des synonymes contextuellement plus appropriés. Une autre approche serait de rectifier le lexique de sorte à ne conserver que les synonymes et antonymes les plus couramment utilisés. Dans cette section, nous présentons des exemples simples avec des listes de synonymes rectifiées et validées manuellement. Cette simplification et validation permet de mieux comprendre le potentiel de Philia. Nous renvoyons le lecteur à la section précédente pour l'utilisation des marqueurs dans le texte.

Le tableau 4.5 regroupe deux exemples simples avec l'utilisation par Philia de tous les synonymes récoltés automatiquement.

TABLEAU 4.5 Exemples de répliques simples en utilisant tous les synonymes.

agent 2 : original	"I'm <i>doing ok</i> , it's <i>thanks</i> to you."
agent 2 : système	"I'm <i>doing good</i> , it 's <i>thanks</i> to you."
agent 2 : système corrigé	"I'm <i>doing well</i> , it 's <i>thanks</i> to you."
agent 2 : original	"I <i>like</i> the <i>flowers</i> ."
agent 2 : système	"I <i>like</i> the <i>spikes</i> ."
agent 2 : système corrigé	"I <u><i>like</i></u> the <u><i>spikes</i></u> ."

Si le premier exemple fonctionne relativement bien et dénote bien une injection de joie, le deuxième exemple illustre bien le problème des synonymes inappropriés. Le substantif *flowers* est remplacé par le mot *spikes*. La liste de synonymes pour le substantif *flower* est constituée de 104 entrées ! Le verbe *like* quant à lui possède 40 entrées, mais pas le verbe *love* !

Le tableau 4.6 regroupe les mêmes trois exemples que le tableau 4.5, mais cette fois-ci avec l'utilisation de listes de synonymes validées manuellement. Nous avons gardé l'entraînement fait avec les synonymes collectés automatiquement, mais nous avons considérablement réduit manuellement la liste des synonymes en essayant de simuler une validation automatique imparfaite. Nous avons aussi rajouté le verbe *love* comme synonyme de *like*. Pour le deuxième exemple, nous n'avons gardé que cinq synonymes pour *like* (*love*, *approve*, *enjoy*, *relish* et *affect*) et neuf synonymes pour *flower* (*bloom*, *blossom*, *dahlia*, *daisy*, *flush*, *herb*, *mistflower*,

orchid et sunflower).

TABLEAU 4.6 Les mêmes exemples que dans le tableau 4.5, cette fois-ci en utilisant une liste de synonymes validée manuellement.

agent 2 : original	"I'm doing ok, it's thanks to you."
agent 2 : système	"I'm <i>doing good</i> , it 's <i>thanks</i> to you."
agent 2 : système corrigé	"I'm <u>doing well</u> , it 's <u>thanks</u> to you."
agent 2 : original	"I like the flowers."
agent 2 : système	"I <i>love</i> the <i>flowers</i> ."
agent 2 : système corrigé	"I <u>love</u> the <u>flowers</u> ."

Cette fois-ci nous obtenons - bien que de manière un peu arbitraire - des résultats nettement plus cohérents. Ce deuxième exemple illustre bien l'importance d'une base de données avec des synonymes (et antonymes) validés de manière contextuelle.

CHAPITRE 5 CONCLUSION ET RECOMMANDATION

Dans ce mémoire, nous avons présenté Philia, un système combinant la recherche opérationnelle et l'apprentissage automatique pour injecter de l'émotion dans des répliques données en modifiant le choix de certains mots utilisés. C'est - à notre connaissance - le seul système de ce type qui combine de la sorte la recherche opérationnelle et l'apprentissage machine. Nous sommes au tout début de telles approches, mais nous estimons que les premiers résultats obtenus dans ce mémoire sont probants et révélateurs du potentiel de cette combinaison, en particulier, telle que nous l'avons présentée.

Reprenons les quatre sous-objectifs que nous nous sommes donnés en début de ce mémoire à la section 1.2 :

S-O1 : Nous avons proposé une métrique capable de mesurer la consonance et la dissonance émotives entre l'utilisateur et le système à l'aide du scoreur combinatoire présenté à la section 3.5.1. De plus, nous avons adapté une métrique d'évaluation dialogale, la métrique FED, qui nous permet de mesurer automatiquement la qualité du dialogue généré et de mesurer dans quelle mesure nous arrivons à injecter de l'émotion dans des répliques existantes. Notre adaptation de la métrique FED est présentée dans la section 4.2.

S-O2 : Nous avons développé une base de données contextuelle capable de recenser l'impact des mots sur l'élicitation d'une émotion donnée. Cette base de données est introduite dans la section 3.3.2.

S-O3 : Notre troisième sous-objectif, qui était de créer ou adapter un algorithme de reconnaissance émotive et sentimentale a été partiellement atteint, car nous n'avons pas créé nous-mêmes un nouvel algorithme, mais avons utilisé une implémentation traduite en Pytorch pour créer nos classificateurs émotionnels et sentimentaux. Ces classificateurs sont introduits dans la section 3.2.2.

S-O4 : Nous avons développé et implémenté un nouvel algorithme de génération de réponses capable de transmettre de l'émotion en étant en concordance ou discordance émotive avec l'utilisateur. Philia fait l'objet du chapitre 3.

Philia est un projet de recherche innovateur. Il n'existe pas, à notre connaissance, de projet similaire basé sur une approche d'hybridation de la recherche opérationnelle avec l'apprentissage automatique telle que nous l'avons proposée. Nous sommes donc sur un terrain nouveau et, disons-le, prometteur.

5.1 Limitations de la solution proposée

Dans cette thèse, nous avons construit une preuve de concept de Philia. Nous avons dû faire des choix et ceux-ci ont forcément limité la portée du projet. Nous discutons ici de quelques-unes des limitations de la solution proposée. Nous les développons bloc par bloc.

Pour le premier bloc de Philia, le bloc des *agents conversationnels* présenté à la section 3.1, nous avons dû nous limiter à utiliser des implémentations d'agents conversationnels existants au lieu de créer nous-mêmes notre propre agent et les utiliser comme des boîtes noires. Il s'agit d'une limitation dans la mesure où nous pourrions intégrer Philia et les agents conversationnels en un seul système qui bénéficierait des connaissances apprises pour générer des réponses plus adaptées avec des émotions mieux ciblées et un contrôle accru.

Pour le deuxième bloc, le bloc de *compréhension du langage naturel* présenté à la section 3.2, lors de la décomposition des répliques, nous n'exploitons pas la pleine capacité de la librairie spaCy pour tenir compte des liens inter-mots d'une phrase. Le système dans son état actuel ne fait que collecter les racines des mots d'une phrase et leurs classes syntaxiques associées. Pour avoir un système capable d'effectuer des transformations plus robustes, il est nécessaire de tenir compte des relations entre les mots d'une phrase. Premièrement, cela nous permettrait d'affiner la sélection des tuples de remplacement lors de l'optimisation. Nous aurions ainsi une meilleure idée du contexte syntaxique dans lesquels ces mots sont employés, ce qui nous donnerait une information supplémentaire quant au sens des mots employés. Comme un seul mot peut avoir plusieurs sens, cela peut soulever une ambiguïté lors du choix des tuples de remplacement. Deuxièmement, en tenant compte des relations des mots dans une phrase, nous pourrions améliorer la reconstruction de celle-ci. Pour l'instant, la reconstruction ne dépend que de la racine et de la classe syntaxique du mot sans tenir compte de sa relation inter-mots dans la phrase. De manière plus générale, il y a aussi un manque de contexte sémantique et notamment un manque de suivi de l'historique du dialogue qui pourrait grandement aider à mieux cibler les transformations.

Les limitations du bloc *vocabulaire* présenté à la section 3.3 sont en partie similaires à celles du bloc de compréhension du langage, c'est-à-dire que sans informations supplémentaires sur les relations inter-mots, le lexique comprenant les synonymes et antonymes est difficile à utiliser. En effet, dans l'implémentation actuelle, nous avons beaucoup de synonymes et antonymes qui ne sont pas utilisables dans des contextes généraux et cela crée des problèmes lors de l'optimisation comme nous avons pu le constater par exemple dans le tableau 4.4. Certains synonymes employés font en sorte que la phrase originale perd de son sens, car ils ont été employés sans tenir compte d'un contexte d'utilisation plus robuste. Une autre faiblesse

du système est que nous comptabilisons les scores des tuples dans la base de données des scores que pour une seule émotion cible et un contexte test général. L'injection d'émotion à l'aide de Philia est donc limitée à cette seule émotion et ce contexte général. Idéalement, il faudrait initialiser la base de données des scores pour toutes les émotions considérées et tous les contextes dialogaux utiles. Une autre limitation de ce bloc est le manque d'analyse de sensibilité des hyper-paramètres utilisés pour les scores émotifs. Pour terminer, les méthodes d'initialisation/entraînement du système sont limitées. Actuellement, nous utilisons les bases de données *Emo WordNet* et *Alexa Topical-Chat* pour initialiser le système, mais cette manière de procéder est loin d'être idéale. En entraînant le système sur une base de données dialogale statique, nous ne pouvons pas réellement mesurer l'impact qu'ont les transformations du système sur la réaction de l'utilisateur. Ceci est problématique, car l'apprentissage de Philia dépend fortement des réactions des utilisateurs lorsqu'ils interagissent avec le système. De plus, en initialisant le système de la sorte, il nous est impossible d'utiliser la boucle rétroactive usager pour affiner les scores de la base de données et corriger les synonymes/antonymes de mauvaise qualité.

Pour le bloc d'*optimisation* présenté dans la section 3.4, nous observons les limitations suivantes. Premièrement, les arrondis dans le calcul des scores pondérés diminuent parfois la capacité du système de discriminer adéquatement les mots qui ont des scores similaires. Deuxièmement, nous n'avons pas pu intégrer les contraintes (3.7) et (3.8) permettant un meilleur contrôle de l'injection des émotions. Idéalement, nous aimerions intégrer des contraintes qui limitent la sélection de tuples de remplacement ayant des scores en dessous d'un certain seuil pour les émotions opposées dans un contexte donné, et qui forcent la sélection de tuples de remplacement qui ont des scores au-dessus d'un certain seuil pour les émotions similaires. Ceci pour nous assurer que les mots choisis par le système ne convoient pas les émotions indésirées et pour nous assurer d'une transmission minimale des émotions similaires à l'émotion cible.

Pour terminer, explorons les limitations de Philia reliées au bloc du *centre de commande* présenté dans la section 3.5. Pour le scoreur individuel, dans l'équation (3.20), nous utilisons les facteurs de pondération α_0 , α_1 et α_2 pour pondérer les scores associés à la détection de l'émotion cible, des émotions similaires et des émotions opposées respectivement. Ces facteurs sont identiques pour toutes les émotions appartenant à l'un de ces trois ensembles. Pour avoir une formule plus générale, il faudrait avoir un facteur de pondération par émotion et non par ensemble. Lors de la mise à jour des scores de la base de données, nous mettons à jour tous les tuples d'une réplique optimisée par le même facteur. Une meilleure approche serait d'intégrer un mécanisme capable d'attribuer un poids d'importance associé à la transmission d'émotion pour chaque tuple et de mettre à jour les scores des tuples en pondérant la valeur obtenue

du scoreur combiné par un poids d'importance. Finalement, la boucle de rétroaction usager implique de manière limitée l'utilisateur. Des questions plus profondes sur la syntaxe des répliques transformées, ou le contexte d'utilisation des mots pourraient aider le système à collecter un éventail plus large d'informations spécifiques à l'utilisation des tuples de remplacement.

5.2 Améliorations futures

Nous sommes au tout début de ce genre d'hybridation. Les possibilités d'amélioration sont donc nombreuses et diverses. Plusieurs pistes de recherche future s'offrent à nous. Tout d'abord, intégrer des agents conversationnels au sein même de Philia pourrait les faire bénéficier des connaissances du système pour générer des répliques plus adéquates pour un contexte donné. Une autre piste d'amélioration est de compléter la base de données actuelle des scores afin de comptabiliser les scores des tuples associés à des émotions différentes, et ce, pour des contextes dialogaux différents. Par la suite, une amélioration prioritaire serait de tenir compte des relations inter-mots d'une réplique. Ceci nous permettrait dans un premier temps de construire une base de données vocabulaire plus raffinée et versatile. Dans un deuxième temps, ce genre d'information supplémentaire aiderait lors de la sélection des tuples de remplacement afin d'assurer l'utilisation de termes plus appropriés par le système. Ces informations supplémentaires pourraient aussi aider à développer des modèles nouveaux qui ne se contenteraient pas uniquement de remplacer les mots d'une réplique par des mots qui élicitent l'émotion désirée de manière plus prononcée, mais qui pourrait régénérer des parties de la réplique originale ou complètement la modifier au besoin.

Terminons par une remarque la nécessité d'une contextualisation plus développée des synonymes. Si les modèles 3.6 ne jouent que sur le remplacement ou non de mots (et de manière plus générale sur le remplacement de sous-phrases ou non), notre étude a démontré qu'il fallait introduire des mécanismes supplémentaires de validation pour contrôler les remplacements possibles. Ces mécanismes peuvent être de différents ordres. Comme mentionné dans la section 4.3.4, nous pouvons valider les possibilités en amont, par exemple, en construisant des listes de synonymes plus pertinentes suivant des contextes syntaxiques, mais surtout sémantiques plus détaillés. Nous pouvons aussi valider les candidats admissibles en aval. Une possibilité est d'utiliser des contraintes supplémentaires du même type que celles ajoutées pour l'injection d'émotion dans les modèles plus sophistiqués ((3.7) et (3.8)), mais cette fois sur des indicateurs de pertinences syntaxiques et/ou sémantiques. Nous pourrions aussi utiliser une approche basée sur l'apprentissage automatique comme l'utilisation de *word embeddings* et/ou l'utilisation de vecteurs internes contenant de l'information pertinente à respecter ou à éviter. D'autres possibilités sont bien sûr envisageables. Nous préconisons une

combinaison de toutes ces approches. L'hybridation de l'apprentissage automatique avec la recherche opérationnelle a un bel avenir devant elle.

RÉFÉRENCES

- [1] H. Camirand, I. Traoré et J. Baulne, “L’enquête québécoise sur la santé de la population, 2014-2015 : pour en savoir plus sur la santé des québécois,” 2016. [En ligne]. Disponible : <https://statistique.quebec.ca/en/fichier/enquete-quebecoise-sur-la-sante-de-la-population-2014-2015-pour-en-savoir-plus-sur-la-sante-des-quebecois-resultats-de-la-deuxieme-edition.pdf>
- [2] F.-E. Lessard, “Enquête sur la santé psychologique étudiante,” 2016. [En ligne]. Disponible : <http://www.faecum.qc.ca/ressources/documentation/avis-memoires-recherches-et-positions-1/enquete-sur-la-sante-psychologique-etudiante>
- [3] A. Malhotra *et al.*, “An exploratory study into the use of an emotionally aware cognitive assistant,” dans *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [4] N. Asghar *et al.*, “Affective neural response generation,” dans *European Conference on Information Retrieval*. Springer, 2018, p. 154–166.
- [5] W. James, “What is an emotion ?” *Mind*, vol. os-IX, n°. 34, p. 188–205, 1884.
- [6] W. Cannon, “The james-lange theory of emotions : A critical examination and an alternative theory.” *American Journal of Psychology*, vol. 39, p. 10–124, 1927.
- [7] M. Arnold, “An excitatory theory of emotion,” 1950.
- [8] R. S. Lazarus, “Progress on a cognitive-motivational-relational theory of emotion,” *American psychologist*, vol. 46, n°. 8, p. 819, 1991.
- [9] I. J. Roseman, “Appraisal determinants of emotions : Constructing a more accurate and comprehensive theory,” *Cognition & Emotion*, vol. 10, n°. 3, p. 241–278, 1996.
- [10] ———, “Cognitive determinants of emotion : A structural theory.” *Review of personality & social psychology*, 1984.
- [11] K. R. Scherer, “Appraisal considered as a process of multilevel sequential checking,” *Appraisal processes in emotion : Theory, methods, research*, vol. 92, n°. 120, p. 57, 2001.
- [12] T. M. Moerland, J. Broekens et C. M. Jonker, “Emotion in reinforcement learning agents and robots : a survey,” *Machine Learning*, vol. 107, n°. 2, p. 443–480, 2018.
- [13] A. Schlesinger, K. P. O’Hara et A. S. Taylor, “Let’s talk about race : Identity, chatbots, and ai,” dans *Proceedings of the 2018 chi conference on human factors in computing systems*, 2018, p. 1–14.

- [14] A. Moors, “Theories of emotion causation : A review,” *Cognition and emotion*, vol. 23, n° 4, p. 625–662, 2009.
- [15] C. Kendra, “Overview of the 6 Major Theories of Emotion,” 2020. [En ligne]. Disponible : <https://www.verywellmind.com/theories-of-emotion-2795717>
- [16] Lumenlearning, “Theories of Emotion | Introduction to Psychology,” 2021. [En ligne]. Disponible : <https://courses.lumenlearning.com/waymaker-psychology/chapter/emotion/>
- [17] A. R. Damasio, “Descartes’ error and the future of human life.” *Scientific American*, vol. 271, n° 4, p. 144–144, 1994.
- [18] S. Schachter et J. Singer, “Cognitive, social, and physiological determinants of emotional state,” *Psychological review*, vol. 69, n° 5, p. 379, 1962.
- [19] N. Frijda, “Moods, emotion episodes, and emotions.” *Handbook of Emotions*, p. 381–403, 1993.
- [20] K. R. Scherer, “Appraisal theory,” *Handbook of Cognition and Emotion*, p. 637–663, 1999.
- [21] S. Folkman et R. S. Lazarus, *Stress, appraisal, and coping*. New York : Springer Publishing Company, 1984.
- [22] SparkNotes, “Emotion : Theories of Emotion,” 2021. [En ligne]. Disponible : <https://www.sparknotes.com/psychology/psych101/emotion/section1/>
- [23] W. James, *The Emotions*. Williams & Wilkins, 1922, vol. 1.
- [24] S. A. Shields et J. Matthew, “History of emotion theory,” 2011.
- [25] L. F. Barrett, “Emotions are real,” *Emotion*, vol. 12, n° 3, p. 413, 2012.
- [26] J. D. Laird et K. Lacasse, “Bodily influences on emotional feelings : Accumulating evidence and extensions of william james’s theory of emotion,” *Emotion Review*, vol. 6, n° 1, p. 27–34, 2014.
- [27] B. H. Friedman, “Feelings and the body : The jamesian perspective on autonomic specificity of emotion,” *Biological psychology*, vol. 84, n° 3, p. 383–393, 2010.
- [28] W. B. Cannon et S. W. Britton, “Studies on the conditions of activity in endocrine glands : Xv. pseudoaffective medulliadrenal secretion,” *American Journal of Physiology-Legacy Content*, vol. 72, n° 2, p. 283–294, 1925.
- [29] M. R. Roxo *et al.*, “The limbic system conception and its historical evolution,” *The Scientific World Journal*, p. 2427–2440, 2011.
- [30] J. E. LeDoux, “Emotion : Clues from the brain,” *Annual review of psychology*, vol. 46, n° 1, p. 209–235, 1995.

