# Recommandation automatique et adaptative d'emojis

Gaël Guibon





Université Aix-Marseille

École Doctorale 184

Faculté des Sciences et Techniques

LIS UMR CNRS 7020 - R2I
Caléa Solutions


Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Mathématique et informatique
Spécialité : Informatique


Gaël GUIBON


# Recommandation automatique et adaptative d'emojis.

Soutenue le 24/05/2019 devant le jury :

| | | |
|---|---|---|
| Chloé CLAVEL | MCF HDR, LTCI Telecom-ParisTech | Rapportrice |
| Horacio SAGGION | Associate Professor, Universitat Pompeu Fabra | Rapporteur |
| Béatrice DAILLE | PR, LS2N, Université de Nantes | Examinatrice |
| Frédéric LANDRAGIN | DR, CNRS, LATTICE | Examinateur |
| Pierre LISON | PhD, Senior Researcher, Norsk Regnesentral | Examinateur |
| Magalie OCHS | MCF HDR, Université Aix-Marseille | Co-directrice de thèse |
| Patrice BELLOT | PR, Université Aix-Marseille | Directeur de thèse |

v0.0.1 20180907
Ce travail s'inscrit dans le cadre d'une thèse CIFRE de l'Association Nationale Recherche et Technologie (ANRT [1]) Commentaires, corrections, et autres remarques sont les bienvenus à :
gael.guibon@lis-lab.fr
Laboratoire des Sciences de L'Information et des Systèmes, LIS UMR 7020, CNRS
Université Aix-Marseille,
Batiment Polytech,
Avenue Escadrille Normandie-Niemen,
13397 MARSEILLE CEDEX 20

1. http://www.anrt.asso.fr/fr/cifre-7843

# Remerciements

Tout d'abord, je tiens à remercier les membres du jury, Chloé Clavel et Horacio Saggion pour avoir accepté d'être rapporteurs de ce manuscrit, ainsi que Béatrice Daille, Frédéric Landragin et Pierre Lison pour leur participation à ce jury en tant qu'examinateurs.

Un grand merci à Patrice Bellot et Magalie Ochs pour l'encadrement tout au long de cette thèse et tous les conseils avisés et les relectures qui ont contribué à améliorer cette thèse.

Je remercie également Saïd Hadjiat et Olivier Lhermite, président et directeur général de Caléa Solutions, pour leur confiance et leur sérénité malgré les difficultés rencontrées. Je remercie la *Mood Team* au complet pour ces moments de discussion au support, et particulièrement l'ensemble des "marseillais" Gaëtan, Florian, Roland, Maxime, Marie, ainsi que Marc, Didier et Julien pour ces moments enrichissants.

Merci aux membres actuels et anciens du laboratoire LIS, doctorants, enseignants, chercheurs ou membres du personnel avec qui j'ai souvent échangé. Merci à Aref et Feda pour toutes ces discussions et tous ces échanges bénéfiques. Merci à Adrian et Sébastien pour leurs conseils précieux qui m'ont aidé à bien démarrer. À tous ceux que je n'ai pas cité, merci à tous pour tous ces moments de rigolade, de soutien moral ou café d'équipe.

Pour son attention, son aide et ses relectures très instructives, je tiens à remercier Laurence d'Alifée de son soutien *de facto* tout au long de la thèse, et aussi tous les membres de l'*english workshop*.

Merci à Sophie, Kim et Matthieu pour m'avoir mis le pied à l'étrier et pour avoir accompagné mes premiers pas dans le monde de la recherche.

Merci à mes parents et à mon frère pour tout le soutien apporté depuis toujours et sans qui rien ne serait arrivé.

* * *

Je tiens à rendre hommage à Isabelle Tellier qui a cru en moi, m'a soutenu et m'a instillé la passion et l'étincelle de la recherche.

# Résumé

Depuis leur apparition en 1999, les emojis ont une popularité grandissante dans les systèmes de communication. Ces petites images pouvant représenter une idée, un concept ou une émotion, se retrouvent disponibles aux utilisateurs dans de nombreux contextes logiciels : messagerie instantanée, courriel, forums et autres réseaux sociaux en tout genre. Leur usage, en hausse constante, a entraîné l'apparition récurrente de nouveaux emojis, allant jusqu'à 2 789 emojis standardisés en fin d'année 2018.