- [31] K. R. Scherer, A. Schorr et T. Johnstone, *Appraisal processes in emotion : Theory, methods, research*. Oxford University Press, 2001.
- [32] A. Alamsyah, W. Rahmah et H. Irawan, “Sentiment analysis based on appraisal theory for marketing intelligence in indonesia’s mobile phone market,” *Journal of Theoretical and Applied Information Technology*, vol. 82, n°. 2, p. 335 – 40, 2015.
- [33] V. Dragos, D. Battistelli et E. Kelodjoue, “Beyond sentiments and opinions : exploring social media with appraisal categories,” dans *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, p. 1851–1858.
- [34] P. Korenek et M. Rimko, “Sentiment analysis on microblog utilizing appraisal theory,” *World Wide Web*, vol. 17, n°. 4, p. 847 – 67, 2014.
- [35] C. Whitelaw, N. Garg et S. Argamon, “Using appraisal groups for sentiment analysis,” dans *Proceedings of the 14th ACM international conference on Information and knowledge management*, 2005, p. 625–631.
- [36] A. Rajendran, C. Zhang et M. Abdul-Mageed, “Happy together : Learning and understanding appraisal from natural language,” 2019.
- [37] P. Jin *et al.*, “Extracting appraisal expressions from short texts,” dans *International Conference on Web-Age Information Management*. Springer, 2015, p. 481–485.
- [38] P. Sequeira, F. S. Melo et A. Paiva, “Emotion-based intrinsic motivation for reinforcement learning agents,” dans *International conference on affective computing and intelligent interaction*. Springer, 2011, p. 326–336.
- [39] R. P. Marinier et J. E. Laird, “Emotion-driven reinforcement learning,” dans *Proceedings of the annual meeting of the cognitive science society*, vol. 30, n°. 30, 2008.
- [40] J. Bratman *et al.*, “Strong mitigation : nesting search for good policies within search for good reward.” dans *AAMAS*. Citeseer, 2012, p. 407–414.
- [41] T. Hester et P. Stone, “Intrinsically motivated model learning for developing curious robots,” *Artificial Intelligence*, vol. 247, p. 170–186, 2017.
- [42] R. Houthoof *et al.*, “Curiosity-driven exploration in deep reinforcement learning via bayesian neural networks,” 2016.
- [43] J. I. Davis, A. Senghas et K. N. Ochsner, “How does facial feedback modulate emotional experience?” *Journal of research in personality*, vol. 43, n°. 5, p. 822–829, 2009.
- [44] D. A. Havas *et al.*, “Cosmetic use of botulinum toxin-a affects processing of emotional language,” *Psychological science*, vol. 21, n°. 7, p. 895–900, 2010.

- [45] F. Strack, L. L. Martin et S. Stepper, “Inhibiting and facilitating conditions of the human smile : a nonobtrusive test of the facial feedback hypothesis,” *Journal of personality and social psychology*, vol. 54, n°. 5, p. 768, 1988.
- [46] Y. Li *et al.*, “Expressing reactive emotion based on multimodal emotion recognition for natural conversation in humanrobot interaction*,” *Advanced Robotics*, vol. 33, n°. 20, p. 1030 – 1041, 2019.
- [47] T. Saha *et al.*, “Emotion aided dialogue act classification for task-independent conversations in a multi-modal framework,” *Cognitive Computation*, vol. 13, n°. 2, p. 277–289, 2021.
- [48] P. Ekman, “An argument for basic emotions,” *Cognition & emotion*, vol. 6, n°. 3-4, p. 169–200, 1992.
- [49] R. S. Lazarus et B. N. Lazarus, *Passion and reason : Making sense of our emotions*. Oxford University Press, USA, 1994.
- [50] A. S. Cowen et D. Keltner, “Self-report captures 27 distinct categories of emotion bridged by continuous gradients,” *Proceedings of the National Academy of Sciences*, vol. 114, n°. 38, p. E7900–E7909, 2017.
- [51] J. Posner, J. A. Russell et B. S. Peterson, “The circumplex model of affect : An integrative approach to affective neuroscience, cognitive development, and psychopathology,” *Development and psychopathology*, vol. 17, n°. 3, p. 715–734, 2005.
- [52] J. A. Russell, “A circumplex model of affect,” *Journal of Personality and Social Psychology*, vol. 39, p. 1161–1178, 1980.
- [53] M. M. Bradley *et al.*, “Remembering pictures : pleasure and arousal in memory.” *Journal of experimental psychology : Learning, Memory, and Cognition*, vol. 18, n°. 2, p. 379, 1992.
- [54] R. Plutchik, “The nature of emotions,” *American Scientist*, vol. 89, p. 344, 2001.
- [55] P. Ekman, “Basic emotions,” *Handbook of cognition and emotion*, vol. 98, n°. 45-60, p. 16, 1999.
- [56] D. Demszky *et al.*, “Goemotions : A dataset of fine-grained emotions,” *arXiv preprint arXiv :2005.00547*, 2020.
- [57] E. Coutinho, E. R. Miranda et A. Cangelosi, “Towards a model for embodied emotions,” dans *2005 portuguese conference on artificial intelligence*. IEEE, 2005, p. 54–63.
- [58] J. Hoey, T. Schroder et A. Alhothali, “Bayesian affect control theory,” dans *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, p. 166–172.

- [59] T. Wilson, J. Wiebe et P. Hoffmann, “Recognizing contextual polarity in phrase-level sentiment analysis,” dans *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 2005, p. 347–354.
- [60] B. Pang et L. Lee, “Opinion mining and sentiment analysis,” *Comput. Linguist*, vol. 35, n^o. 2, p. 311–312, 2009.
- [61] P. R. Kleinginna et A. M. Kleinginna, “A categorized list of emotion definitions, with suggestions for a consensual definition,” *Motivation and Emotion*, vol. 5, n^o. 4, p. 345–379, 1981.
- [62] E. Cambria *et al.*, “New avenues in opinion mining and sentiment analysis,” *IEEE Intelligent systems*, vol. 28, n^o. 2, p. 15–21, 2013.
- [63] C. Hutto et E. Gilbert, “Vader : A parsimonious rule-based model for sentiment analysis of social media text,” dans *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, n^o. 1, 2014.
- [64] L. Im Tan *et al.*, “Rule-based sentiment analysis for financial news,” dans *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, p. 1601–1606.
- [65] H. Liu et M. Cocea, “Fuzzy rule based systems for interpretable sentiment analysis,” dans *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE, 2017, p. 129–136.
- [66] M. Z. Asghar *et al.*, “Lexicon-enhanced sentiment analysis framework using rule-based classification scheme,” *PloS one*, vol. 12, n^o. 2, p. e0171649, 2017.
- [67] W. Li et H. Xu, “Text-based emotion classification using emotion cause extraction,” *Expert Systems with Applications*, vol. 41, n^o. 4, p. 1742–1749, 2014.
- [68] A. Almahdawi et W. J. Teahan, “Emotion recognition in text using ppm,” dans *International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, 2017, p. 149–155.
- [69] J. Tao, “Context based emotion detection from text input,” 2004.
- [70] C. Ma, H. Prendinger et M. Ishizuka, “Emotion estimation and reasoning based on affective textual interaction.” Springer Berlin Heidelberg, 2005, p. 622–628.
- [71] X. Glorot, A. Bordes et Y. Bengio, “Domain adaptation for large-scale sentiment classification : A deep learning approach,” dans *ICML*, 2011.
- [72] O. Romero Llobart, “Using machine learning techniques for sentiment analysis,” 2017.
- [73] R. Xia, C. Zong et S. Li, “Ensemble of feature sets and classification algorithms for sentiment classification,” *Information sciences*, vol. 181, n^o. 6, p. 1138–1152, 2011.

- [74] C. Troussas *et al.*, “Sentiment analysis of facebook statuses using naive bayes classifier for language learning,” dans *IISA 2013*. IEEE, 2013, p. 1–6.
- [75] A. F. A. Nasir *et al.*, “Text-based emotion prediction system using machine learning approach,” dans *IOP Conference Series : Materials Science and Engineering*, vol. 769, n° 1. IOP Publishing, 2020, p. 012022.
- [76] S. Aman et S. Szpakowicz, “Identifying expressions of emotion in text,” dans *International Conference on Text, Speech and Dialogue*. Springer, 2007, p. 196–205.
- [77] A. Severyn et A. Moschitti, “Twitter sentiment analysis with deep convolutional neural networks,” dans *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 2015, p. 959–962.
- [78] D. Stojanovski *et al.*, “Deep neural network architecture for sentiment analysis and emotion identification of twitter messages,” *Multimedia Tools and Applications*, vol. 77, n° 24, p. 32 213 – 42, 2018.
- [79] E. Batbaatar, M. Li et K. H. Ryu, “Semantic-emotion neural network for emotion recognition from text,” *IEEE Access*, vol. 7, p. 111 866 – 78, 2019.
- [80] N. Colneric et J. Demsar, “Emotion recognition on twitter : Comparative study and training a unison model,” *IEEE Transactions on Affective Computing*, 2018.
- [81] S. Zhou, Q. Chen et X. Wang, “Fuzzy deep belief networks for semi-supervised sentiment classification,” *Neurocomputing*, vol. 131, p. 312–322, 2014.
- [82] M. Ghiassi, J. Skinner et D. Zimbra, “Twitter brand sentiment analysis : A hybrid system using n-gram analysis and dynamic artificial neural network,” *Expert Systems with applications*, vol. 40, n° 16, p. 6266–6282, 2013.
- [83] A. C. Pandey, D. S. Rajpoot et M. Saraswat, “Twitter sentiment analysis using hybrid cuckoo search method,” *Information Processing & Management*, vol. 53, n° 4, p. 764–779, 2017.
- [84] F. H. Khan, S. Bashir et U. Qamar, “Tom : Twitter opinion mining framework using hybrid classification scheme,” *Decision support systems*, vol. 57, p. 245–257, 2014.
- [85] S. Gievska, K. Koroveshovski et T. Chavdarova, “A hybrid approach for emotion detection in support of affective interaction,” dans *2014 IEEE International Conference on Data Mining Workshop*. IEEE, 2014, p. 352–359.
- [86] Y. Gao et W. Zhu, “Detecting affective states from text based on a multi-component emotion model,” *Computer Speech and Language*, vol. 36, p. 42 – 57, 2016.
- [87] S. Shaheen *et al.*, “Emotion recognition from text based on automatically generated rules,” dans *2014 IEEE International Conference on Data Mining Workshop*. IEEE, 2014, p. 383–392.

- [88] A.-C. Le *et al.*, “Integrating grammatical features into cnn model for emotion classification,” dans *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*. IEEE, 2018, p. 243–249.
- [89] S. N. Shivhare, S. Garg et A. Mishra, “Emotionfinder : Detecting emotion from blogs and textual documents,” dans *International Conference on Computing, Communication & Automation*. IEEE, 2015, p. 52–57.
- [90] C. D. Elliott, “The affective reasoner : a process model of emotions in a multiagent system,” Thèse de doctorat, Northwestern University, 1992.
- [91] C. Strapparava et R. Mihalcea, “Learning to identify emotions in text,” dans *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, p. 1556–1560.
- [92] S. Somasundaran, J. Wiebe et J. Ruppenhofer, “Discourse level opinion interpretation,” dans *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, 2008, p. 801–808.
- [93] L. Iliadis *et al.*, “Recognizing emotion presence in natural language sentences,” ser. Engineering Applications of Neural Networks. Springer Berlin Heidelberg, 2013, p. 30–39.
- [94] O. Udochukwu et Y. He, “A rule-based approach to implicit emotion detection in text.” Springer International Publishing, 2015, p. 197–203.
- [95] F. Benamara *et al.*, “Sentiment analysis : Adjectives and adverbs are better than adjectives alone.” *ICWSM*, vol. 7, p. 203–206, 2007.
- [96] H. Liu, H. Lieberman et T. Selker, “A model of textual affect sensing using real-world knowledge,” dans *Proceedings of the 8th international conference on Intelligent user interfaces*, 2003, p. 125–132.
- [97] S. Y. M. Lee, Y. Chen et C.-R. Huang, “A text-driven rule-based system for emotion cause detection,” dans *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, 2010, p. 45–53.
- [98] R. McDonald *et al.*, “Structured models for fine-to-coarse sentiment analysis,” dans *Proceedings of the 45th annual meeting of the association of computational linguistics*, 2007, p. 432–439.
- [99] M. Schröder, “Emotional speech synthesis : A review,” dans *Seventh European Conference on Speech Communication and Technology*, 2001.
- [100] T. Nasukawa et J. Yi, “Sentiment analysis : Capturing favorability using natural language processing,” dans *Proceedings of the 2nd international conference on Knowledge capture*, 2003, p. 70–77.

- [101] I. McCowan *et al.*, “The ami meeting corpus,” dans *Proceedings of the 5th international conference on methods and techniques in behavioral research*, vol. 88. Citeseer, 2005, p. 100.
- [102] D. B. Lenat, “Cyc : A large-scale investment in knowledge infrastructure,” *Communications of the ACM*, vol. 38, n^o. 11, p. 33–38, 1995.
- [103] P. Singh *et al.*, “The public acquisition of commonsense knowledge,” dans *Proceedings of AAAI Spring Symposium : Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, 2002.
- [104] E. T. Mueller, “Thoughttreasure : A natural language/commonsense platform,” 1998.
- [105] H. Liu et P. Singh, “Conceptnet—a practical commonsense reasoning tool-kit,” *BT technology journal*, vol. 22, n^o. 4, p. 211–226, 2004.
- [106] M.-C. De Marneffe *et al.*, “Generating typed dependency parses from phrase structure parses.” dans *Lrec*, vol. 6, 2006, p. 449–454.
- [107] C. Strapparava, A. Valitutti *et al.*, “Wordnet affect : an affective extension of wordnet.” dans *Lrec*, vol. 4, n^o. 1083-1086. Lisbon, 2004, p. 40.
- [108] A. Ortony, G. L. Clore et A. Collins, *The cognitive structure of emotions*. Cambridge university press, 1990.
- [109] L. K. Hansen *et al.*, “Good friends, bad news - affect and virality in twitter,” dans *Future Information Technology*, J. J. Park, L. T. Yang et C. Lee, édit. Berlin, Heidelberg : Springer Berlin Heidelberg, 2011, p. 34–43.
- [110] A. Esuli et F. Sebastiani, “Determining the semantic orientation of terms through gloss classification,” dans *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, ser. CIKM '05. New York, NY, USA : Association for Computing Machinery, 2005, p. 617–624.
- [111] K.-J. Chen *et al.*, “Sinica corpus : Design methodology for balanced corpora,” dans *Proceedings of the 11th Pacific Asia Conference on Language, Information and Computation*, 1996, p. 167–176.
- [112] J. Kleres, “Emotions and narrative analysis : A methodological approach,” *Journal for the theory of social behaviour*, vol. 41, n^o. 2, p. 182–202, 2011.
- [113] J. R. Martin et P. R. White, *The language of evaluation*. Springer, 2003, vol. 2.
- [114] Z. S. Harris, “Distributional structure,” *Word*, vol. 10, n^o. 2-3, p. 146–162, 1954.
- [115] N. Cristianini, J. Shawe-Taylor *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

- [116] E. Kontopoulos *et al.*, “Ontology-based sentiment analysis of twitter posts,” *Expert systems with applications*, vol. 40, n^o. 10, p. 4065–4074, 2013.
- [117] R. Studer, V. R. Benjamins et D. Fensel, “Knowledge engineering : Principles and methods,” *Data & knowledge engineering*, vol. 25, n^o. 1-2, p. 161–197, 1998.
- [118] J. Ramos *et al.*, “Using tf-idf to determine word relevance in document queries,” dans *Proceedings of the first instructional conference on machine learning*, vol. 242, n^o. 1. Citeseer, 2003, p. 29–48.
- [119] T. Mikolov *et al.*, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv :1301.3781*, 2013.
- [120] Y. Bengio *et al.*, “A neural probabilistic language model,” *Journal of machine learning research*, vol. 3, n^o. Feb, p. 1137–1155, 2003.
- [121] C. O. Alm, D. Roth et R. Sproat, “Emotions from text : machine learning for text-based emotion prediction,” dans *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 2005, p. 579–586.
- [122] W. Wang *et al.*, “Harnessing twitter" big data" for automatic emotion identification,” dans *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*. IEEE, 2012, p. 587–592.
- [123] S. Aman et S. Szpakowicz, “Using roget’s thesaurus for fine-grained emotion recognition,” dans *Proceedings of the Third International Joint Conference on Natural Language Processing : Volume-I*, 2008.
- [124] T. Danisman et A. Alpkocak, “Feeler : Emotion classification of text using vector space model,” dans *AISB 2008 convention communication, interaction and social intelligence*, vol. 1, 2008, p. 53.
- [125] S. Tan et J. Zhang, “An empirical study of sentiment analysis for chinese documents,” *Expert Systems with applications*, vol. 34, n^o. 4, p. 2622–2629, 2008.
- [126] D. Ghazi, D. Inkpen et S. Szpakowicz, “Hierarchical versus flat classification of emotions in text,” dans *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, 2010, p. 140–146.
- [127] A. Go, L. Huang et R. Bhayani, “Twitter sentiment analysis,” *Entropy*, vol. 17, p. 252, 2009.
- [128] A. Go, R. Bhayani et L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N project report, Stanford*, vol. 1, n^o. 12, p. 2009, 2009.
- [129] N. Zainuddin et A. Selamat, “Sentiment analysis using support vector machine,” dans *2014 international conference on computer, communications, and control technology (I4CT)*. IEEE, 2014, p. 333–337.