Le parcours de bibliothèques d'emojis ou l'utilisation de moteur de recherche intégré n'est plus suffisant pour permettre à l'utilisateur de maximiser leur utilisation ; une recommandation d'emojis adaptée est nécessaire. Pour cela, nous présentons nos travaux de recherche axés sur le thème de la recommandation d'emojis. Ces travaux ont pour objectif de créer un système de recommandation automatique d'emojis adapté à un contexte conversationnel informel et privé. Ce système doit améliorer l'expérience utilisateur et la qualité de la communication, en plus de pouvoir prendre en compte d'éventuels nouveaux emojis.

Dans le cadre de cette thèse, nous contribuons tout d'abord en montrant les limites d'usage réel d'une prédiction d'emojis ainsi que la nécessité de prédire des notions plus générales. Nous vérifions également si l'usage réel des emojis représentant une expression faciale d'émotion correspond à l'existant théorique. Enfin, nous abordons les pistes d'évaluation d'un tel système par l'insuffisance des métriques, et l'importance d'une interface utilisateur dédiée.

Pour ce faire, nous utilisons une approche orientée apprentissage automatique à la fois supervisée et non supervisée, ainsi que la conception de modèles de langues ou, plus précisément, de plongements lexicaux. Plusieurs composantes de ce travail ont donné lieu à des publications nationales et internationales, dont le prix du meilleur logiciel reproductible à la conférence CICLing 2018, ainsi que celui du meilleur poster pour la catégorie des médias sociaux SONAMA à la conférence SAC 2018.

# Abstract

The first emojis were created in 1999. Since then, their propularity constantly raised in communication systems. Being images representing either an idea, a concept, or an emotion, emojis are available to the users in multiple software contexts : instant messaging, emails, forums, and other types of social medias. Their usage grew constantly and, associated to the constant addition of new emojis, there are now more than 2,789 standard emojis since winter 2018.

To access a specific emoji, scrolling through huge emoji librairies or using a emoji search engines is not enough to maximize their usage and their diversity. An emoji recommendation system is required. To answer this need, we present our research work focused on the emoji recommendation topic. The objectives are to create an emoji recommender system adapted to a private and informal conversationnal context. This system must enhance the user experience, the communication quality, and take into account possible new emerging emojis.

Our first contribution is to show the limits of a emoji prediction for the real usage case, and to demonstrate the need of a more global recommandation. We also verify the correlation between the real usage of emojis representing facial expressions and a related theory on facial expressions. We also tackle the evaluation part of this system, with the metrics' limits and the importance of a dedicated user interface.

The approach is based on supervised and unsupervised machine learning, associated to language models such as word embeddings. Several parts of this work were published in national and international conferences, including the best software award and best poster award for the social media track.

# Publications

*Articles de conférences internationales*

1. (Best Poster for SONAMA (social media) track Award) : Gaël Guibon, Magalie Ochs, Patrice Bellot. Emoji Recommendation in Private Instant Messages. SAC '18 Proceedings of the 33rd Annual ACM Symposium on Applied Computing, Apr 2018, Pau, France. ACM Press, 2018, SAC '18 Proceedings of the 33rd Annual ACM Symposium on Applied Computing.
2. (Best software and reproductibilty Award) : Gaël Guibon, Magalie Ochs, Patrice Bellot. From Emoji Usage to Categorical Emoji Prediction. 19th International Conference on Computational Linguistics and Intelligent Text Processing (CICLING 2018), Mar 2018, Hanoï, Vietnam.
3. Gaël Guibon, Magalie Ochs, Patrice Bellot. LIS at SemEval-2018 Task 2 : Mixing Word Embeddings and Bag of Features for Multilingual Emoji Prediction. SemEVAL, Jun 2018, New-Orleans, United States.

*Articles de conférences nationales*

1. Gaël Guibon, Magalie Ochs, Patrice Bellot. De l'usage réel des emojis à une prédiction de leurs catégories. In Actes de la conférence Traitement Automatique de la Langue Naturelle, TALN 2018 (p. 539)
2. Gaël Guibon, Magalie Ochs, Patrice Bellot. Une plateforme de recommandation automatique d'emojis. Traitement Automatique de la Langue Naturelle, Jun 2017, Orléans, France
3. Gaël Guibon, Magalie Ochs, Patrice Bellot. Prédiction automatique d'emojis sentimentaux. COnférence en Recherche d'Information et Applications (CORIA), Mar 2017, Marseille, France.
4. Gaël Guibon, Magalie Ochs, Patrice Bellot. From Emojis to Sentiment Analysis. WACAI 2016, Jun 2016, Brest, France. WACAI 2016, 2016