- [130] B. Pang, L. Lee et S. Vaithyanathan, “Thumbs up? sentiment classification using machine learning techniques,” *arXiv preprint cs/0205070*, 2002.
- [131] Z. Tan *et al.*, “Research on the text emotion of multinomial naïve bayes integration algorithm,” dans *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE, 2019, p. 107–111.
- [132] P. Gamallo et M. Garcia, “Citius : A naivebayes strategy for sentiment analysis on english tweets,” dans *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Citeseer, 2014.
- [133] L. Dey *et al.*, “Sentiment analysis of review datasets using naive bayes and k-nn classifier,” *arXiv preprint arXiv :1610.09982*, 2016.
- [134] B. Y. Pratama et R. Sarno, “Personality classification based on twitter text using naive bayes, knn and svm,” dans *2015 International Conference on Data and Software Engineering (ICoDSE)*. IEEE, 2015, p. 170–174.
- [135] A. Carlson *et al.*, “The snow learning architecture (technical report uiucdcs-r-99-2101),” *Dept. of CS, UIUC*, 1999.
- [136] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and psychological measurement*, vol. 20, n^o. 1, p. 37–46, 1960.
- [137] P. J. Stone, D. C. Dunphy et M. S. Smith, “The general inquirer : A computer approach to content analysis.” 1966.
- [138] C. Strapparava et R. Mihalcea, “Semeval-2007 task 14 : Affective text,” dans *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, 2007, p. 70–74.
- [139] G. L. Clore, A. Ortony et M. A. Foss, “The psychological foundations of the affective lexicon.” *Journal of personality and social psychology*, vol. 53, n^o. 4, p. 751, 1987.
- [140] D. Ghazi, D. Inkpen et S. Szpakowicz, “Hierarchical approach to emotion recognition and classification in texts,” dans *Canadian Conference on Artificial Intelligence*. Springer, 2010, p. 40–50.
- [141] E. C. O. Alm, *Affect in* text and speech*. University of Illinois at Urbana-Champaign, 2008.
- [142] P. D. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” *arXiv preprint cs/0212032*, 2002.
- [143] K. Church et P. Hanks, “Word association norms, mutual information, and lexicography,” *Computational linguistics*, vol. 16, n^o. 1, p. 22–29, 1990.

- [144] Q. Ye, Z. Zhang et R. Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert systems with applications*, vol. 36, n^o. 3, p. 6527–6535, 2009.
- [145] D. Tang, B. Qin et T. Liu, “Document modeling with gated recurrent neural network for sentiment classification,” dans *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, p. 1422–1432.
- [146] J. Xu *et al.*, “Cached long short-term memory neural networks for document-level sentiment classification,” *arXiv preprint arXiv :1610.04989*, 2016.
- [147] Y. Wang *et al.*, “Attention-based lstm for aspect-level sentiment classification,” dans *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, p. 606–615.
- [148] H. Zhou *et al.*, “Emotional chatting machine : Emotional conversation generation with internal and external memory,” dans *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [149] C. Baziotis *et al.*, “Ntua-slp at semeval-2018 task 1 : Predicting affective content in tweets with deep attentive rnns and transfer learning,” *arXiv preprint arXiv :1804.06658*, 2018.
- [150] N. Majumder *et al.*, “Dialoguernn : An attentive rnn for emotion detection in conversations,” *arXiv preprint arXiv :1811.00405*, 2018.
- [151] J. Wang *et al.*, “Dimensional sentiment analysis using a regional cnn-lstm model,” dans *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2 : Short Papers)*, 2016, p. 225–230.
- [152] R. Johnson et T. Zhang, “Effective use of word order for text categorization with convolutional neural networks,” *arXiv preprint arXiv :1412.1058*, 2014.
- [153] D. Tang, B. Qin et T. Liu, “Learning semantic representations of users and products for document level sentiment classification,” dans *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1 : Long Papers)*, 2015, p. 1014–1023.
- [154] D. Ghosal *et al.*, “Dialoguegn : A graph convolutional neural network for emotion recognition in conversation,” *arXiv preprint arXiv :1908.11540*, 2019.
- [155] R. Socher *et al.*, “Semi-supervised recursive autoencoders for predicting sentiment distributions,” dans *Proceedings of the 2011 conference on empirical methods in natural language processing*, 2011, p. 151–161.

- [156] Z. Yang *et al.*, “Hierarchical attention networks for document classification,” dans *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics : human language technologies*, 2016, p. 1480–1489.
- [157] H. Chen *et al.*, “Neural sentiment classification with user and product attention,” dans *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, p. 1650–1659.
- [158] Z. Wang *et al.*, “A bilingual attention network for code-switched emotion prediction,” dans *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics : Technical Papers*, 2016, p. 1624–1634.
- [159] J. Zheng, “A novel computer-aided emotion recognition of text method based on word-embedding and bi-lstm,” dans *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*. IEEE, 2019, p. 176–180.
- [160] M. Abdul-Mageed et L. Ungar, “Emonet : Fine-grained emotion detection with gated recurrent neural networks,” dans *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1 : Long papers)*, 2017, p. 718–728.
- [161] L. Gui *et al.*, “A question answering approach to emotion cause extraction,” *arXiv preprint arXiv :1708.05482*, 2017.
- [162] Y. Kim, “Convolutional neural networks for sentence classification,” *CoRR*, vol. abs/1408.5882, 2014.
- [163] D. Hazarika *et al.*, “Conversational transfer learning for emotion recognition,” *Information Fusion*, vol. 65, p. 1–12, 2021.
- [164] D. Ghosal *et al.*, “Cosmic : Commonsense knowledge for emotion identification in conversations,” *arXiv preprint arXiv :2010.02795*, 2020.
- [165] C. Huang *et al.*, “Seq2emo for multi-label emotion classification based on latent variable chains transformation,” *arXiv preprint arXiv :1911.02147*, 2019.
- [166] W. Ragheb *et al.*, “Attention-based modeling for emotion detection and classification in textual conversations,” *arXiv preprint arXiv :1906.07020*, 2019.
- [167] J. Devlin *et al.*, “Bert : Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv :1810.04805*, 2018.
- [168] P. Vincent *et al.*, “Extracting and composing robust features with denoising autoencoders,” dans *Proceedings of the 25th international conference on Machine learning*, 2008, p. 1096–1103.
- [169] D. T. Ho et T. H. Cao, “A high-order hidden markov model for emotion detection from textual data,” dans *Pacific rim knowledge acquisition workshop*. Springer, 2012, p. 94–105.

- [170] Y. Liu *et al.*, “Roberta : A robustly optimized bert pretraining approach,” *arXiv preprint arXiv :1907.11692*, 2019.
- [171] A. Bosselut *et al.*, “Comet : Commonsense transformers for automatic knowledge graph construction,” *arXiv preprint arXiv :1906.05317*, 2019.
- [172] J. Read, “Using emoticons to reduce dependency in machine learning techniques for sentiment classification,” dans *Proceedings of the ACL student research workshop*, 2005, p. 43–48.
- [173] S. Baccianella, A. Esuli et F. Sebastiani, “Sentiwordnet 3.0 : an enhanced lexical resource for sentiment analysis and opinion mining.” dans *Lrec*, vol. 10, n°. 2010, 2010, p. 2200–2204.
- [174] B. Liu *et al.*, “Text classification by labeling words,” dans *AAAI*, vol. 4, 2004, p. 425–430.
- [175] Y. Ma *et al.*, “A survey on empathetic dialogue systems,” *Information Fusion*, vol. 64, p. 50–70, 2020.
- [176] C. Huang *et al.*, “Automatic dialogue generation with expressed emotions,” dans *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies, Volume 2 (Short Papers)*, 2018, p. 49–54.
- [177] H. Rashkin *et al.*, “I know the feeling : Learning to converse with empathy,” 2018.
- [178] A. Vaswani *et al.*, “Attention is all you need,” dans *Advances in neural information processing systems*, 2017, p. 5998–6008.
- [179] X. Kong *et al.*, “An adversarial approach to high-quality, sentiment-controlled neural dialogue generation,” *arXiv preprint arXiv :1901.07129*, 2019.
- [180] J. Li *et al.*, “A diversity-promoting objective function for neural conversation models,” *arXiv preprint arXiv :1510.03055*, 2015.
- [181] T. Niu et M. Bansal, “Polite dialogue generation without parallel data,” *Transactions of the Association for Computational Linguistics*, vol. 6, p. 373–389, 2018.
- [182] Y. Peng *et al.*, “Topic-enhanced emotional conversation generation with attention mechanism,” *Knowledge-Based Systems*, vol. 163, p. 429–437, 2019.
- [183] P. Fung *et al.*, “Empathetic dialog systems,” dans *The international conference on language resources and evaluation. European Language Resources Association*, 2018.
- [184] X. Zhou et W. Y. Wang, “Mojitalk : Generating emotional responses at scale,” *arXiv preprint arXiv :1711.04090*, 2017.

- [185] K. Sohn, H. Lee et X. Yan, “Learning structured output representation using deep conditional generative models,” *Advances in neural information processing systems*, vol. 28, p. 3483–3491, 2015.
- [186] S. Ghosh *et al.*, “Affect-lm : A neural language model for customizable affective text generation,” *arXiv preprint arXiv :1704.06851*, 2017.
- [187] J. Xu *et al.*, “Unpaired sentiment-to-sentiment translation : A cycled reinforcement learning approach,” *arXiv preprint arXiv :1805.05181*, 2018.
- [188] D. Peng *et al.*, “Human–machine dialogue modelling with the fusion of word-and sentence-level emotions,” *Knowledge-Based Systems*, vol. 192, p. 105319, 2020.
- [189] R. Zhang *et al.*, “Emotional text generation based on cross-domain sentiment transfer,” *IEEE Access*, vol. 7, p. 100 081–100 089, 2019.
- [190] R. S. Sutton et A. G. Barto, *Reinforcement learning : An introduction*. MIT press, 2018.
- [191] M. A. Wiering et M. Van Otterlo, “Reinforcement learning,” *Adaptation, learning, and optimization*, vol. 12, n^o. 3, 2012.
- [192] K. L. Chung, “Markov chains,” *Springer-Verlag, New York*, 1967.
- [193] O. Omidvar et D. L. Elliott, *Neural systems for control*. Elsevier, 1997.
- [194] M. L. Littman, “Value-function reinforcement learning in markov games,” *Cognitive systems research*, vol. 2, n^o. 1, p. 55–66, 2001.
- [195] J. Boyan et A. W. Moore, “Generalization in reinforcement learning : Safely approximating the value function,” *Advances in neural information processing systems*, p. 369–376, 1995.
- [196] M. G. Lagoudakis et R. Parr, “Least-squares policy iteration,” *The Journal of Machine Learning Research*, vol. 4, p. 1107–1149, 2003.
- [197] H. Van Hasselt, A. Guez et D. Silver, “Deep reinforcement learning with double q-learning,” dans *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, n^o. 1, 2016.
- [198] C. J. Watkins et P. Dayan, “Q-learning,” *Machine learning*, vol. 8, n^o. 3-4, p. 279–292, 1992.
- [199] A. Lazaric, M. Restelli et A. Bonarini, “Reinforcement learning in continuous action spaces through sequential monte carlo methods,” *Advances in neural information processing systems*, vol. 20, p. 833–840, 2007.

- [200] T. Jaakkola, S. P. Singh et M. I. Jordan, “Reinforcement learning algorithm for partially observable markov decision problems,” *Advances in neural information processing systems*, p. 345–352, 1995.
- [201] S. C. Gadanho et J. Hallam, *Emotion-triggered learning for autonomous robots*. Citeseer, 1998.
- [202] N. Goerke, *EMOBOT : A robot control architecture based on emotion-like internal values*. INTECH Open Access Publisher Rijeka, 2006.
- [203] R. Von Haugwitz, Y. Kitamura et K. Takashima, “Modulating reinforcement-learning parameters using agent emotions,” dans *The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems*. IEEE, 2012, p. 1281–1285.
- [204] I. Cos *et al.*, “Hedonic value : Enhancing adaptation for motivated agents,” *Adaptive Behavior*, vol. 21, n^o. 6, p. 465–483, 2013.
- [205] P.-Y. Oudeyer et F. Kaplan, “What is intrinsic motivation ? a typology of computational approaches,” *Frontiers in neurorobotics*, vol. 1, p. 6, 2009.
- [206] C. Yu *et al.*, “Emotional multiagent reinforcement learning in spatial social dilemmas,” *IEEE transactions on neural networks and learning systems*, vol. 26, n^o. 12, p. 3083–3096, 2015.
- [207] S. Bozinovski, G. Stojanov et L. Bozinovska, “Emotion, embodiment, and consequence driven systems,” dans *Proc AAAI fall symposium on embodied cognition and action*, 1996, p. 12–17.
- [208] A. Matsuda, H. Misawa et K. Horio, “Decision making based on reinforcement learning and emotion learning for social behavior,” dans *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*. IEEE, 2011, p. 2714–2719.
- [209] E. Jacobs, J. Broekens et C. Jonker, “Emergent dynamics of joy, distress, hope and fear in reinforcement learning agents,” dans *Adaptive learning agents workshop at AAMAS2014*, 2014.
- [210] T. M. Moerland, J. Broekens et C. M. Jonker, “Fear and hope emerge from anticipation in model-based reinforcement learning.” dans *IJCAI*, 2016, p. 848–854.
- [211] N. Kubota et S. Wakisaka, “Emotional model based on computational intelligence for partner robots,” dans *Modeling machine emotions for realizing intelligence*. Springer, 2010, p. 89–108.
- [212] C. L. Hull, “Principles of behavior : An introduction to behavior theory.” 1943.

- [213] D. Canamero, “A hormonal model of emotions for behavior control,” *VUB AI-Lab Memo*, vol. 2006, p. 1–10, 1997.
- [214] ———, “Modeling motivations and emotions as a basis for intelligent behavior,” dans *Proceedings of the first international conference on Autonomous agents*, 1997, p. 148–155.
- [215] G. Konidaris et A. Barto, “An adaptive robot motivational system,” dans *International Conference on Simulation of Adaptive Behavior*. Springer, 2006, p. 346–356.
- [216] M. Keramati et B. S. Gutkin, “A reinforcement learning theory for homeostatic regulation,” *Advances in neural information processing systems*, n^o. 24, p. 82–90, 2011.
- [217] M. A. Salichs et M. Malfaz, “A new approach to modeling emotions and their use on a decision-making system for artificial agents,” *IEEE Transactions on affective computing*, vol. 3, n^o. 1, p. 56–68, 2011.
- [218] I. J. Roseman, M. S. Spindel et P. E. Jose, “Appraisals of emotion-eliciting events : Testing a theory of discrete emotions,” *Journal of personality and social psychology*, vol. 59, n^o. 5, p. 899, 1990.
- [219] H. Williams *et al.*, “Emotion inspired adaptive robotic path planning,” dans *2015 IEEE Congress on Evolutionary Computation (CEC)*, 2015, p. 3004–3011.
- [220] C. Hasson, P. Gaussier et S. Boucenna, “Emotions as a dynamical system : the interplay between the meta-control and communication function of emotions,” *Paladyn*, vol. 2, n^o. 3, p. 111–125, 2011.
- [221] M. Ficocelli, J. Terao et G. Nejat, “Promoting interactions between humans and robots using robotic emotional behavior,” *IEEE transactions on cybernetics*, vol. 46, n^o. 12, p. 2911–2923, 2015.
- [222] X. Li *et al.*, “How to evaluate your dialogue models : A review of approaches,” *arXiv preprint arXiv :2108.01369*, 2021.
- [223] S. Mehri et M. Eskenazi, “U_{sr} : An unsupervised and reference free evaluation metric for dialog generation,” *arXiv preprint arXiv :2005.00456*, 2020.
- [224] K. Papineni *et al.*, “Bleu : a method for automatic evaluation of machine translation,” dans *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, p. 311–318.
- [225] C.-Y. Lin, “Rouge : A package for automatic evaluation of summaries,” dans *Text summarization branches out*, 2004, p. 74–81.
- [226] M. Denkowski et A. Lavie, “Meteor universal : Language specific translation evaluation for any target language,” dans *Proceedings of the ninth workshop on statistical machine translation*, 2014, p. 376–380.

- [227] C.-W. Liu *et al.*, “How not to evaluate your dialogue system : An empirical study of unsupervised evaluation metrics for dialogue response generation,” *arXiv preprint arXiv :1603.08023*, 2016.
- [228] E. Dinan *et al.*, “The second conversational intelligence challenge (convai2),” *arXiv preprint arXiv :1902.00098*, 2019.
- [229] G. A. Miller, “Wordnet : a lexical database for english,” *Communications of the ACM*, vol. 38, n^o. 11, p. 39–41, 1995.
- [230] V. Rus et M. Lintean, “An optimal assessment of natural language student input using word-to-word similarity metrics,” dans *International Conference on Intelligent Tutoring Systems*. Springer, 2012, p. 675–676.
- [231] J. Wieting *et al.*, “Towards universal paraphrastic sentence embeddings,” *arXiv preprint arXiv :1511.08198*, 2015.
- [232] T. K. Landauer et S. T. Dumais, “A solution to plato’s problem : The latent semantic analysis theory of acquisition, induction, and representation of knowledge.” *Psychological review*, vol. 104, n^o. 2, p. 211, 1997.
- [233] G. Forgues *et al.*, “Bootstrapping dialog systems with word embeddings,” dans *Nips, modern machine learning and natural language processing workshop*, vol. 2, 2014.
- [234] R. Kiros *et al.*, “Skip-thought vectors,” dans *Advances in neural information processing systems*, 2015, p. 3294–3302.
- [235] T. Zhang *et al.*, “Bertscore : Evaluating text generation with bert,” *arXiv preprint arXiv :1904.09675*, 2019.
- [236] T. Zhao, R. Zhao et M. Eskenazi, “Learning discourse-level diversity for neural dialog models using conditional variational autoencoders,” *arXiv preprint arXiv :1703.10960*, 2017.
- [237] M. A. Walker *et al.*, “Paradise : A framework for evaluating spoken dialogue agents,” *arXiv preprint cmp-lg/9704004*, 1997.
- [238] J. Deriu *et al.*, “Survey on evaluation methods for dialogue systems,” *Artificial Intelligence Review*, vol. 54, n^o. 1, p. 755–810, 2021.
- [239] S. Mehri et M. Eskenazi, “Unsupervised evaluation of interactive dialog with dialogpt,” *arXiv preprint arXiv :2006.12719*, 2020.
- [240] C. Tao *et al.*, “Ruber : An unsupervised method for automatic evaluation of open-domain dialog systems,” dans *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

- [241] K. Gopalakrishnan *et al.*, “Topical-chat : Towards knowledge-grounded open-domain conversations.” dans *INTERSPEECH*, 2019, p. 1891–1895.
- [242] S. Zhang *et al.*, “Personalizing dialogue agents : I have a dog, do you have pets too?” *arXiv preprint arXiv :1801.07243*, 2018.
- [243] Y. Zhang *et al.*, “Dialogpt : Large-scale generative pre-training for conversational response generation,” *arXiv preprint arXiv :1911.00536*, 2019.
- [244] S. Roller *et al.*, “Recipes for building an open-domain chatbot,” 2020.
- [245] M. Honnibal et I. Montani, “spacy 2 : Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing,” *To appear*, vol. 7, n° 1, p. 411–420, 2017.
- [246] G. Badaro *et al.*, “Emowordnet : Automatic expansion of emotion lexicon using english wordnet,” dans *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, 2018, p. 86–93.
- [247] L. Perron et V. Furnon, “Or-tools,” Google, 2019. [En ligne]. Disponible : <https://developers.google.com/optimization/>
- [248] M. Eskenazi *et al.*, “Beyond turing : Intelligent agents centered on the user,” *arXiv preprint arXiv :1901.06613*, 2019.