# Table des matières

**Remerciements** 7

**Résumé** 9

**Abstract** 11

**Publications** 13

**Liste des figures** 19

**Liste des algorithmes** 21

**Liste des tableaux** 21

**Liste des abréviations** 25

**Introduction générale** 25

**1 Contexte théorique** 31
- 1.1 Introduction 31
- 1.2 Les fonctions de l'emoji 33
- 1.3 Interprétation des emojis 35
- 1.4 Usage des emojis 37
- 1.5 Recommandation ou suggestion ? 40
- 1.6 Représentation des emojis dans une perspective de recommandation 40
- 1.7 Conclusion 42
- 1.8 Synthèse 43

**2 État de l'art : prédiction et recommandation d'emojis** 45
- 2.1 De la compréhension des emojis aux ressources 45
  - 2.1.1 Similarité des emojis 45
  - 2.1.2 Sémantique des emojis 46
- 2.2 Ressources disponibles 49
  - 2.2.1 Inventaires : normalisation et description 49
  - 2.2.2 Claviers 50
- 2.3 Modèles prédictifs 50
  - 2.3.1 Emoji, sentiments et émotions 50
  - 2.3.2 Prédiction d'emoji 51
- 2.4 Discussion 52
- 2.5 Synthèse 55

**3 Prédiction automatique d'emojis** **57**
3.1 Introduction 57
3.2 Substitution lexicale par approche générative 58
3.2.1 Évaluation 64
3.2.2 Discussion 66
3.3 Enrichissement extra-linguistique par approche discriminante 67
3.3.1 Corpus de messagerie sociale privée 67
3.3.2 Ensemble des caractéristiques 69
3.3.3 Classification multi-étiquettes pour la prédiction d'emojis 72
3.3.3.1 Prédiction des emojis sentimentaux dans un corpus dédié 72
3.3.3.2 Prédiction des emojis dans un corpus étendu 78
3.4 Limites de la prédiction directe d'emojis 81
3.5 Conclusion 82
3.6 Synthèse 84

**4 Catégorisation d'emojis émotionnels** **85**
4.1 Introduction 85
4.2 Approche méthodologique 86
4.3 Catégorisation d'emojis par plongements lexicaux existants 88
4.4 De la nécessité de plongements spécifiques pour une catégorisation d'emojis 90
4.4.1 Corpus de tweets 91
4.4.2 Plongements lexicaux adéquats 93
4.4.2.1 Plongements lexicaux en skip-gram 93
4.4.2.2 Plongements lexicaux en sac de mots continu 95
4.5 Évaluation des plongements lexicaux par comparaison à une typologie des expressions faciales émotionnelles humaines 97
4.5.1 Partitionnement de référence 98
4.5.2 Métriques d'évaluations quantitatives des partitionnements automatiques 99
4.5.3 Évaluation quantitative des partitionnements automatiques 99
4.5.4 Analyse des erreurs 101
4.6 Conclusion 101
4.7 Synthèse 103

**5 Recommandation d'emojis par prédiction de leur catégorie** **105**
5.1 Introduction 105
5.2 Corpus d'apprentissage pour la recommandation d'emojis 107
5.2.1 Données du corpus 107
5.2.2 Pré-traitement des données 107
5.3 Encodage des données 108
5.4 Élagage et remplissage des messages 109

5.5 Prédiction de catégories d'emojis 109
5.6 Impact de la temporalité 116
5.7 Utilisation du système prédictif pour une recommandation 119
5.8 Limites 121
5.8.1 Multiple usages des emojis dans le corpus 121
5.8.2 De la différence entre tweets et messages informels privés 123
5.9 Conclusion 124
5.10 Synthèse 126

**6 Évaluation subjective par l'utilisateur 127**
6.1 Introduction 127
6.2 Évaluation du système de recommandation d'emojis pour l'enrichissement extra-linguistique 128
6.2.1 Conception de la plateforme de prédiction d'emojis 129
6.2.2 Résultats de l'évaluation du système de recommandation par prédictions combinées d'emojis 131
6.3 Évaluation du système de recommandation automatique de catégories d'emojis 132
6.3.1 Conception de la plateforme de conversations instantanées 132
6.3.1.1 Protocole expérimental 134
6.3.1.2 Questionnaire pré-expérience 134
6.3.1.3 Questionnaire post-expérience pour l'évaluation subjective de la recommandation 135
6.3.2 Résultats 135
6.3.2.1 Sarcasme, ironie 138
6.4 Limites 139
6.5 Conclusion et perspectives 140
6.6 Synthèse 142