**ANNEXE A RÉSULTATS DÉTAILLÉS OBTENUS LORS DE
L'ÉVALUATION DU SYSTÈME, INITIALISATION FAITE AVEC UN
APPORT SENTIMENTAL À 0.25**

TABLEAU A.1 Résultats obtenus en appliquant un facteur pénalisant de 0 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental 0.25, Pénalité = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.478	-0.490	4.84	2.44
Engagement	-0.625	-0.632	-0.640	-0.99	-2.33
Originalité	-0.328	-0.297	-0.307	9.47	6.35
Pertinence	0.206	0.185	0.191	-10.09	-7.32
Compréhension interloc.	0.036	-0.014	-0.007	-138.41	-119.05
Sémantiquement approprié	0.374	0.356	0.364	-4.86	-2.56
Compréhension réplique	0.329	0.294	0.305	-10.54	-7.35
Fluidité	0.354	0.322	0.334	-9.06	-5.80
Cohérence	0.142	0.138	0.139	-2.58	-1.86
Récupération d'erreurs	-0.654	-0.665	-0.670	-1.66	-2.46
Constance	0.560	0.575	0.579	2.77	3.46
Diversité	0.392	0.388	0.392	-1.23	-0.03
Profondeur	0.077	0.100	0.098	29.33	26.19
Sympathie	-0.382	-0.385	-0.392	-1.02	-2.72
Compréhension générale	0.008	-0.023	-0.017	-404.78	-325.54
Flexibilité	-0.099	-0.102	-0.105	-3.40	-6.22
Informatif	-0.245	-0.222	-0.230	9.34	6.11
Curiosité	-0.574	-0.546	-0.560	4.95	2.40
Score individuel moyen	0.571	0.677	0.673	18.51	17.86
Précision émotion moyenne	0.506	0.609	0.605	20.15	19.41
Score combiné moyen	0.571	0.677	0.673	18.51	17.86

TABLEAU A.2 Résultats obtenus en appliquant un facteur pénalisant de 3 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental 0.25, Pénalité = 3			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.478	-0.489	4.80	2.63
Engagement	-0.625	-0.633	-0.640	-1.16	-2.36
Originalité	-0.328	-0.297	-0.305	9.40	6.93
Pertinence	0.206	0.185	0.190	-10.16	-7.61
Compréhension interloc.	0.036	-0.015	-0.009	-141.09	-126.59
Sémantiquement approprié	0.374	0.356	0.364	-4.88	-2.56
Compréhension réplique	0.329	0.293	0.303	-10.99	-7.93
Fluidité	0.354	0.321	0.332	-9.27	-6.24
Cohérence	0.142	0.137	0.139	-3.37	-1.94
Récupération d'erreurs	-0.654	-0.667	-0.672	-1.91	-2.77
Constance	0.560	0.576	0.580	2.84	3.65
Diversité	0.392	0.388	0.392	-1.22	-0.22
Profondeur	0.077	0.099	0.098	28.75	27.28
Sympathie	-0.382	-0.387	-0.392	-1.35	-2.85
Compréhension générale	0.008	-0.024	-0.020	-421.25	-362.96
Flexibilité	-0.099	-0.103	-0.106	-4.52	-7.49
Informatif	-0.245	-0.223	-0.228	9.18	7.07
Curiosité	-0.574	-0.546	-0.559	4.84	2.64
Score individuel moyen	0.571	0.703	0.677	23.05	18.51
Précision émotion moyenne	0.506	0.629	0.620	24.18	22.34
Score combiné moyen	0.571	0.703	0.677	23.05	18.51

TABLEAU A.3 Résultats obtenus en appliquant un facteur pénalisant de 6 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental 0.25, Pénalité = 6			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.477	-0.488	5.08	2.80
Engagement	-0.625	-0.631	-0.638	-0.96	-2.10
Originalité	-0.328	-0.296	-0.306	9.68	6.83
Pertinence	0.206	0.186	0.190	-9.73	-7.56
Compréhension interloc.	0.036	-0.014	-0.008	-140.13	-121.16
Sémantiquement approprié	0.374	0.356	0.363	-4.87	-2.76
Compréhension réplique	0.329	0.293	0.302	-10.99	-8.16
Fluidité	0.354	0.322	0.332	-9.19	-6.34
Cohérence	0.142	0.137	0.138	-3.11	-2.45
Récupération d'erreurs	-0.654	-0.664	-0.668	-1.44	-2.18
Constance	0.560	0.575	0.579	2.74	3.43
Diversité	0.392	0.388	0.391	-1.10	-0.29
Profondeur	0.077	0.099	0.097	28.64	25.39
Sympathie	-0.382	-0.386	-0.391	-1.10	-2.53
Compréhension générale	0.008	-0.024	-0.018	-413.54	-339.81
Flexibilité	-0.099	-0.103	-0.105	-4.71	-6.49
Informatif	-0.245	-0.223	-0.229	9.19	6.61
Curiosité	-0.574	-0.546	-0.559	4.90	2.69
Score individuel moyen	0.571	0.685	0.651	19.81	13.96
Précision émotion moyenne	0.506	0.605	0.590	19.41	16.48
Score combiné moyen	0.571	0.685	0.651	19.81	13.96

TABLEAU A.4 Résultats obtenus en appliquant un facteur pénalisant de 9 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental 0.25, Pénalité = 9			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.477	-0.488	5.04	2.82
Engagement	-0.625	-0.631	-0.638	-0.97	-2.03
Originalité	-0.328	-0.296	-0.305	9.70	7.08
Pertinence	0.206	0.185	0.191	-9.92	-7.29
Compréhension interloc.	0.036	-0.015	-0.008	-141.12	-121.26
Sémantiquement approprié	0.374	0.355	0.363	-5.03	-2.89
Compréhension réplique	0.329	0.292	0.303	-11.09	-7.92
Fluidité	0.354	0.321	0.332	-9.44	-6.36
Cohérence	0.142	0.138	0.140	-2.98	-1.45
Récupération d'erreurs	-0.654	-0.664	-0.669	-1.53	-2.31
Constance	0.560	0.575	0.578	2.84	3.34
Diversité	0.392	0.387	0.390	-1.32	-0.49
Profondeur	0.077	0.099	0.097	28.42	26.09
Sympathie	-0.382	-0.386	-0.391	-1.08	-2.42
Compréhension générale	0.008	-0.024	-0.018	-413.97	-334.71
Flexibilité	-0.099	-0.103	-0.105	-3.88	-6.33
Informatif	-0.245	-0.223	-0.228	9.17	6.96
Curiosité	-0.574	-0.546	-0.558	4.92	2.87
Score individuel moyen	0.571	0.699	0.673	22.40	17.86
Précision émotion moyenne	0.506	0.633	0.629	24.91	24.18
Score combiné moyen	0.571	0.699	0.673	22.40	17.86

TABLEAU A.5 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (a)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.479	-0.490	4.72	2.51
Engagement	-0.625	-0.631	-0.638	-0.88	-2.07
Originalité	-0.328	-0.299	-0.307	9.02	6.33
Pertinence	0.206	0.186	0.191	-9.63	-7.18
Compréhension interloc.	0.036	-0.013	-0.007	-136.29	-119.47
Sémantiquement approprié	0.374	0.355	0.363	-4.99	-2.84
Compréhension réplique	0.329	0.294	0.303	-10.63	-7.88
Fluidité	0.354	0.322	0.332	-9.11	-6.37
Cohérence	0.142	0.137	0.139	-3.22	-2.06
Récupération d'erreurs	-0.654	-0.665	-0.670	-1.61	-2.49
Constance	0.560	0.575	0.579	2.73	3.42
Diversité	0.392	0.388	0.392	-1.14	-0.01
Profondeur	0.077	0.100	0.098	29.28	27.20
Sympathie	-0.382	-0.385	-0.390	-0.85	-2.34
Compréhension générale	0.008	-0.023	-0.017	-398.58	-327.29
Flexibilité	-0.099	-0.101	-0.103	-2.01	-4.33
Informatif	-0.245	-0.223	-0.229	8.92	6.60
Curiosité	-0.574	-0.545	-0.557	5.02	2.96
Score individuel moyen	0.571	0.727	0.714	27.27	25.00
Précision émotion moyenne	0.506	0.657	0.644	29.67	27.11
Score combiné moyen	0.571	0.727	0.714	27.27	25.00

TABLEAU A.6 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (b)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.479	-0.492	4.59	2.15
Engagement	-0.625	-0.631	-0.639	-0.94	-2.21
Originalité	-0.328	-0.299	-0.307	8.93	6.32
Pertinence	0.206	0.186	0.191	-9.66	-7.02
Compréhension interloc.	0.036	-0.013	-0.006	-137.47	-116.32
Sémantiquement approprié	0.374	0.356	0.364	-4.79	-2.50
Compréhension réplique	0.329	0.294	0.305	-10.60	-7.27
Fluidité	0.354	0.322	0.334	-9.04	-5.78
Cohérence	0.142	0.137	0.138	-3.13	-2.59
Récupération d'erreurs	-0.654	-0.665	-0.670	-1.71	-2.36
Constance	0.560	0.575	0.580	2.77	3.56
Diversité	0.392	0.388	0.393	-1.06	0.07
Profondeur	0.077	0.100	0.098	29.62	26.19
Sympathie	-0.382	-0.385	-0.391	-0.94	-2.55
Compréhension générale	0.008	-0.023	-0.016	-404.36	-314.43
Flexibilité	-0.099	-0.101	-0.104	-2.66	-5.34
Informatif	-0.245	-0.224	-0.231	8.83	5.90
Curiosité	-0.574	-0.546	-0.560	4.89	2.50
Score individuel moyen	0.571	0.731	0.701	27.92	22.73
Précision émotion moyenne	0.506	0.662	0.659	30.77	30.04
Score combiné moyen	0.571	0.731	0.701	27.92	22.73

TABLEAU A.7 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (c)

Apport sentimental 0.25, Variable Interne = 0				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.475	-0.489	5.41	2.60
Engagement	-0.625	-0.629	-0.638	-0.51	-1.99
Originalité	-0.328	-0.297	-0.308	9.37	6.11
Pertinence	0.206	0.185	0.192	-10.01	-6.36
Compréhension interloc.	0.036	-0.014	-0.006	-140.38	-116.46
Sémantiquement approprié	0.374	0.355	0.365	-4.96	-2.29
Compréhension réplique	0.329	0.293	0.305	-10.74	-7.16
Fluidité	0.354	0.322	0.334	-9.18	-5.63
Cohérence	0.142	0.137	0.139	-3.55	-1.97
Récupération d'erreurs	-0.654	-0.663	-0.670	-1.42	-2.45
Constance	0.560	0.575	0.580	2.80	3.65
Diversité	0.392	0.386	0.392	-1.54	0.02
Profondeur	0.077	0.100	0.098	29.26	27.19
Sympathie	-0.382	-0.384	-0.391	-0.61	-2.62
Compréhension générale	0.008	-0.024	-0.017	-414.69	-329.67
Flexibilité	-0.099	-0.100	-0.103	-1.18	-4.10
Informatif	-0.245	-0.223	-0.232	8.94	5.27
Curiosité	-0.574	-0.545	-0.561	5.04	2.28
Score individuel moyen	0.571	0.718	0.701	25.65	22.73
Précision émotion moyenne	0.506	0.664	0.649	31.14	28.21
Score combiné moyen	0.571	0.718	0.701	25.65	22.73

TABLEAU A.8 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (a)

Apport sentimental 0.25, Variable Interne = 0.1				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.479	-0.490	4.77	2.44
Engagement	-0.625	-0.631	-0.639	-0.83	-2.22
Originalité	-0.328	-0.298	-0.308	9.11	6.18
Pertinence	0.206	0.185	0.191	-9.78	-6.90
Compréhension interloc.	0.036	-0.013	-0.007	-136.33	-120.62
Sémantiquement approprié	0.374	0.355	0.364	-4.91	-2.53
Compréhension réplique	0.329	0.294	0.303	-10.68	-7.69
Fluidité	0.354	0.322	0.332	-9.11	-6.20
Cohérence	0.142	0.137	0.139	-3.27	-1.70
Récupération d'erreurs	-0.654	-0.664	-0.671	-1.51	-2.57
Constance	0.560	0.574	0.580	2.65	3.63
Diversité	0.392	0.387	0.393	-1.26	0.20
Profondeur	0.077	0.100	0.098	29.14	27.07
Sympathie	-0.382	-0.384	-0.391	-0.77	-2.49
Compréhension générale	0.008	-0.022	-0.017	-395.80	-331.05
Flexibilité	-0.099	-0.101	-0.103	-2.04	-4.63
Informatif	-0.245	-0.223	-0.230	8.96	6.21
Curiosité	-0.574	-0.545	-0.558	5.08	2.87
Score individuel moyen	0.571	0.724	0.705	26.62	23.38
Précision émotion moyenne	0.506	0.657	0.660	29.67	30.40
Score combiné moyen	0.571	0.724	0.705	26.62	23.38

TABLEAU A.9 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (b)

Apport sentimental 0.25, Variable Interne = 0.1				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.488	5.37	2.96
Engagement	-0.625	-0.628	-0.637	-0.49	-1.82
Originalité	-0.328	-0.297	-0.307	9.49	6.55
Pertinence	0.206	0.185	0.191	-9.98	-6.96
Compréhension interloc.	0.036	-0.014	-0.007	-139.94	-119.91
Sémantiquement approprié	0.374	0.355	0.365	-4.94	-2.43
Compréhension réplique	0.329	0.293	0.304	-10.77	-7.52
Fluidité	0.354	0.322	0.334	-9.12	-5.82
Cohérence	0.142	0.137	0.139	-3.51	-2.35
Récupération d'erreurs	-0.654	-0.663	-0.668	-1.40	-2.12
Constance	0.560	0.575	0.580	2.70	3.56
Diversité	0.392	0.386	0.391	-1.58	-0.39
Profondeur	0.077	0.100	0.099	28.95	27.77
Sympathie	-0.382	-0.384	-0.390	-0.61	-2.26
Compréhension générale	0.008	-0.024	-0.017	-417.29	-321.26
Flexibilité	-0.099	-0.100	-0.103	-1.59	-4.73
Informatif	-0.245	-0.224	-0.230	8.89	6.15
Curiosité	-0.574	-0.545	-0.559	5.06	2.63
Score individuel moyen	0.571	0.729	0.716	27.60	25.32
Précision émotion moyenne	0.506	0.666	0.662	31.50	30.77
Score combiné moyen	0.571	0.729	0.716	27.60	25.32

TABLEAU A.10 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (a)

Apport sentimental 0.25, Variable Interne = 0.2				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.479	-0.490	4.69	2.45
Engagement	-0.625	-0.631	-0.638	-0.86	-1.96
Originalité	-0.328	-0.299	-0.307	9.02	6.36
Pertinence	0.206	0.186	0.191	-9.62	-7.24
Compréhension interloc.	0.036	-0.013	-0.005	-135.60	-113.91
Sémantiquement approprié	0.374	0.356	0.363	-4.85	-2.79
Compréhension réplique	0.329	0.294	0.304	-10.58	-7.66
Fluidité	0.354	0.322	0.333	-9.03	-6.09
Cohérence	0.142	0.138	0.140	-3.03	-1.55
Récupération d'erreurs	-0.654	-0.664	-0.669	-1.58	-2.23
Constance	0.560	0.575	0.577	2.71	3.12
Diversité	0.392	0.388	0.391	-1.19	-0.38
Profondeur	0.077	0.100	0.098	29.71	26.27
Sympathie	-0.382	-0.385	-0.390	-0.87	-2.13
Compréhension générale	0.008	-0.023	-0.016	-399.15	-313.62
Flexibilité	-0.099	-0.100	-0.103	-1.71	-4.24
Informatif	-0.245	-0.224	-0.229	8.73	6.54
Curiosité	-0.574	-0.545	-0.558	4.98	2.87
Score individuel moyen	0.571	0.725	0.718	26.95	25.65
Précision émotion moyenne	0.506	0.657	0.659	29.67	30.04
Score combiné moyen	0.571	0.725	0.718	26.95	25.65

TABLEAU A.11 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (b)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0.2			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.489	5.34	2.65
Engagement	-0.625	-0.629	-0.639	-0.55	-2.15
Originalité	-0.328	-0.298	-0.307	9.31	6.38
Pertinence	0.206	0.185	0.190	-10.04	-7.37
Compréhension interloc.	0.036	-0.014	-0.008	-139.88	-122.25
Sémantiquement approprié	0.374	0.355	0.365	-4.95	-2.34
Compréhension réplique	0.329	0.294	0.304	-10.66	-7.39
Fluidité	0.354	0.322	0.334	-9.07	-5.76
Cohérence	0.142	0.137	0.139	-3.57	-2.20
Récupération d'erreurs	-0.654	-0.664	-0.672	-1.45	-2.73
Constance	0.560	0.575	0.581	2.83	3.87
Diversité	0.392	0.386	0.392	-1.60	-0.13
Profondeur	0.077	0.100	0.098	29.19	27.18
Sympathie	-0.382	-0.384	-0.391	-0.64	-2.51
Compréhension générale	0.008	-0.024	-0.019	-415.81	-348.69
Flexibilité	-0.099	-0.100	-0.104	-1.49	-5.25
Informatif	-0.245	-0.223	-0.231	8.94	5.94
Curiosité	-0.574	-0.545	-0.560	4.99	2.49
Score individuel moyen	0.571	0.720	0.699	25.97	22.40
Précision émotion moyenne	0.506	0.662	0.659	30.77	30.04
Score combiné moyen	0.571	0.720	0.699	25.97	22.40

TABLEAU A.12 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (a)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0.3			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.479	-0.492	4.72	2.17
Engagement	-0.625	-0.633	-0.642	-1.24	-2.70
Originalité	-0.328	-0.298	-0.307	9.34	6.44
Pertinence	0.206	0.185	0.191	-9.86	-6.95
Compréhension interloc.	0.036	-0.015	-0.007	-142.77	-120.35
Sémantiquement approprié	0.374	0.355	0.364	-4.95	-2.71
Compréhension réplique	0.329	0.293	0.303	-10.94	-7.91
Fluidité	0.354	0.321	0.332	-9.29	-6.26
Cohérence	0.142	0.138	0.138	-3.09	-2.42
Récupération d'erreurs	-0.654	-0.666	-0.671	-1.84	-2.58
Constance	0.560	0.576	0.580	2.91	3.62
Diversité	0.392	0.388	0.393	-1.15	0.16
Profondeur	0.077	0.099	0.097	28.47	25.27
Sympathie	-0.382	-0.387	-0.393	-1.49	-3.12
Compréhension générale	0.008	-0.024	-0.019	-419.37	-351.61
Flexibilité	-0.099	-0.102	-0.106	-3.12	-7.42
Informatif	-0.245	-0.223	-0.231	9.16	6.00
Curiosité	-0.574	-0.547	-0.562	4.71	2.18
Score individuel moyen	0.571	0.681	0.664	19.16	16.23
Précision émotion moyenne	0.506	0.616	0.610	21.61	20.51
Score combiné moyen	0.571	0.681	0.664	19.16	16.23

TABLEAU A.13 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (b)

Apport sentimental 0.25, Variable Interne = 0.3				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.485	5.35	3.39
Engagement	-0.625	-0.632	-0.638	-1.10	-2.08
Originalité	-0.328	-0.296	-0.306	9.67	6.91
Pertinence	0.206	0.185	0.190	-10.01	-7.72
Compréhension interloc.	0.036	-0.016	-0.010	-145.20	-127.07
Sémantiquement approprié	0.374	0.355	0.362	-4.93	-3.06
Compréhension réplique	0.329	0.293	0.301	-11.00	-8.34
Fluidité	0.354	0.321	0.331	-9.25	-6.63
Cohérence	0.142	0.137	0.138	-3.47	-2.45
Récupération d'erreurs	-0.654	-0.666	-0.669	-1.78	-2.31
Constance	0.560	0.576	0.579	3.00	3.45
Diversité	0.392	0.387	0.390	-1.43	-0.50
Profondeur	0.077	0.099	0.097	28.33	25.76
Sympathie	-0.382	-0.387	-0.391	-1.37	-2.56
Compréhension générale	0.008	-0.025	-0.019	-428.08	-357.28
Flexibilité	-0.099	-0.103	-0.104	-4.21	-5.62
Informatif	-0.245	-0.223	-0.229	9.08	6.66
Curiosité	-0.574	-0.548	-0.558	4.60	2.78
Score individuel moyen	0.571	0.686	0.688	20.13	20.45
Précision émotion moyenne	0.506	0.620	0.620	22.34	22.34
Score combiné moyen	0.571	0.686	0.688	20.13	20.45

TABLEAU A.14 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (a)

Apport sentimental 0.25, Variable Interne = 0.4				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.478	-0.489	4.89	2.69
Engagement	-0.625	-0.633	-0.639	-1.22	-2.18
Originalité	-0.328	-0.295	-0.305	9.98	7.03
Pertinence	0.206	0.185	0.191	-9.97	-7.09
Compréhension interloc.	0.036	-0.015	-0.006	-141.26	-117.91
Sémantiquement approprié	0.374	0.356	0.364	-4.78	-2.71
Compréhension réplique	0.329	0.293	0.303	-10.84	-7.85
Fluidité	0.354	0.321	0.332	-9.28	-6.32
Cohérence	0.142	0.138	0.140	-2.71	-1.59
Récupération d'erreurs	-0.654	-0.665	-0.669	-1.73	-2.26
Constance	0.560	0.576	0.578	2.90	3.33
Diversité	0.392	0.388	0.392	-1.16	-0.17
Profondeur	0.077	0.099	0.096	27.88	24.28
Sympathie	-0.382	-0.386	-0.391	-1.31	-2.37
Compréhension générale	0.008	-0.024	-0.017	-414.54	-321.34
Flexibilité	-0.099	-0.103	-0.106	-4.43	-7.12
Informatif	-0.245	-0.222	-0.229	9.51	6.63
Curiosité	-0.574	-0.547	-0.559	4.76	2.63
Score individuel moyen	0.571	0.690	0.683	20.78	19.48
Précision émotion moyenne	0.506	0.627	0.644	23.81	27.11
Score combiné moyen	0.571	0.690	0.683	20.78	19.48

TABLEAU A.15 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (b)

Apport sentimental 0.25, Variable Interne = 0.4				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.487	5.32	3.10
Engagement	-0.625	-0.631	-0.638	-0.94	-2.04
Originalité	-0.328	-0.296	-0.306	9.81	6.88
Pertinence	0.206	0.185	0.191	-10.10	-7.22
Compréhension interloc.	0.036	-0.016	-0.008	-144.41	-122.23
Sémantiquement approprié	0.374	0.355	0.363	-5.09	-2.95
Compréhension réplique	0.329	0.292	0.302	-11.14	-8.13
Fluidité	0.354	0.321	0.331	-9.41	-6.51
Cohérence	0.142	0.137	0.139	-3.21	-2.00
Récupération d'erreurs	-0.654	-0.664	-0.668	-1.57	-2.08
Constance	0.560	0.576	0.578	2.89	3.31
Diversité	0.392	0.387	0.390	-1.30	-0.54
Profondeur	0.077	0.099	0.097	28.39	25.51
Sympathie	-0.382	-0.386	-0.391	-1.22	-2.60
Compréhension générale	0.008	-0.024	-0.018	-419.67	-340.75
Flexibilité	-0.099	-0.103	-0.105	-3.90	-6.70
Informatif	-0.245	-0.223	-0.230	9.12	6.31
Curiosité	-0.574	-0.546	-0.558	4.91	2.73
Score individuel moyen	0.571	0.692	0.683	21.10	19.48
Précision émotion moyenne	0.506	0.612	0.607	20.88	19.78
Score combiné moyen	0.571	0.692	0.683	21.10	19.48

TABLEAU A.16 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (a)

Apport sentimental 0.25, Variable Interne = 0.5				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.477	-0.488	5.07	2.95
Engagement	-0.625	-0.632	-0.638	-1.07	-1.96
Originalité	-0.328	-0.297	-0.306	9.54	6.84
Pertinence	0.206	0.184	0.190	-10.30	-7.55
Compréhension interloc.	0.036	-0.016	-0.008	-144.18	-121.98
Sémantiquement approprié	0.374	0.355	0.361	-5.04	-3.33
Compréhension réplique	0.329	0.292	0.300	-11.24	-8.60
Fluidité	0.354	0.320	0.329	-9.52	-7.01
Cohérence	0.142	0.138	0.139	-2.97	-2.14
Récupération d'erreurs	-0.654	-0.665	-0.669	-1.65	-2.27
Constance	0.560	0.574	0.577	2.61	3.19
Diversité	0.392	0.386	0.390	-1.60	-0.58
Profondeur	0.077	0.099	0.096	27.75	24.13
Sympathie	-0.382	-0.387	-0.391	-1.34	-2.51
Compréhension générale	0.008	-0.025	-0.019	-426.93	-348.60
Flexibilité	-0.099	-0.103	-0.106	-4.73	-7.53
Informatif	-0.245	-0.221	-0.228	9.77	7.18
Curiosité	-0.574	-0.545	-0.557	5.01	3.04
Score individuel moyen	0.571	0.688	0.679	20.45	18.83
Précision émotion moyenne	0.506	0.612	0.620	20.88	22.34
Score combiné moyen	0.571	0.688	0.679	20.45	18.83

TABLEAU A.17 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (b)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0.5			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.477	-0.491	4.98	2.25
Engagement	-0.625	-0.633	-0.643	-1.21	-2.87
Originalité	-0.328	-0.297	-0.308	9.58	6.07
Pertinence	0.206	0.185	0.190	-10.13	-7.35
Compréhension interloc.	0.036	-0.015	-0.008	-143.20	-123.50
Sémantiquement approprié	0.374	0.354	0.363	-5.37	-2.83
Compréhension réplique	0.329	0.291	0.302	-11.50	-8.13
Fluidité	0.354	0.320	0.331	-9.77	-6.51
Cohérence	0.142	0.137	0.139	-3.69	-2.32
Récupération d'erreurs	-0.654	-0.664	-0.671	-1.57	-2.51
Constance	0.560	0.575	0.580	2.72	3.67
Diversité	0.392	0.388	0.393	-1.16	0.09
Profondeur	0.077	0.098	0.096	27.22	24.39
Sympathie	-0.382	-0.386	-0.393	-1.30	-2.91
Compréhension générale	0.008	-0.025	-0.019	-425.40	-346.65
Flexibilité	-0.099	-0.103	-0.106	-4.38	-7.35
Informatif	-0.245	-0.223	-0.230	9.30	6.40
Curiosité	-0.574	-0.546	-0.560	4.93	2.47
Score individuel moyen	0.571	0.677	0.640	18.51	12.01
Précision émotion moyenne	0.506	0.607	0.590	19.78	16.48
Score combiné moyen	0.571	0.677	0.640	18.51	12.01

TABLEAU A.18 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (c)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0.5			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.488	5.26	2.89
Engagement	-0.625	-0.632	-0.639	-1.02	-2.15
Originalité	-0.328	-0.296	-0.306	9.67	6.71
Pertinence	0.206	0.184	0.190	-10.32	-7.54
Compréhension interloc.	0.036	-0.015	-0.008	-143.43	-121.33
Sémantiquement approprié	0.374	0.355	0.364	-4.90	-2.75
Compréhension réplique	0.329	0.293	0.302	-11.02	-8.06
Fluidité	0.354	0.321	0.332	-9.34	-6.24
Cohérence	0.142	0.138	0.140	-2.63	-1.65
Récupération d'erreurs	-0.654	-0.665	-0.669	-1.73	-2.34
Constance	0.560	0.576	0.579	3.00	3.46
Diversité	0.392	0.387	0.391	-1.37	-0.45
Profondeur	0.077	0.099	0.096	28.15	24.85
Sympathie	-0.382	-0.386	-0.391	-1.25	-2.48
Compréhension générale	0.008	-0.024	-0.017	-412.57	-330.34
Flexibilité	-0.099	-0.102	-0.105	-3.84	-6.81
Informatif	-0.245	-0.223	-0.229	9.30	6.52
Curiosité	-0.574	-0.547	-0.560	4.70	2.52
Score individuel moyen	0.571	0.681	0.677	19.16	18.51
Précision émotion moyenne	0.506	0.607	0.610	19.78	20.51
Score combiné moyen	0.571	0.681	0.677	19.16	18.51

TABLEAU A.19 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (a)

Apport sentimental 0.25, Variable Interne = 0.6				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.479	-0.490	4.75	2.52
Engagement	-0.625	-0.633	-0.641	-1.19	-2.49
Originalité	-0.328	-0.298	-0.306	9.20	6.67
Pertinence	0.206	0.185	0.190	-10.13	-7.68
Compréhension interloc.	0.036	-0.015	-0.009	-142.54	-124.96
Sémantiquement approprié	0.374	0.355	0.363	-4.90	-2.85
Compréhension réplique	0.329	0.293	0.302	-10.81	-8.25
Fluidité	0.354	0.322	0.331	-9.14	-6.66
Cohérence	0.142	0.138	0.139	-2.98	-2.04
Récupération d'erreurs	-0.654	-0.666	-0.672	-1.88	-2.73
Constance	0.560	0.576	0.579	2.97	3.50
Diversité	0.392	0.388	0.392	-1.07	0.00
Profondeur	0.077	0.099	0.097	27.92	25.71
Sympathie	-0.382	-0.387	-0.392	-1.37	-2.83
Compréhension générale	0.008	-0.024	-0.020	-413.32	-360.79
Flexibilité	-0.099	-0.103	-0.107	-4.72	-8.77
Informatif	-0.245	-0.223	-0.229	9.05	6.74
Curiosité	-0.574	-0.547	-0.558	4.79	2.72
Score individuel moyen	0.571	0.692	0.666	21.10	16.56
Précision émotion moyenne	0.506	0.614	0.603	21.25	19.05
Score combiné moyen	0.571	0.692	0.666	21.10	16.56

TABLEAU A.20 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (b)

Apport sentimental 0.25, Variable Interne = 0.6				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.486	5.31	3.30
Engagement	-0.625	-0.632	-0.638	-1.08	-1.97
Originalité	-0.328	-0.297	-0.305	9.62	7.12
Pertinence	0.206	0.184	0.189	-10.34	-8.02
Compréhension interloc.	0.036	-0.016	-0.008	-144.51	-123.82
Sémantiquement approprié	0.374	0.356	0.363	-4.83	-2.90
Compréhension réplique	0.329	0.293	0.302	-11.02	-8.26
Fluidité	0.354	0.321	0.332	-9.26	-6.41
Cohérence	0.142	0.138	0.138	-3.09	-2.74
Récupération d'erreurs	-0.654	-0.665	-0.668	-1.61	-2.12
Constance	0.560	0.576	0.579	2.95	3.39
Diversité	0.392	0.388	0.391	-1.25	-0.28
Profondeur	0.077	0.099	0.096	28.33	24.37
Sympathie	-0.382	-0.386	-0.391	-1.17	-2.46
Compréhension générale	0.008	-0.025	-0.019	-429.76	-350.94
Flexibilité	-0.099	-0.103	-0.105	-3.99	-6.79
Informatif	-0.245	-0.223	-0.230	9.10	6.30
Curiosité	-0.574	-0.546	-0.558	4.87	2.72
Score individuel moyen	0.571	0.657	0.644	14.94	12.66
Précision émotion moyenne	0.506	0.609	0.603	20.15	19.05
Score combiné moyen	0.571	0.657	0.644	14.94	12.66

TABLEAU A.21 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (a)

Apport sentimental 0.25, Variable Interne = 0.7				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.478	-0.490	4.85	2.40
Engagement	-0.625	-0.633	-0.642	-1.26	-2.69
Originalité	-0.328	-0.296	-0.306	9.76	6.71
Pertinence	0.206	0.186	0.191	-9.59	-7.24
Compréhension interloc.	0.036	-0.016	-0.009	-145.52	-126.46
Sémantiquement approprié	0.374	0.356	0.364	-4.83	-2.56
Compréhension réplique	0.329	0.293	0.302	-10.99	-8.12
Fluidité	0.354	0.321	0.332	-9.33	-6.31
Cohérence	0.142	0.138	0.139	-2.66	-2.12
Récupération d'erreurs	-0.654	-0.667	-0.673	-1.97	-2.84
Constance	0.560	0.577	0.580	3.10	3.69
Diversité	0.392	0.388	0.393	-1.03	0.12
Profondeur	0.077	0.100	0.097	29.51	24.97
Sympathie	-0.382	-0.387	-0.394	-1.48	-3.17
Compréhension générale	0.008	-0.025	-0.019	-425.68	-347.48
Flexibilité	-0.099	-0.103	-0.107	-4.37	-8.49
Informatif	-0.245	-0.222	-0.229	9.48	6.46
Curiosité	-0.574	-0.546	-0.560	4.90	2.40
Score individuel moyen	0.571	0.685	0.655	19.81	14.61
Précision émotion moyenne	0.506	0.622	0.607	22.71	19.78
Score combiné moyen	0.571	0.685	0.655	19.81	14.61

TABLEAU A.22 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (b)

Apport sentimental 0.25, Variable Interne = 0.7				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.475	-0.486	5.43	3.37
Engagement	-0.625	-0.632	-0.638	-0.99	-2.03
Originalité	-0.328	-0.296	-0.303	9.82	7.83
Pertinence	0.206	0.184	0.189	-10.48	-7.93
Compréhension interloc.	0.036	-0.017	-0.011	-147.70	-130.07
Sémantiquement approprié	0.374	0.355	0.363	-5.06	-2.86
Compréhension réplique	0.329	0.292	0.301	-11.27	-8.35
Fluidité	0.354	0.321	0.332	-9.37	-6.39
Cohérence	0.142	0.136	0.137	-4.21	-3.60
Récupération d'erreurs	-0.654	-0.665	-0.669	-1.60	-2.29
Constance	0.560	0.576	0.580	2.90	3.66
Diversité	0.392	0.386	0.390	-1.73	-0.60
Profondeur	0.077	0.099	0.097	27.89	25.60
Sympathie	-0.382	-0.386	-0.392	-1.25	-2.73
Compréhension générale	0.008	-0.025	-0.020	-436.85	-365.67
Flexibilité	-0.099	-0.103	-0.106	-4.57	-7.32
Informatif	-0.245	-0.222	-0.228	9.69	7.02
Curiosité	-0.574	-0.546	-0.558	4.85	2.75
Score individuel moyen	0.571	0.685	0.673	19.81	17.86
Précision émotion moyenne	0.506	0.622	0.612	22.71	20.88
Score combiné moyen	0.571	0.685	0.673	19.81	17.86

TABLEAU A.23 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (a)

Apport sentimental 0.25, Variable Interne = 0.8				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.478	-0.490	4.88	2.58
Engagement	-0.625	-0.631	-0.638	-0.98	-2.02
Originalité	-0.328	-0.297	-0.307	9.66	6.52
Pertinence	0.206	0.185	0.191	-10.13	-7.16
Compréhension interloc.	0.036	-0.014	-0.005	-140.49	-115.36
Sémantiquement approprié	0.374	0.355	0.363	-5.07	-3.01
Compréhension réplique	0.329	0.292	0.302	-11.06	-8.07
Fluidité	0.354	0.321	0.331	-9.40	-6.44
Cohérence	0.142	0.138	0.139	-2.80	-2.09
Récupération d'erreurs	-0.654	-0.665	-0.669	-1.67	-2.30
Constance	0.560	0.575	0.576	2.72	2.95
Diversité	0.392	0.387	0.392	-1.26	-0.02
Profondeur	0.077	0.099	0.095	28.22	23.52
Sympathie	-0.382	-0.386	-0.391	-1.31	-2.47
Compréhension générale	0.008	-0.024	-0.018	-422.60	-336.48
Flexibilité	-0.099	-0.104	-0.106	-5.15	-7.70
Informatif	-0.245	-0.222	-0.230	9.48	6.19
Curiosité	-0.574	-0.546	-0.559	4.92	2.66
Score individuel moyen	0.571	0.672	0.646	17.53	12.99
Précision émotion moyenne	0.506	0.603	0.588	19.05	16.12
Score combiné moyen	0.571	0.672	0.646	17.53	12.99

TABLEAU A.24 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (b)