**Conclusion générale 142**

**Bibliographie 149**

**ANNEXES 160**
A EmoReco : plateforme d'évaluation 160
A.1 Vue et contrôleurs 160
A.2 Modèles 162
A.3 Services et prédiction 162
B Déploiement Android 163
B.1 Motivation 163
B.2 Déploiement 164
B.3 Optimisation de taille 167
B.3.1 Optimisation de la taille du modèle 167
B.3.2 Optimisation de la taille de la librairie 167

B.4 Performances 168
B.4.1 Performances du modèle 168
B.4.2 Utilisation des ressources 169
B.5 Conclusion 170

# Liste des figures


0.1 Emojis originels (NTT DoCoMo, 1999) 25

1.1 Emojis aux interprétations les plus différentes 37
1.2 Distribution des usages d'emojis par catégories. (*SwiftKey Emoji Report* 2015). 39

2.1 Exemple de noms propres associés à l'emoji dans The Emoji Dictionary 49

3.1 Étapes de prédiction d'emojis selon le mot en cours et le mot précédent 59
3.2 Sélection de l'humeur (*mood*) par le biais de 38 emojis 71

4.1 Espace vectoriel d'emojis réduit à 2 dimensions 88
4.2 Groupes obtenus à l'aide de l'espace vectoriel existant 90
4.3 Représentation en deux dimensions par TSNE de l'espace vectoriel d'emojis obtenu par architecture skip-gram. 93
4.4 Groupes obtenus en partitionnant automatiquement les vecteurs d'emojis de la figure 4.3 à l'aide de K-moyennes. 94
4.5 Espace vectoriel d'emojis obtenu par sac de mots continus, réduit à deux dimensions. 95
4.6 Groupes obtenus en partitionnant automatiquement les vecteurs d'emojis de la figure 4.5 à l'aide de K-moyennes. 96

5.1 Réseau CNN-LSTM utilisé pour prédire les catégories d'emojis. 112
5.2 Comportement de la convolution mise en place. 113
5.3 Courbe de perte du CNN-LSTM sur 20 itérations. 117
5.4 Matrice de confusion du CNN-LSTM. L'ordre des étiquettes est le même que celui des scores détaillés 5.6, de gauche à droite et haut en bas. 120
5.5 Interface initiale de suggestion d'emojis avec distance d'édition et lexique dans l'application source créant un biais dans l'obtention du corpus. 122

6.1 Architecture globale de l'application 129
6.2 Interface de test utilisateur. La prédiction utilisée reprend les modèles appris au chapitre 3. 130
6.3 Emojis sélectionnables manuellement et possibilité d'indiquer l'absence d'emoji adéquat. 131
6.4 Impression écran du salon d'accueil développé. 132
6.5 Affichage des bulles d'emojis recommandés accompagnés du bouton de sélection manuelle. 133
6.6 Guide intégré pour s'assurer de l'usage attendu 135

6.7 Questionnaire d'évaluation présenté à chaque utilisateur après 5 scénarios. L'utilisateur indique sa réponse à travers une échelle de Likert de 5 points. 136
6.8 Sélection manuelle ou par recommandation des emojis selon les contextes émotionnels. 137
6.9 Résultats de l'évaluation par questionnaire. 138
6.10 Exemple de recommandation prenant en compte une ironie possible. 139
6.11 Exemple de recommandation à sens unique 139
A.1 Organisation principale de l'application 160
A.2 Composants VueJS et modèle de vue. 161
B.1 Système déployé sous Android 163
B.2 Utilisation de l'espace disque dans l'application finale 168
B.3 Analyse des consommations globales 169
B.4 Analyse de la consommation du processeur 170
B.5 Analyse de la consommation en mémoire 171

# Liste des algorithmes

1 Élagage et remplissage utilisés pour normaliser la taille des vecteurs. 110

# Liste des tableaux

1.1 L'emoji *face with rolling eyes* : des émotions et polarités opposées 37
1.2 Emoji *dancer* : différentes représentations selon le contexte logiciel 38
1.3 Exemples de représentations exclusives entre emojis et *kaomojis*. 39