Apport sentimental 0.25, Variable Interne = 0.8				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.475	-0.486	5.52	3.37
Engagement	-0.625	-0.631	-0.638	-0.96	-2.03
Originalité	-0.328	-0.296	-0.305	9.84	6.94
Pertinence	0.206	0.184	0.189	-10.60	-8.22
Compréhension interloc.	0.036	-0.016	-0.010	-146.09	-129.36
Sémantiquement approprié	0.374	0.355	0.362	-5.15	-3.04
Compréhension réplique	0.329	0.291	0.300	-11.49	-8.87
Fluidité	0.354	0.320	0.329	-9.61	-7.07
Cohérence	0.142	0.137	0.139	-3.60	-1.92
Récupération d'erreurs	-0.654	-0.665	-0.669	-1.58	-2.26
Constance	0.560	0.575	0.579	2.67	3.44
Diversité	0.392	0.386	0.389	-1.63	-0.80
Profondeur	0.077	0.099	0.097	27.57	25.76
Sympathie	-0.382	-0.386	-0.391	-1.21	-2.58
Compréhension générale	0.008	-0.026	-0.021	-438.57	-378.33
Flexibilité	-0.099	-0.102	-0.105	-3.58	-6.90
Informatif	-0.245	-0.222	-0.227	9.67	7.39
Curiosité	-0.574	-0.545	-0.556	5.11	3.12
Score individuel moyen	0.571	0.683	0.677	19.48	18.51
Précision émotion moyenne	0.506	0.609	0.597	20.15	17.95
Score combiné moyen	0.571	0.683	0.677	19.48	18.51

TABLEAU A.25 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (a)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0.9			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.487	5.19	2.99
Engagement	-0.625	-0.631	-0.637	-0.89	-1.87
Originalité	-0.328	-0.296	-0.305	9.87	6.98
Pertinence	0.206	0.185	0.190	-10.01	-7.36
Compréhension interloc.	0.036	-0.015	-0.007	-142.12	-118.89
Sémantiquement approprié	0.374	0.355	0.363	-5.04	-2.88
Compréhension réplique	0.329	0.292	0.302	-11.20	-8.11
Fluidité	0.354	0.320	0.331	-9.64	-6.59
Cohérence	0.142	0.137	0.139	-3.21	-2.06
Récupération d'erreurs	-0.654	-0.664	-0.667	-1.46	-2.04
Constance	0.560	0.574	0.577	2.64	3.07
Diversité	0.392	0.388	0.391	-1.15	-0.38
Profondeur	0.077	0.099	0.096	27.66	23.95
Sympathie	-0.382	-0.386	-0.389	-1.08	-2.08
Compréhension générale	0.008	-0.024	-0.016	-414.11	-317.53
Flexibilité	-0.099	-0.103	-0.104	-3.91	-5.37
Informatif	-0.245	-0.222	-0.228	9.61	6.88
Curiosité	-0.574	-0.545	-0.558	5.06	2.88
Score individuel moyen	0.571	0.677	0.675	18.51	18.18
Précision émotion moyenne	0.506	0.616	0.609	21.61	20.15
Score combiné moyen	0.571	0.677	0.675	18.51	18.18

TABLEAU A.26 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (b)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 0.9			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.475	-0.487	5.45	3.15
Engagement	-0.625	-0.630	-0.638	-0.79	-1.96
Originalité	-0.328	-0.296	-0.305	9.72	7.14
Pertinence	0.206	0.183	0.189	-10.81	-8.15
Compréhension interloc.	0.036	-0.016	-0.010	-145.96	-129.15
Sémantiquement approprié	0.374	0.354	0.362	-5.26	-3.09
Compréhension réplique	0.329	0.292	0.301	-11.25	-8.34
Fluidité	0.354	0.320	0.331	-9.54	-6.64
Cohérence	0.142	0.136	0.137	-4.04	-3.12
Récupération d'erreurs	-0.654	-0.664	-0.670	-1.47	-2.36
Constance	0.560	0.574	0.578	2.56	3.31
Diversité	0.392	0.386	0.390	-1.71	-0.66
Profondeur	0.077	0.098	0.096	26.96	24.42
Sympathie	-0.382	-0.386	-0.392	-1.13	-2.75
Compréhension générale	0.008	-0.025	-0.021	-437.42	-375.39
Flexibilité	-0.099	-0.102	-0.105	-3.37	-6.14
Informatif	-0.245	-0.222	-0.228	9.57	6.94
Curiosité	-0.574	-0.544	-0.557	5.15	2.99
Score individuel moyen	0.571	0.666	0.651	16.56	13.96
Précision émotion moyenne	0.506	0.610	0.609	20.51	20.15
Score combiné moyen	0.571	0.666	0.651	16.56	13.96

TABLEAU A.27 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (a)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.477	-0.489	5.12	2.76
Engagement	-0.625	-0.632	-0.641	-1.08	-2.55
Originalité	-0.328	-0.297	-0.306	9.62	6.75
Pertinence	0.206	0.185	0.191	-10.16	-7.23
Compréhension interloc.	0.036	-0.015	-0.010	-143.15	-127.80
Sémantiquement approprié	0.374	0.355	0.363	-5.08	-2.78
Compréhension réplique	0.329	0.291	0.302	-11.36	-8.27
Fluidité	0.354	0.320	0.331	-9.57	-6.61
Cohérence	0.142	0.137	0.139	-3.19	-2.30
Récupération d'erreurs	-0.654	-0.665	-0.670	-1.66	-2.44
Constance	0.560	0.575	0.580	2.81	3.63
Diversité	0.392	0.388	0.392	-1.18	-0.21
Profondeur	0.077	0.099	0.097	28.01	25.38
Sympathie	-0.382	-0.387	-0.394	-1.45	-3.15
Compréhension générale	0.008	-0.025	-0.019	-426.41	-356.76
Flexibilité	-0.099	-0.103	-0.106	-4.17	-6.99
Informatif	-0.245	-0.223	-0.229	9.26	6.79
Curiosité	-0.574	-0.545	-0.559	4.99	2.62
Score individuel moyen	0.571	0.670	0.666	17.21	16.56
Précision émotion moyenne	0.506	0.607	0.601	19.78	18.68
Score combiné moyen	0.571	0.670	0.666	17.21	16.56

TABLEAU A.28 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (b)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.477	-0.488	5.06	2.90
Engagement	-0.625	-0.632	-0.637	-1.01	-1.94
Originalité	-0.328	-0.297	-0.305	9.57	7.21
Pertinence	0.206	0.186	0.191	-9.73	-6.93
Compréhension interloc.	0.036	-0.015	-0.007	-141.64	-120.10
Sémantiquement approprié	0.374	0.355	0.362	-5.00	-3.13
Compréhension réplique	0.329	0.292	0.302	-11.09	-8.13
Fluidité	0.354	0.321	0.332	-9.42	-6.40
Cohérence	0.142	0.138	0.139	-2.69	-2.31
Récupération d'erreurs	-0.654	-0.665	-0.668	-1.71	-2.06
Constance	0.560	0.575	0.576	2.66	2.86
Diversité	0.392	0.388	0.391	-1.22	-0.37
Profondeur	0.077	0.099	0.097	28.17	25.09
Sympathie	-0.382	-0.387	-0.391	-1.34	-2.53
Compréhension générale	0.008	-0.024	-0.018	-417.48	-342.43
Flexibilité	-0.099	-0.103	-0.105	-4.71	-6.68
Informatif	-0.245	-0.222	-0.228	9.55	6.89
Curiosité	-0.574	-0.546	-0.557	4.95	2.97
Score individuel moyen	0.571	0.664	0.657	16.23	14.94
Précision émotion moyenne	0.506	0.592	0.596	16.85	17.58
Score combiné moyen	0.571	0.664	0.657	16.23	14.94

TABLEAU A.29 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (c)

Caractéristiques:	Apport sentimental 0.25, Variable Interne = 1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.476	-0.486	5.35	3.19
Engagement	-0.625	-0.631	-0.638	-0.86	-2.05
Originalité	-0.328	-0.296	-0.306	9.66	6.76
Pertinence	0.206	0.184	0.189	-10.52	-7.96
Compréhension interloc.	0.036	-0.016	-0.009	-144.14	-125.20
Sémantiquement approprié	0.374	0.355	0.364	-5.04	-2.71
Compréhension réplique	0.329	0.292	0.303	-11.05	-7.93
Fluidité	0.354	0.321	0.332	-9.33	-6.29
Cohérence	0.142	0.137	0.138	-3.68	-2.55
Récupération d'erreurs	-0.654	-0.664	-0.669	-1.45	-2.23
Constance	0.560	0.575	0.579	2.68	3.42
Diversité	0.392	0.386	0.391	-1.64	-0.43
Profondeur	0.077	0.098	0.096	26.65	23.86
Sympathie	-0.382	-0.385	-0.391	-1.02	-2.39
Compréhension générale	0.008	-0.024	-0.020	-423.05	-359.68
Flexibilité	-0.099	-0.103	-0.106	-3.93	-7.28
Informatif	-0.245	-0.222	-0.229	9.34	6.72
Curiosité	-0.574	-0.546	-0.559	4.88	2.64
Score individuel moyen	0.571	0.679	0.668	18.83	16.88
Précision émotion moyenne	0.506	0.596	0.586	17.58	15.75
Score combiné moyen	0.571	0.679	0.668	18.83	16.88

**ANNEXE B RÉSULTATS DÉTAILLÉS OBTENUS LORS DE
L'ÉVALUATION DU SYSTÈME, INITIALISATION FAITE AVEC UN
APPORT SENTIMENTAL NUL**

TABLEAU B.1 Résultats obtenus en appliquant un facteur pénalisant de 0 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental nul, Pénalité = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.484	6.18	3.76
Engagement	-0.625	-0.628	-0.636	-0.40	-1.64
Originalité	-0.328	-0.294	-0.303	10.43	7.53
Pertinence	0.206	0.184	0.190	-10.54	-7.47
Compréhension interloc.	0.036	-0.015	-0.007	-141.28	-118.89
Sémantiquement approprié	0.374	0.355	0.364	-4.95	-2.70
Compréhension réplique	0.329	0.293	0.304	-10.82	-7.69
Fluidité	0.354	0.322	0.333	-9.07	-6.00
Cohérence	0.142	0.135	0.137	-4.54	-3.21
Récupération d'erreurs	-0.654	-0.659	-0.664	-0.78	-1.57
Constance	0.560	0.572	0.575	2.26	2.72
Diversité	0.392	0.386	0.389	-1.67	-0.76
Profondeur	0.077	0.096	0.094	23.86	21.92
Sympathie	-0.382	-0.385	-0.390	-0.90	-2.23
Compréhension générale	0.008	-0.023	-0.017	-406.35	-327.91
Flexibilité	-0.099	-0.103	-0.105	-4.38	-6.37
Informatif	-0.245	-0.223	-0.231	9.28	5.87
Curiosité	-0.574	-0.546	-0.558	4.88	2.79
Score individuel moyen	0.579	0.673	0.668	16.35	15.38
Précision émotion moyenne	0.506	0.603	0.609	19.05	20.15
Score combiné moyen	0.579	0.673	0.668	16.35	15.38

TABLEAU B.2 Résultats obtenus en appliquant un facteur pénalisant de 3 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental nul, Pénalité = 3			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.92	3.52
Engagement	-0.625	-0.629	-0.635	-0.53	-1.50
Originalité	-0.328	-0.294	-0.304	10.29	7.27
Pertinence	0.206	0.184	0.190	-10.61	-7.79
Compréhension interloc.	0.036	-0.015	-0.006	-142.34	-117.77
Sémantiquement approprié	0.374	0.355	0.363	-4.96	-2.81
Compréhension réplique	0.329	0.293	0.303	-10.91	-7.74
Fluidité	0.354	0.322	0.333	-9.08	-6.04
Cohérence	0.142	0.136	0.137	-4.02	-3.21
Récupération d'erreurs	-0.654	-0.661	-0.664	-1.00	-1.57
Constance	0.560	0.573	0.575	2.34	2.72
Diversité	0.392	0.386	0.390	-1.52	-0.52
Profondeur	0.077	0.095	0.092	23.34	19.47
Sympathie	-0.382	-0.385	-0.389	-0.95	-2.01
Compréhension générale	0.008	-0.024	-0.017	-412.41	-329.61
Flexibilité	-0.099	-0.104	-0.106	-5.07	-7.53
Informatif	-0.245	-0.222	-0.231	9.44	5.94
Curiosité	-0.574	-0.547	-0.558	4.78	2.76
Score individuel moyen	0.579	0.679	0.660	17.31	14.10
Précision émotion moyenne	0.506	0.601	0.588	18.68	16.12
Score combiné moyen	0.579	0.679	0.660	17.31	14.10

TABLEAU B.3 Résultats obtenus en appliquant un facteur pénalisant de 6 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental nul, Pénalité = 6			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.472	-0.484	6.13	3.76
Engagement	-0.625	-0.627	-0.635	-0.32	-1.60
Originalité	-0.328	-0.295	-0.305	10.17	7.21
Pertinence	0.206	0.183	0.189	-10.77	-7.82
Compréhension interloc.	0.036	-0.015	-0.009	-143.02	-124.74
Sémantiquement approprié	0.374	0.354	0.362	-5.23	-3.08
Compréhension réplique	0.329	0.292	0.302	-11.16	-8.01
Fluidité	0.354	0.321	0.332	-9.35	-6.36
Cohérence	0.142	0.136	0.137	-4.34	-3.68
Récupération d'erreurs	-0.654	-0.660	-0.665	-0.82	-1.69
Constance	0.560	0.573	0.575	2.33	2.80
Diversité	0.392	0.385	0.390	-1.77	-0.65
Profondeur	0.077	0.096	0.094	24.15	21.14
Sympathie	-0.382	-0.385	-0.389	-0.80	-2.01
Compréhension générale	0.008	-0.024	-0.019	-419.63	-349.97
Flexibilité	-0.099	-0.104	-0.106	-5.30	-7.39
Informatif	-0.245	-0.221	-0.229	9.83	6.51
Curiosité	-0.574	-0.545	-0.558	5.02	2.75
Score individuel moyen	0.579	0.666	0.649	15.06	12.18
Précision émotion moyenne	0.506	0.603	0.590	19.05	16.48
Score combiné moyen	0.579	0.666	0.649	15.06	12.18

TABLEAU B.4 Résultats obtenus en appliquant un facteur pénalisant de 9 sur la base de données des scores initialisée en négligeant l'apport sentimental au score individuel des répliques.

Caractéristiques:	Apport sentimental nul, Pénalité = 9			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.472	-0.486	6.13	3.38
Engagement	-0.625	-0.627	-0.636	-0.19	-1.65
Originalité	-0.328	-0.294	-0.307	10.27	6.44
Pertinence	0.206	0.183	0.190	-10.76	-7.79
Compréhension interloc.	0.036	-0.014	-0.007	-139.30	-118.33
Sémantiquement approprié	0.374	0.354	0.362	-5.28	-3.09
Compréhension réplique	0.329	0.293	0.303	-10.92	-7.87
Fluidité	0.354	0.321	0.332	-9.29	-6.36
Cohérence	0.142	0.136	0.138	-3.95	-2.51
Récupération d'erreurs	-0.654	-0.659	-0.665	-0.74	-1.71
Constance	0.560	0.572	0.575	2.28	2.80
Diversité	0.392	0.385	0.390	-1.83	-0.55
Profondeur	0.077	0.096	0.093	24.13	20.83
Sympathie	-0.382	-0.384	-0.390	-0.69	-2.33
Compréhension générale	0.008	-0.023	-0.018	-399.61	-335.22
Flexibilité	-0.099	-0.102	-0.104	-3.37	-5.49
Informatif	-0.245	-0.221	-0.230	9.99	6.33
Curiosité	-0.574	-0.545	-0.557	5.13	2.95
Score individuel moyen	0.579	0.679	0.660	17.31	14.10
Précision émotion moyenne	0.506	0.620	0.599	22.34	18.32
Score combiné moyen	0.579	0.679	0.660	17.31	14.10

TABLEAU B.5 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.484	5.86	3.64
Engagement	-0.625	-0.627	-0.635	-0.29	-1.47
Originalité	-0.328	-0.296	-0.304	9.79	7.27
Pertinence	0.206	0.185	0.191	-10.04	-6.94
Compréhension interloc.	0.036	-0.012	-0.005	-134.09	-112.97
Sémantiquement approprié	0.374	0.355	0.364	-4.96	-2.67
Compréhension réplique	0.329	0.295	0.305	-10.40	-7.16
Fluidité	0.354	0.323	0.334	-8.77	-5.62
Cohérence	0.142	0.136	0.137	-4.36	-3.42
Récupération d'erreurs	-0.654	-0.658	-0.662	-0.53	-1.15
Constance	0.560	0.572	0.574	2.15	2.60
Diversité	0.392	0.386	0.391	-1.55	-0.46
Profondeur	0.077	0.095	0.094	23.41	21.65
Sympathie	-0.382	-0.384	-0.388	-0.77	-1.79
Compréhension générale	0.008	-0.021	-0.015	-382.65	-299.48
Flexibilité	-0.099	-0.101	-0.103	-2.50	-4.06
Informatif	-0.245	-0.223	-0.231	9.20	5.97
Curiosité	-0.574	-0.545	-0.558	4.98	2.82
Score individuel moyen	0.579	0.699	0.690	20.83	19.23
Précision émotion moyenne	0.506	0.631	0.633	24.54	24.91
Score combiné moyen	0.579	0.699	0.690	20.83	19.23

TABLEAU B.6 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.89	3.44
Engagement	-0.625	-0.628	-0.635	-0.35	-1.54
Originalité	-0.328	-0.296	-0.306	9.89	6.85
Pertinence	0.206	0.185	0.191	-10.21	-7.26
Compréhension interloc.	0.036	-0.012	-0.006	-135.03	-116.24
Sémantiquement approprié	0.374	0.355	0.363	-5.06	-2.88
Compréhension réplique	0.329	0.294	0.304	-10.55	-7.43
Fluidité	0.354	0.323	0.333	-8.85	-5.94
Cohérence	0.142	0.135	0.137	-4.83	-3.68
Récupération d'erreurs	-0.654	-0.658	-0.663	-0.60	-1.36
Constance	0.560	0.571	0.574	2.09	2.62
Diversité	0.392	0.386	0.391	-1.58	-0.47
Profondeur	0.077	0.095	0.093	22.90	20.55
Sympathie	-0.382	-0.384	-0.389	-0.73	-1.96
Compréhension générale	0.008	-0.022	-0.016	-389.63	-314.08
Flexibilité	-0.099	-0.101	-0.103	-2.76	-4.83
Informatif	-0.245	-0.223	-0.231	9.19	6.04
Curiosité	-0.574	-0.546	-0.558	4.91	2.78
Score individuel moyen	0.579	0.707	0.699	22.12	20.83
Précision émotion moyenne	0.506	0.633	0.623	24.91	23.08
Score combiné moyen	0.579	0.707	0.699	22.12	20.83

TABLEAU B.7 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0. (c)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.484	6.28	3.66
Engagement	-0.625	-0.625	-0.635	0.05	-1.60
Originalité	-0.328	-0.295	-0.305	10.26	7.22
Pertinence	0.206	0.183	0.190	-10.79	-7.65
Compréhension interloc.	0.036	-0.013	-0.006	-137.84	-117.98
Sémantiquement approprié	0.374	0.354	0.363	-5.24	-2.86
Compréhension réplique	0.329	0.293	0.304	-10.81	-7.53
Fluidité	0.354	0.322	0.333	-9.18	-5.94
Cohérence	0.142	0.135	0.137	-4.68	-3.75
Récupération d'erreurs	-0.654	-0.659	-0.665	-0.72	-1.66
Constance	0.560	0.573	0.577	2.38	3.07
Diversité	0.392	0.385	0.390	-1.99	-0.66
Profondeur	0.077	0.096	0.094	23.61	21.23
Sympathie	-0.382	-0.384	-0.390	-0.75	-2.29
Compréhension générale	0.008	-0.023	-0.017	-400.90	-330.39
Flexibilité	-0.099	-0.102	-0.104	-3.19	-5.35
Informatif	-0.245	-0.222	-0.231	9.66	6.03
Curiosité	-0.574	-0.546	-0.560	4.91	2.49
Score individuel moyen	0.579	0.703	0.686	21.47	18.59
Précision émotion moyenne	0.506	0.642	0.625	26.74	23.44
Score combiné moyen	0.579	0.703	0.686	21.47	18.59

TABLEAU B.8 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.486	5.90	3.24
Engagement	-0.625	-0.627	-0.637	-0.20	-1.87
Originalité	-0.328	-0.296	-0.306	9.88	6.90
Pertinence	0.206	0.185	0.190	-10.17	-7.37
Compréhension interloc.	0.036	-0.012	-0.007	-134.27	-119.95
Sémantiquement approprié	0.374	0.355	0.363	-5.07	-2.84
Compréhension réplique	0.329	0.294	0.304	-10.47	-7.60
Fluidité	0.354	0.323	0.333	-8.85	-6.00
Cohérence	0.142	0.135	0.137	-5.02	-3.39
Récupération d'erreurs	-0.654	-0.658	-0.666	-0.57	-1.81
Constance	0.560	0.571	0.575	2.07	2.81
Diversité	0.392	0.386	0.391	-1.56	-0.37
Profondeur	0.077	0.095	0.094	23.39	22.05
Sympathie	-0.382	-0.384	-0.391	-0.65	-2.43
Compréhension générale	0.008	-0.022	-0.018	-385.92	-332.75
Flexibilité	-0.099	-0.101	-0.103	-2.43	-4.62
Informatif	-0.245	-0.223	-0.231	9.18	5.80
Curiosité	-0.574	-0.546	-0.559	4.96	2.65
Score individuel moyen	0.579	0.703	0.690	21.47	19.23
Précision émotion moyenne	0.506	0.638	0.625	26.01	23.44
Score combiné moyen	0.579	0.703	0.690	21.47	19.23

TABLEAU B.9 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.1. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.484	6.31	3.63
Engagement	-0.625	-0.625	-0.636	0.08	-1.70
Originalité	-0.328	-0.294	-0.305	10.39	7.10
Pertinence	0.206	0.183	0.190	-10.91	-7.82
Compréhension interloc.	0.036	-0.013	-0.008	-137.60	-122.34
Sémantiquement approprié	0.374	0.354	0.363	-5.26	-2.81
Compréhension réplique	0.329	0.293	0.303	-10.76	-7.77
Fluidité	0.354	0.322	0.332	-9.12	-6.16
Cohérence	0.142	0.135	0.136	-4.61	-4.20
Récupération d'erreurs	-0.654	-0.659	-0.667	-0.70	-1.92
Constance	0.560	0.573	0.579	2.43	3.38
Diversité	0.392	0.384	0.391	-2.02	-0.36
Profondeur	0.077	0.096	0.095	24.15	22.50
Sympathie	-0.382	-0.384	-0.391	-0.71	-2.52
Compréhension générale	0.008	-0.022	-0.018	-395.90	-334.32
Flexibilité	-0.099	-0.102	-0.105	-3.21	-6.75
Informatif	-0.245	-0.222	-0.231	9.62	6.00
Curiosité	-0.574	-0.546	-0.561	4.87	2.28
Score individuel moyen	0.579	0.699	0.699	20.83	20.83
Précision émotion moyenne	0.506	0.638	0.636	26.01	25.64
Score combiné moyen	0.579	0.699	0.699	20.83	20.83

TABLEAU B.10 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.2			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.483	5.89	3.82
Engagement	-0.625	-0.627	-0.633	-0.30	-1.22
Originalité	-0.328	-0.296	-0.305	9.85	7.22
Pertinence	0.206	0.185	0.190	-10.24	-7.69
Compréhension interloc.	0.036	-0.013	-0.006	-136.11	-115.91
Sémantiquement approprié	0.374	0.355	0.362	-5.10	-3.18
Compréhension réplique	0.329	0.294	0.303	-10.67	-7.86
Fluidité	0.354	0.322	0.332	-8.95	-6.29
Cohérence	0.142	0.135	0.136	-4.63	-4.29
Récupération d'erreurs	-0.654	-0.659	-0.661	-0.71	-1.05
Constance	0.560	0.572	0.573	2.13	2.31
Diversité	0.392	0.387	0.389	-1.44	-0.79
Profondeur	0.077	0.095	0.092	23.56	19.54
Sympathie	-0.382	-0.384	-0.388	-0.63	-1.79
Compréhension générale	0.008	-0.022	-0.016	-387.59	-309.50
Flexibilité	-0.099	-0.102	-0.103	-3.44	-4.68
Informatif	-0.245	-0.223	-0.230	9.24	6.28
Curiosité	-0.574	-0.546	-0.557	4.92	2.96
Score individuel moyen	0.579	0.707	0.701	22.12	21.15
Précision émotion moyenne	0.506	0.642	0.646	26.74	27.47
Score combiné moyen	0.579	0.707	0.701	22.12	21.15

TABLEAU B.11 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.2. (b)

Apport sentimental nul, Variable Interne = 0.2				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.484	6.30	3.67
Engagement	-0.625	-0.625	-0.635	0.06	-1.58
Originalité	-0.328	-0.294	-0.305	10.32	7.11
Pertinence	0.206	0.183	0.190	-10.83	-7.78
Compréhension interloc.	0.036	-0.014	-0.007	-137.95	-120.18
Sémantiquement approprié	0.374	0.354	0.362	-5.33	-3.06
Compréhension réplique	0.329	0.293	0.302	-10.82	-8.02
Fluidité	0.354	0.322	0.332	-9.20	-6.32
Cohérence	0.142	0.135	0.135	-4.78	-4.71
Récupération d'erreurs	-0.654	-0.659	-0.666	-0.74	-1.82
Constance	0.560	0.573	0.578	2.43	3.30
Diversité	0.392	0.385	0.391	-1.91	-0.43
Profondeur	0.077	0.096	0.093	24.04	20.87
Sympathie	-0.382	-0.384	-0.391	-0.69	-2.53
Compréhension générale	0.008	-0.023	-0.018	-401.09	-336.31
Flexibilité	-0.099	-0.102	-0.106	-3.43	-7.18
Informatif	-0.245	-0.222	-0.231	9.62	5.84
Curiosité	-0.574	-0.546	-0.560	4.87	2.38
Score individuel moyen	0.579	0.703	0.699	21.47	20.83
Précision émotion moyenne	0.506	0.642	0.620	26.74	22.34
Score combiné moyen	0.579	0.703	0.699	21.47	20.83

TABLEAU B.12 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (a)

Apport sentimental nul, Variable Interne = 0.3				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.486	5.85	3.34
Engagement	-0.625	-0.629	-0.638	-0.58	-1.98
Originalité	-0.328	-0.296	-0.305	9.92	6.92
Pertinence	0.206	0.184	0.190	-10.40	-7.46
Compréhension interloc.	0.036	-0.014	-0.008	-140.08	-122.73
Sémantiquement approprié	0.374	0.356	0.363	-4.89	-2.83
Compréhension réplique	0.329	0.294	0.303	-10.68	-7.92
Fluidité	0.354	0.322	0.333	-8.97	-6.11
Cohérence	0.142	0.136	0.137	-3.92	-3.25
Récupération d'erreurs	-0.654	-0.662	-0.667	-1.12	-1.99
Constance	0.560	0.575	0.578	2.75	3.31
Diversité	0.392	0.386	0.392	-1.75	-0.07
Profondeur	0.077	0.096	0.095	24.82	22.46
Sympathie	-0.382	-0.385	-0.391	-0.98	-2.43
Compréhension générale	0.008	-0.023	-0.018	-409.85	-341.33
Flexibilité	-0.099	-0.104	-0.107	-5.31	-8.92
Informatif	-0.245	-0.222	-0.231	9.39	5.77
Curiosité	-0.574	-0.547	-0.560	4.69	2.39
Score individuel moyen	0.579	0.666	0.651	15.06	12.50
Précision émotion moyenne	0.506	0.594	0.594	17.22	17.22
Score combiné moyen	0.579	0.666	0.651	15.06	12.50

TABLEAU B.13 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.3. (b)

Apport sentimental nul, Variable Interne = 0.3				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.472	-0.484	6.10	3.72
Engagement	-0.625	-0.628	-0.636	-0.39	-1.75
Originalité	-0.328	-0.294	-0.305	10.39	7.14
Pertinence	0.206	0.184	0.189	-10.60	-8.22
Compréhension interloc.	0.036	-0.015	-0.008	-141.36	-123.77
Sémantiquement approprié	0.374	0.354	0.363	-5.18	-3.02
Compréhension réplique	0.329	0.292	0.302	-11.12	-8.27
Fluidité	0.354	0.321	0.331	-9.31	-6.42
Cohérence	0.142	0.136	0.136	-4.12	-4.02
Récupération d'erreurs	-0.654	-0.659	-0.664	-0.78	-1.53
Constance	0.560	0.573	0.576	2.31	2.95
Diversité	0.392	0.384	0.389	-2.02	-0.81
Profondeur	0.077	0.095	0.093	22.86	20.85
Sympathie	-0.382	-0.386	-0.391	-1.21	-2.59
Compréhension générale	0.008	-0.023	-0.019	-406.80	-351.55
Flexibilité	-0.099	-0.103	-0.106	-4.55	-7.68
Informatif	-0.245	-0.222	-0.230	9.52	6.24
Curiosité	-0.574	-0.547	-0.558	4.75	2.79
Score individuel moyen	0.579	0.675	0.657	16.67	13.46
Précision émotion moyenne	0.506	0.612	0.605	20.88	19.41
Score combiné moyen	0.579	0.675	0.657	16.67	13.46

TABLEAU B.14 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (a)

Apport sentimental nul, Variable Interne = 0.4				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.79	3.52
Engagement	-0.625	-0.628	-0.637	-0.50	-1.87
Originalité	-0.328	-0.295	-0.306	9.98	6.83
Pertinence	0.206	0.183	0.189	-10.92	-8.10
Compréhension interloc.	0.036	-0.015	-0.009	-141.43	-124.90
Sémantiquement approprié	0.374	0.354	0.362	-5.22	-3.15
Compréhension réplique	0.329	0.292	0.302	-11.06	-8.19
Fluidité	0.354	0.321	0.332	-9.26	-6.38
Cohérence	0.142	0.136	0.135	-4.19	-4.71
Récupération d'erreurs	-0.654	-0.661	-0.665	-0.98	-1.68
Constance	0.560	0.573	0.576	2.42	2.98
Diversité	0.392	0.386	0.390	-1.71	-0.49
Profondeur	0.077	0.096	0.093	23.73	19.88
Sympathie	-0.382	-0.385	-0.391	-1.03	-2.50
Compréhension générale	0.008	-0.024	-0.020	-413.44	-359.19
Flexibilité	-0.099	-0.104	-0.107	-4.98	-8.08
Informatif	-0.245	-0.222	-0.230	9.32	6.24
Curiosité	-0.574	-0.546	-0.559	4.87	2.65
Score individuel moyen	0.579	0.662	0.640	14.42	10.58
Précision émotion moyenne	0.506	0.594	0.566	17.22	11.72
Score combiné moyen	0.579	0.662	0.640	14.42	10.58

TABLEAU B.15 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.4. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.4			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.485	6.26	3.47
Engagement	-0.625	-0.627	-0.637	-0.24	-1.89
Originalité	-0.328	-0.294	-0.305	10.45	7.04
Pertinence	0.206	0.184	0.190	-10.57	-7.48
Compréhension interloc.	0.036	-0.014	-0.008	-139.86	-121.63
Sémantiquement approprié	0.374	0.355	0.365	-5.09	-2.48
Compréhension réplique	0.329	0.293	0.304	-11.02	-7.67
Fluidité	0.354	0.322	0.334	-9.16	-5.78
Cohérence	0.142	0.136	0.138	-3.97	-2.77
Récupération d'erreurs	-0.654	-0.659	-0.666	-0.78	-1.74
Constance	0.560	0.572	0.576	2.26	3.01
Diversité	0.392	0.385	0.390	-1.93	-0.51
Profondeur	0.077	0.095	0.094	22.66	21.09
Sympathie	-0.382	-0.386	-0.392	-1.08	-2.65
Compréhension générale	0.008	-0.023	-0.017	-403.20	-331.37
Flexibilité	-0.099	-0.103	-0.105	-4.47	-6.86
Informatif	-0.245	-0.221	-0.230	9.94	6.17
Curiosité	-0.574	-0.546	-0.561	4.80	2.24
Score individuel moyen	0.579	0.685	0.672	18.27	16.03
Précision émotion moyenne	0.506	0.612	0.614	20.88	21.25
Score combiné moyen	0.579	0.685	0.672	18.27	16.03

TABLEAU B.16 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.5			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.484	5.81	3.71
Engagement	-0.625	-0.628	-0.633	-0.44	-1.29
Originalité	-0.328	-0.296	-0.306	9.95	6.78
Pertinence	0.206	0.184	0.190	-10.40	-7.50
Compréhension interloc.	0.036	-0.014	-0.007	-139.80	-119.35
Sémantiquement approprié	0.374	0.354	0.362	-5.20	-3.13
Compréhension réplique	0.329	0.293	0.303	-10.84	-7.86
Fluidité	0.354	0.322	0.332	-9.09	-6.24
Cohérence	0.142	0.136	0.137	-4.25	-3.61
Récupération d'erreurs	-0.654	-0.661	-0.665	-1.02	-1.58
Constance	0.560	0.573	0.574	2.36	2.59
Diversité	0.392	0.387	0.390	-1.41	-0.51
Profondeur	0.077	0.096	0.094	23.91	21.36
Sympathie	-0.382	-0.386	-0.390	-1.09	-2.16
Compréhension générale	0.008	-0.023	-0.018	-405.55	-333.86
Flexibilité	-0.099	-0.103	-0.105	-4.79	-6.46
Informatif	-0.245	-0.222	-0.230	9.59	6.17
Curiosité	-0.574	-0.546	-0.559	4.82	2.68
Score individuel moyen	0.579	0.659	0.657	13.78	13.46
Précision émotion moyenne	0.506	0.592	0.583	16.85	15.02
Score combiné moyen	0.579	0.659	0.657	13.78	13.46

TABLEAU B.17 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (b)

Apport sentimental nul, Variable Interne = 0.5				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.91	3.55
Engagement	-0.625	-0.627	-0.635	-0.23	-1.53
Originalité	-0.328	-0.295	-0.306	10.26	6.78
Pertinence	0.206	0.184	0.190	-10.60	-7.67
Compréhension interloc.	0.036	-0.014	-0.007	-139.10	-120.40
Sémantiquement approprié	0.374	0.355	0.362	-4.96	-3.10
Compréhension réplique	0.329	0.293	0.302	-10.77	-8.07
Fluidité	0.354	0.322	0.331	-9.20	-6.56
Cohérence	0.142	0.137	0.137	-3.28	-3.22
Récupération d'erreurs	-0.654	-0.660	-0.664	-0.84	-1.51
Constance	0.560	0.572	0.576	2.27	2.89
Diversité	0.392	0.386	0.389	-1.75	-0.80
Profondeur	0.077	0.096	0.093	23.99	19.79
Sympathie	-0.382	-0.385	-0.391	-0.91	-2.40
Compréhension générale	0.008	-0.023	-0.018	-403.01	-343.92
Flexibilité	-0.099	-0.103	-0.106	-4.41	-7.22
Informatif	-0.245	-0.222	-0.230	9.43	6.09
Curiosité	-0.574	-0.546	-0.558	4.94	2.77
Score individuel moyen	0.579	0.651	0.666	12.50	15.06
Précision émotion moyenne	0.506	0.592	0.588	16.85	16.12
Score combiné moyen	0.579	0.651	0.666	12.50	15.06

TABLEAU B.18 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.5. (c)

Apport sentimental nul, Variable Interne = 0.5				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.482	6.35	4.08
Engagement	-0.625	-0.626	-0.634	-0.16	-1.37
Originalité	-0.328	-0.293	-0.303	10.78	7.65
Pertinence	0.206	0.183	0.189	-10.87	-8.11
Compréhension interloc.	0.036	-0.015	-0.008	-143.28	-122.52
Sémantiquement approprié	0.374	0.355	0.362	-5.14	-3.18
Compréhension réplique	0.329	0.292	0.302	-11.20	-8.22
Fluidité	0.354	0.321	0.331	-9.27	-6.41
Cohérence	0.142	0.136	0.136	-4.29	-4.22
Récupération d'erreurs	-0.654	-0.660	-0.664	-0.87	-1.58
Constance	0.560	0.572	0.574	2.26	2.59
Diversité	0.392	0.384	0.388	-2.19	-1.17
Profondeur	0.077	0.094	0.092	22.11	19.41
Sympathie	-0.382	-0.386	-0.390	-1.05	-2.25
Compréhension générale	0.008	-0.024	-0.019	-412.03	-348.81
Flexibilité	-0.099	-0.103	-0.105	-4.54	-6.49
Informatif	-0.245	-0.221	-0.229	9.82	6.77
Curiosité	-0.574	-0.547	-0.557	4.76	2.88
Score individuel moyen	0.579	0.647	0.655	11.86	13.14
Précision émotion moyenne	0.506	0.588	0.575	16.12	13.55
Score combiné moyen	0.579	0.647	0.655	11.86	13.14

TABLEAU B.19 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (a)