3.1 Détails du corpus de texte utilisé pour entraîner les HMM 60
3.2 Format indicatif du lexique utilisé 62
3.3 Format du corpus d'évaluation pour le modèle génératif 64
3.4 Données du corpus d'évaluation par reproduction du contexte précédant un emoji de remplacement 65
3.5 Résultats globaux en macro selon différents filtres 65
3.6 Exemple factice d'une phrase représentée par la paire emojis|texte 68
3.7 Caractéristiques des deux corpus utilisés (l'un dédié aux emojis sentimentaux, l'autre étendu à tous les emojis). *valeurs prédites avec SentiStrength (Thelwall, Buckley et al., 2010). **valeurs prédites avec Echo (Hamdan, Bellot et al., 2015) uniquement pour le corpus dédié. 69
3.8 Moyennes des performances de prédiction d'emojis sentimentaux avec *ML-Random Forest*. *La lemmatisation est ignorée pour les sacs de caractères. 73
3.9 Les cinq caractéristiques discriminantes selon les scores d'importance des *RandomForest*. Avec différents ensembles de caractéristiques utilisées. 76
3.10 Impact empirique de chaque caractéristique par rapport à la "*baseline*" constituée d'un sac de caractères (moyennes obtenues pour 3 exécutions). *Scores prédits avec SentiStrength 77
3.11 Moyennes des prédictions d'emojis sentimentaux avec Random Forest et sac de caractères 79
3.12 Les cinq caractéristiques discriminantes calculées par les *RandomForest*, avec sac de mots ou de caractères. Pour ce dernier, les scores stagnent à 0.0058 au delà de la 5$^{e}$ position. 80

3.13 Comparaison des performances d'un même modèle général sur les deux jeux d'étiquettes (moyennes de 3 exécutions aux découpages aléatoires mais comparables) 81

4.1 Paramètres fixes des K-moyennes. 89
4.2 Corpus de tweets contenant des emojis 91
4.3 Partitionnement de référence établi manuellement pour le calcul des métriques. 98
4.4 Résultats comparatifs en variant uniquement le partitionnement appliqué. 100
4.5 Groupes d'emojis obtenus par regroupement spectral sur des plongements lexicaux en sac de mots continus. Les étiquettes proviennent des catégories d'expression de l'émotion d'Ekman. 100

5.1 Exemple de donnée après pré-traitements énumérés ci-dessus (répétition typographique, etc.). 108
5.2 Quadrillage ciblé d'hyper paramètres pour le CNN-LSTM. Les résultats utilisés sont des moyennes d'une validation croisée à 5 plis. 113
5.3 Moyennes macro de la précision, du rappel et de la F-Mesure pour chaque classifieur. C'est moyennes sont issues d'une validation croisée à 5 plis. 115
5.4 Architectures connexes pour estimer l'impact du LSTM sur la sortie du CNN. 118
5.5 Comparaison des performances avec des architectures connexes. 118
5.6 Résultats détaillés de la prédiction du CNN-LSTM par catégorie sur une séparation aléatoire du corpus. Les scores moyens sont pondérés en fonction du nombre d'occurrences (*support*). 119

6.1 Résultats de l'évaluation par analyse de la sélection d'emojis. 135

# Liste des abréviations

- **CNN** Réseau de neurone convolutif (*Convolutional Neural Network*)
- **HMM** modèles de Markov cachés (*Hidden Markov Model*)
- **LSTM** *Long-Short Term Memory network*
- **MRR** *Mean Reciprocal Rank*
- **SHS** Sciences Humaines et Sociales
- **SVM** Machine à vecteurs de support (*Support Vector Machine*)
- **TAL** Traitement Automatique des Langues
- **UX** Expérience utililsateur (*User eXperience*)

# Introduction générale

Les emojis trouvent leurs origines en 1999 lors de la préparation de l'*i-mode*, le premier système mobile majeur permettant un accès à internet. Ce système, ainsi que les emojis qui y étaient alors intégrés, fut créé par *NTT DoCoMo* [2]. La création des emojis est attribuée à Shigetaka Kurita, un des employés de l'entreprise. Pour la création des 176 premiers emojis dont certains sont visibles en Figure 0.1, Mr. Kurita fut inspiré par les caractères chinois, les bandes dessinées japonaises (aussi appelées *manga*) et les panneaux de signalisation.

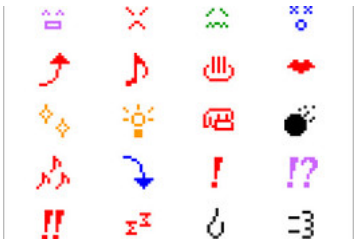

Figure 0.1. – Emojis originels (NTT DoCoMo, 1999)

Depuis leur création, l'usage des emojis n'a cessé d'augmenter et certains emojis (images) remplacent désormais certaines émoticônes (caractères ASCII). L'usage des emojis s'est désormais répandu à une pluralité de contextes logiciels dans lesquels ils apparaissent. Ces contextes sont variés mais les plus importants s'inscrivent dans un cadre d'échange social. Les contextes logiciels sont :

— les forums : avec bien souvent l'importation d'emojis personnalisés ;
— les clients web d'emails (*webmails*) : Gmail ou encore Outlook possèdent leurs claviers virtuels d'emojis, amenant par la même occasion les emojis dans le monde à connotation plus sérieuse des emails ;
— les applications de messagerie instantanée : WhatsApp, WeChat et Telegram possèdent chacun des emojis particuliers mais aussi des *stickers* (images statiques prédéfinies) ou autres images animées (GIF) ;
— les plateformes de réseaux sociaux : Facebook [3] utilise les emojis dans les réactions et commentaires, là où Instagram [4] et Twitter [5] leur permettent une association avec des mots-dièses au sein des commentaires, ainsi que dans les publicités affichées. Ces plateformes participent énormément à la création d'un usage global normalisé des emojis ;

2. https://www.nttdocomo.co.jp/
3. https://www.facebook.com
4. https://www.instagram.com/
5. https://twitter.com/

— les autres médias divers et variés : films sur les emojis, publicités ciblées pour les nouvelles générations, sites de rencontres pour l'évaluation de critères, *etc.*

Bien que ces contextes logiciels soient nombreux et importants vis-à-vis de l'évolution de l'utilisation des emojis, c'est surtout par la popularité du premier iPhone d'Apple [6], qui embarquait déjà des emojis à l'apparence améliorée, que leur usage a connu un véritable essor. Cette importante utilisation a entraîné la création constante de nombreux nouveaux emojis, au point d'atteindre 2789 emojis au cours de l'année 2018 [7], avec de nombreux à venir.

## Contexte doctoral

Le travail présenté dans ce manuscrit s'inscrit dans le cadre d'une thèse CIFRE associant le Laboratoire d'Informatique et Systèmes (LIS-CNRS UMR 7020) et l'entreprise Caléa Solutions.

Caléa Solutions est une petite entreprise (*startup*) marseillaise de moins de vingt employés possédant une application de messagerie mobile : Mood Messenger. Cette application Android permet d'envoyer des messages par SMS et internet, mais sa particularité consiste en l'enrichissement de l'expérience SMS, ainsi que la prédiction à la volée d'emojis lors de la frappe du message par l'utilisateur. C'est cette dernière fonctionnalité, au centre de la stratégie économique de l'entreprise, qui justifie la mise en place d'une thèse CIFRE : il est apparu nécessaire pour l'entreprise d'investir à moyen terme pour une prédiction d'emojis plus intelligente et adaptative.

Au niveau du laboratoire l'intérêt de cette CIFRE était d'explorer des méthodes de Traitement Automatique des Langues (TAL) et d'apprentissage automatique pour la prédiction d'objets liés à l'analyse de sentiments, que l'on trouve dans l'historique des travaux de l'équipe DiMaG (aujourd'hui R2I), équipe d'accueil au sein du laboratoire pour cette thèse.

## Problématiques et objectifs

L'utilisation grandissante des emojis fait émerger plusieurs problématiques, qu'elles soient théoriques ou applicatives.

6. https://www.apple.com/
7. http://unicode.org/emoji/charts/emoji-counts.html

**Au niveau théorique** tout d'abord, l'emoji est un objet complexe qu'il n'est pas trivial de définir et dont la définition dépend du point de vue dans lequel on se pose. Un emoji sera ainsi considéré comme un ensemble de pixels, ou une matrice à plusieurs dimensions, du point de vue d'une analyse de l'image. Au contraire, un emoji sera considéré comme le référent d'une émotion ou d'une idée pour un linguiste, reste la question de savoir s'il s'agit d'une représentation abstraite ou de la somme de plusieurs concepts. Toujours au niveau théorique, l'emoji pose la question du rôle qu'il joue dans la conversation, de l'apport qu'il donne comparé à une conversation sans emojis, et ainsi de définir l'intérêt théorique d'une recommandation d'emojis dans un contexte conversationnel.