Apport sentimental nul, Variable Interne = 0.6				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.92	3.44
Engagement	-0.625	-0.628	-0.635	-0.37	-1.59
Originalité	-0.328	-0.296	-0.306	9.95	6.91
Pertinence	0.206	0.184	0.190	-10.51	-7.71
Compréhension interloc.	0.036	-0.015	-0.007	-140.79	-119.85
Sémantiquement approprié	0.374	0.355	0.363	-5.10	-2.99
Compréhension réplique	0.329	0.293	0.303	-10.91	-7.86
Fluidité	0.354	0.322	0.332	-9.21	-6.33
Cohérence	0.142	0.137	0.136	-3.79	-3.89
Récupération d'erreurs	-0.654	-0.660	-0.665	-0.87	-1.60
Constance	0.560	0.573	0.576	2.39	2.86
Diversité	0.392	0.386	0.389	-1.76	-0.91
Profondeur	0.077	0.096	0.093	23.77	20.00
Sympathie	-0.382	-0.385	-0.390	-1.00	-2.25
Compréhension générale	0.008	-0.024	-0.019	-413.57	-347.01
Flexibilité	-0.099	-0.104	-0.107	-5.16	-7.97
Informatif	-0.245	-0.222	-0.229	9.63	6.67
Curiosité	-0.574	-0.546	-0.558	4.95	2.78
Score individuel moyen	0.579	0.666	0.655	15.06	13.14
Précision émotion moyenne	0.506	0.596	0.586	17.58	15.75
Score combiné moyen	0.579	0.666	0.655	15.06	13.14

TABLEAU B.20 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.6. (b)

Apport sentimental nul, Variable Interne = 0.6				Différence (%)	
Caractéristiques:	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.482	6.26	4.01
Engagement	-0.625	-0.628	-0.635	-0.39	-1.50
Originalité	-0.328	-0.293	-0.303	10.58	7.67
Pertinence	0.206	0.183	0.189	-10.85	-8.12
Compréhension interloc.	0.036	-0.016	-0.008	-143.59	-123.09
Sémantiquement approprié	0.374	0.354	0.362	-5.40	-3.21
Compréhension réplique	0.329	0.291	0.301	-11.40	-8.30
Fluidité	0.354	0.321	0.331	-9.47	-6.55
Cohérence	0.142	0.135	0.136	-4.61	-3.88
Récupération d'erreurs	-0.654	-0.660	-0.665	-0.90	-1.61
Constance	0.560	0.573	0.576	2.41	2.90
Diversité	0.392	0.385	0.390	-1.82	-0.73
Profondeur	0.077	0.094	0.093	22.05	19.89
Sympathie	-0.382	-0.386	-0.390	-1.17	-2.29
Compréhension générale	0.008	-0.024	-0.018	-415.59	-340.62
Flexibilité	-0.099	-0.103	-0.105	-4.66	-6.31
Informatif	-0.245	-0.221	-0.229	9.77	6.86
Curiosité	-0.574	-0.547	-0.558	4.77	2.84
Score individuel moyen	0.579	0.649	0.660	12.18	14.10
Précision émotion moyenne	0.506	0.583	0.588	15.02	16.12
Score combiné moyen	0.579	0.649	0.660	12.18	14.10

TABLEAU B.21 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.7			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.92	3.39
Engagement	-0.625	-0.627	-0.637	-0.27	-1.81
Originalité	-0.328	-0.296	-0.305	9.91	7.05
Pertinence	0.206	0.183	0.188	-10.74	-8.32
Compréhension interloc.	0.036	-0.015	-0.009	-140.96	-126.21
Sémantiquement approprié	0.374	0.354	0.363	-5.38	-2.99
Compréhension réplique	0.329	0.292	0.302	-11.13	-8.27
Fluidité	0.354	0.321	0.331	-9.40	-6.53
Cohérence	0.142	0.136	0.136	-4.20	-4.02
Récupération d'erreurs	-0.654	-0.660	-0.665	-0.84	-1.73
Constance	0.560	0.571	0.577	2.10	3.15
Diversité	0.392	0.384	0.389	-2.05	-0.76
Profondeur	0.077	0.095	0.093	23.30	19.91
Sympathie	-0.382	-0.385	-0.393	-0.95	-2.89
Compréhension générale	0.008	-0.024	-0.019	-416.86	-355.05
Flexibilité	-0.099	-0.104	-0.107	-5.31	-8.79
Informatif	-0.245	-0.222	-0.230	9.63	6.39
Curiosité	-0.574	-0.545	-0.559	5.02	2.55
Score individuel moyen	0.579	0.666	0.651	15.06	12.50
Précision émotion moyenne	0.506	0.586	0.586	15.75	15.75
Score combiné moyen	0.579	0.666	0.651	15.06	12.50

TABLEAU B.22 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.7. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.7			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.470	-0.483	6.38	3.86
Engagement	-0.625	-0.627	-0.637	-0.22	-1.79
Originalité	-0.328	-0.293	-0.303	10.69	7.54
Pertinence	0.206	0.184	0.189	-10.64	-7.87
Compréhension interloc.	0.036	-0.014	-0.008	-140.11	-121.91
Sémantiquement approprié	0.374	0.355	0.363	-5.07	-2.91
Compréhension réplique	0.329	0.292	0.302	-11.08	-8.18
Fluidité	0.354	0.322	0.332	-9.19	-6.29
Cohérence	0.142	0.137	0.138	-3.60	-2.94
Récupération d'erreurs	-0.654	-0.659	-0.665	-0.73	-1.64
Constance	0.560	0.572	0.576	2.20	2.90
Diversité	0.392	0.384	0.390	-2.17	-0.71
Profondeur	0.077	0.095	0.093	22.34	20.15
Sympathie	-0.382	-0.385	-0.391	-0.80	-2.43
Compréhension générale	0.008	-0.023	-0.018	-407.49	-343.33
Flexibilité	-0.099	-0.103	-0.106	-4.55	-7.49
Informatif	-0.245	-0.220	-0.228	10.41	7.04
Curiosité	-0.574	-0.546	-0.559	4.96	2.69
Score individuel moyen	0.579	0.660	0.660	14.10	14.10
Précision émotion moyenne	0.506	0.597	0.586	17.95	15.75
Score combiné moyen	0.579	0.660	0.660	14.10	14.10

TABLEAU B.23 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.8			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.472	-0.485	6.08	3.44
Engagement	-0.625	-0.627	-0.636	-0.31	-1.76
Originalité	-0.328	-0.294	-0.306	10.34	6.89
Pertinence	0.206	0.184	0.190	-10.66	-7.77
Compréhension interloc.	0.036	-0.014	-0.007	-139.99	-120.92
Sémantiquement approprié	0.374	0.355	0.363	-5.02	-2.97
Compréhension réplique	0.329	0.293	0.302	-10.91	-8.03
Fluidité	0.354	0.322	0.332	-9.13	-6.30
Cohérence	0.142	0.136	0.137	-4.30	-3.43
Récupération d'erreurs	-0.654	-0.659	-0.664	-0.68	-1.55
Constance	0.560	0.573	0.576	2.30	2.93
Diversité	0.392	0.385	0.390	-2.01	-0.54
Profondeur	0.077	0.095	0.093	22.77	20.87
Sympathie	-0.382	-0.385	-0.391	-1.02	-2.46
Compréhension générale	0.008	-0.023	-0.017	-402.44	-330.48
Flexibilité	-0.099	-0.103	-0.106	-4.65	-7.74
Informatif	-0.245	-0.221	-0.230	9.79	6.27
Curiosité	-0.574	-0.545	-0.559	5.02	2.65
Score individuel moyen	0.579	0.651	0.649	12.50	12.18
Précision émotion moyenne	0.506	0.590	0.592	16.48	16.85
Score combiné moyen	0.579	0.651	0.649	12.50	12.18

TABLEAU B.24 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.8. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.8			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.481	6.37	4.30
Engagement	-0.625	-0.628	-0.633	-0.39	-1.22
Originalité	-0.328	-0.294	-0.303	10.55	7.71
Pertinence	0.206	0.183	0.189	-11.03	-8.21
Compréhension interloc.	0.036	-0.015	-0.007	-142.71	-120.97
Sémantiquement approprié	0.374	0.354	0.361	-5.35	-3.52
Compréhension réplique	0.329	0.291	0.301	-11.44	-8.59
Fluidité	0.354	0.320	0.330	-9.54	-6.79
Cohérence	0.142	0.136	0.137	-4.20	-3.30
Récupération d'erreurs	-0.654	-0.659	-0.661	-0.73	-1.02
Constance	0.560	0.572	0.571	2.16	2.11
Diversité	0.392	0.385	0.388	-1.91	-1.05
Profondeur	0.077	0.094	0.091	21.08	17.86
Sympathie	-0.382	-0.385	-0.388	-0.93	-1.82
Compréhension générale	0.008	-0.023	-0.018	-403.26	-338.31
Flexibilité	-0.099	-0.103	-0.104	-4.33	-5.89
Informatif	-0.245	-0.221	-0.228	9.88	7.08
Curiosité	-0.574	-0.546	-0.556	4.87	3.10
Score individuel moyen	0.579	0.664	0.662	14.74	14.42
Précision émotion moyenne	0.506	0.579	0.583	14.29	15.02
Score combiné moyen	0.579	0.664	0.662	14.74	14.42

TABLEAU B.25 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.9			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.485	5.91	3.39
Engagement	-0.625	-0.627	-0.636	-0.22	-1.74
Originalité	-0.328	-0.296	-0.306	9.87	6.62
Pertinence	0.206	0.184	0.190	-10.46	-7.78
Compréhension interloc.	0.036	-0.014	-0.007	-138.84	-119.51
Sémantiquement approprié	0.374	0.354	0.363	-5.30	-2.99
Compréhension réplique	0.329	0.292	0.303	-11.12	-7.97
Fluidité	0.354	0.321	0.332	-9.50	-6.29
Cohérence	0.142	0.137	0.138	-3.64	-2.48
Récupération d'erreurs	-0.654	-0.659	-0.666	-0.73	-1.74
Constance	0.560	0.571	0.576	2.11	2.99
Diversité	0.392	0.385	0.390	-1.89	-0.72
Profondeur	0.077	0.096	0.094	24.08	22.23
Sympathie	-0.382	-0.384	-0.391	-0.71	-2.37
Compréhension générale	0.008	-0.023	-0.018	-400.23	-334.58
Flexibilité	-0.099	-0.103	-0.106	-4.87	-7.46
Informatif	-0.245	-0.221	-0.230	9.77	6.45
Curiosité	-0.574	-0.545	-0.558	5.04	2.72
Score individuel moyen	0.579	0.675	0.659	16.67	13.78
Précision émotion moyenne	0.506	0.605	0.596	19.41	17.58
Score combiné moyen	0.579	0.675	0.659	16.67	13.78

TABLEAU B.26 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 0.9. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 0.9			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.470	-0.482	6.41	4.03
Engagement	-0.625	-0.627	-0.635	-0.19	-1.56
Originalité	-0.328	-0.293	-0.304	10.78	7.51
Pertinence	0.206	0.183	0.189	-10.76	-7.83
Compréhension interloc.	0.036	-0.015	-0.009	-142.49	-126.40
Sémantiquement approprié	0.374	0.353	0.362	-5.58	-3.16
Compréhension réplique	0.329	0.291	0.302	-11.37	-8.24
Fluidité	0.354	0.320	0.331	-9.68	-6.59
Cohérence	0.142	0.135	0.136	-4.58	-4.09
Récupération d'erreurs	-0.654	-0.658	-0.665	-0.54	-1.62
Constance	0.560	0.571	0.576	2.03	3.01
Diversité	0.392	0.384	0.390	-2.16	-0.66
Profondeur	0.077	0.095	0.093	22.32	20.68
Sympathie	-0.382	-0.385	-0.391	-0.87	-2.53
Compréhension générale	0.008	-0.023	-0.018	-401.61	-344.92
Flexibilité	-0.099	-0.103	-0.105	-3.88	-6.47
Informatif	-0.245	-0.221	-0.228	10.00	6.98
Curiosité	-0.574	-0.545	-0.558	5.08	2.73
Score individuel moyen	0.579	0.672	0.646	16.03	11.54
Précision émotion moyenne	0.506	0.609	0.571	20.15	12.82
Score combiné moyen	0.579	0.672	0.646	16.03	11.54

TABLEAU B.27 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (a)

Caractéristiques:	Apport sentimental nul, Variable Interne = 1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.473	-0.486	5.82	3.31
Engagement	-0.625	-0.626	-0.636	-0.17	-1.75
Originalité	-0.328	-0.296	-0.306	9.82	6.89
Pertinence	0.206	0.185	0.191	-10.08	-7.27
Compréhension interloc.	0.036	-0.014	-0.009	-138.06	-124.58
Sémantiquement approprié	0.374	0.356	0.364	-4.84	-2.58
Compréhension réplique	0.329	0.294	0.304	-10.46	-7.51
Fluidité	0.354	0.323	0.333	-8.93	-6.02
Cohérence	0.142	0.138	0.138	-3.00	-2.77
Récupération d'erreurs	-0.654	-0.659	-0.665	-0.79	-1.72
Constance	0.560	0.573	0.578	2.34	3.34
Diversité	0.392	0.385	0.390	-1.82	-0.58
Profondeur	0.077	0.097	0.096	24.95	23.79
Sympathie	-0.382	-0.384	-0.391	-0.67	-2.59
Compréhension générale	0.008	-0.022	-0.018	-393.03	-334.80
Flexibilité	-0.099	-0.102	-0.106	-3.69	-7.45
Informatif	-0.245	-0.223	-0.229	9.25	6.48
Curiosité	-0.574	-0.546	-0.559	4.95	2.58
Score individuel moyen	0.579	0.672	0.655	16.03	13.14
Précision émotion moyenne	0.506	0.607	0.607	19.78	19.78
Score combiné moyen	0.579	0.672	0.655	16.03	13.14

TABLEAU B.28 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (b)

Caractéristiques:	Apport sentimental nul, Variable Interne = 1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.472	-0.485	6.02	3.56
Engagement	-0.625	-0.626	-0.635	-0.13	-1.52
Originalité	-0.328	-0.296	-0.306	9.90	6.88
Pertinence	0.206	0.183	0.189	-10.95	-7.96
Compréhension interloc.	0.036	-0.014	-0.007	-140.27	-120.01
Sémantiquement approprié	0.374	0.354	0.363	-5.25	-2.78
Compréhension réplique	0.329	0.292	0.304	-11.07	-7.67
Fluidité	0.354	0.321	0.333	-9.35	-6.12
Cohérence	0.142	0.137	0.138	-3.69	-2.51
Récupération d'erreurs	-0.654	-0.660	-0.665	-0.92	-1.61
Constance	0.560	0.572	0.576	2.19	2.87
Diversité	0.392	0.385	0.389	-1.96	-0.79
Profondeur	0.077	0.096	0.095	23.92	22.50
Sympathie	-0.382	-0.385	-0.390	-0.90	-2.21
Compréhension générale	0.008	-0.024	-0.018	-423.20	-341.58
Flexibilité	-0.099	-0.103	-0.105	-4.20	-6.39
Informatif	-0.245	-0.222	-0.229	9.47	6.48
Curiosité	-0.574	-0.545	-0.558	5.01	2.75
Score individuel moyen	0.579	0.644	0.649	11.22	12.18
Précision émotion moyenne	0.506	0.584	0.594	15.38	17.22
Score combiné moyen	0.579	0.644	0.649	11.22	12.18

TABLEAU B.29 Résultats obtenus en forçant la valeur de la variable interne émotionnelle égale à 1. (c)

Caractéristiques:	Apport sentimental nul, Variable Interne = 1			Différence (%)	
	Scores originaux	Scores système	Scores système corrigé	Système	Système corrigé
Intérêt	-0.503	-0.471	-0.482	6.32	4.05
Engagement	-0.625	-0.627	-0.633	-0.25	-1.21
Originalité	-0.328	-0.293	-0.304	10.81	7.41
Pertinence	0.206	0.183	0.189	-10.82	-7.90
Compréhension interloc.	0.036	-0.015	-0.007	-141.52	-119.04
Sémantiquement approprié	0.374	0.354	0.361	-5.41	-3.31
Compréhension réplique	0.329	0.292	0.302	-11.17	-8.26
Fluidité	0.354	0.321	0.331	-9.29	-6.47
Cohérence	0.142	0.135	0.137	-4.62	-3.32
Récupération d'erreurs	-0.654	-0.658	-0.663	-0.63	-1.30
Constance	0.560	0.572	0.574	2.14	2.50
Diversité	0.392	0.384	0.389	-2.06	-0.89
Profondeur	0.077	0.094	0.092	22.07	19.03
Sympathie	-0.382	-0.386	-0.390	-1.30	-2.18
Compréhension générale	0.008	-0.023	-0.017	-403.95	-330.64
Flexibilité	-0.099	-0.103	-0.106	-3.95	-7.38
Informatif	-0.245	-0.222	-0.229	9.57	6.83
Curiosité	-0.574	-0.546	-0.558	4.83	2.80
Score individuel moyen	0.579	0.646	0.651	11.54	12.50
Précision émotion moyenne	0.506	0.592	0.579	16.85	14.29
Score combiné moyen	0.579	0.646	0.651	11.54	12.50

ANNEXE C PSEUDO-CODE DE LA BOUCLE PRINCIPALE DE PHILIA

Algorithm 5 Pseudocode de Philia

```

Initialisation :
nlu = NLU()
chatbots = Chatbots()
model = Model()
scores_factory = ScoresFactory()
// Collecte des informations des bases de données :
context = scores_factory.get_context()
emotion = scores_factory.get_emotion()
scores = scores_factory.get_scores(emotion =
emotion, context = context)
Début de la boucle :
previous_input = 'Hi'
while True do
    // Initialisation du dialogue par l'utilisateur :
    user_input = input()
    if user_input == 'out' then
        break
    end if
    // Nettoyage de la réplique usager et décomposition de la réplique système précédente :
    user_input = nlu.clean(input = user_input)
    previous_input_decomposition =
nlu.decompose(input = previous_input)
    // Vérification croisée des tuples avec notre base de données :
    scores.check(input = previous_input_decomposition)
    // Calcul des scores des répliques :
    user_score = scores.scorer(input = user_input)
    system_score = scores.scorer(input = previous_input)
    combined_score =
scores.combined_scorer(system_score =
system_score, user_score = user_score)
    // Mise à jour des scores des tuples et de la variable interne :
    scores.update(input = previous_input_decomposition, score =
combined_score)
    scores.iv_update(score = combined_score)
    // répliques des agents conversationnels :
    replies = chatbot.replies(input = user_input)
    system_answer = model.solve(replies =
replies, scores = scores)
    previous_input = system_answer
    corrected_answer = nlu.correct(input =
system_answer)
    output(corrected_answer)
end while

```