**Au niveau applicatif**, la recommandation d'emojis propose une solution à la problématique du parcours de bibliothèques d'emojis. Dans une interface classique sans recommandation, l'utilisateur doit parcourir de longues bibliothèques d'emojis pour sélectionner celui qui lui convient. Cette étape est fastidieuse et est opposée à la volonté d'obtenir une expérience utilisateur (UX) fluide. De cela en découle une autre problématique applicative : comment estimer que la recommandation effectuée est efficace ? Comment évaluer la nécessité de plusieurs options proposées selon un contexte donné ? Comment s'adapter à l'utilisateur et à ses habitudes ?

Ces problématiques sont récentes et l'émergence des emojis de par le monde en fait apparaître de nouvelles régulièrement.

Les travaux de thèse présentés dans ce manuscrit possèdent plusieurs objectifs majeurs :

1. Identifier une méthode adéquate à la recommandation automatique d'emojis pour l'amélioration de l'expérience utilisateur ;
2. Parmi l'ensemble des emojis, traiter les emojis porteurs de sentiment afin de faire transparaître leurs nuances dans la recommandation ;
3. Obtenir un système considérant chaque type d'usages et de fonctions des emojis sentimentaux dans une conversation.

## Méthodologie proposée

Pour résoudre certains de ces objectifs ou répondre aux problématiques énoncées, des travaux récents ont été conduits par d'autres équipes de recherche. Des modèles de prédiction d'emojis ont été proposés pour prédire 5 à 20 emojis les plus utilisés, ainsi que diverses manières de représenter numériquement un emoji (décrits au Chapitre 2). Ces travaux présentent toutefois certaines limites.

Par exemple, l'ensemble des modèles proposés reposent sur une prédiction, et non une recommandation d'emojis. De plus, les travaux de recherche sur les emojis prennent rarement en compte les différents usages tels que la fonction référentielle ou la fonction expressive (Section 1.2) et se concentrent sur un ensemble restreint d'emojis.

Les travaux de recherche présentés dans ce manuscrit se distinguent des recherches existantes par une attention portée sur la recommandation à travers une approche par apprentissage automatique supervisé et non supervisé. La tâche que nous traitons principalement est donc celle de la classification d'éléments textuels, qu'il s'agisse aussi bien du texte des messages que des emojis. Nous posons ainsi la question de la définition d'un emoji en cherchant à savoir par ses différents usages si l'emoji est un élément textuel au même sens que peut l'être un message ou un mot.

## Contributions principales

La méthodologie proposée dans ce manuscrit répond aux problématiques posées en contribuant de plusieurs manières.

— Tout d'abord, l'attention est portée sur la recommandation d'emojis dans un contexte conversationnel informel et privé. Un corpus de messages privés pour appliquer et mettre au point le système de recommandation.

— La méthodologie permet de prendre en compte les différents types d'usages d'emojis associés à leurs fonctions dans la conversation. Plusieurs approches de prédiction sont ainsi dédiées à la prise en compte de ces fonctions.

— Cette méthodologie n'est pas figée et s'adapte à l'évolution des usages d'emojis en générant automatiquement les classes à prédire. Ainsi, c'est en s'éloignant de la prédiction directe d'emojis que cette méthodologie permet l'intégration directe de nouveaux emojis ou d'emojis propriétaires à l'entreprise, comme c'est le cas avec Mood Messenger.

— Finalement, il est montré que l'analyse de l'évaluation objective par métriques de performances ne reflète pas la qualité réelle de la recommandation. Une interface utilisateur reconstituant le contexte conversationnel privé est mise en place pour proposer une évaluation subjective de la recommandation d'emojis.

## Organisation du manuscrit

Le manuscrit est organisé de la manière suivante.

Dans le premier chapitre (Chapitre 1), nous situons le cadre théorique de ces travaux de recherche. Nous rappelons ainsi les différentes définitions de l'emoji. D'un point de vue sémiotique tout d'abord, dans lequel l'emoji est considéré comme un signe avec trois composantes distinctes. D'un point de vue linguistique ensuite, dans lequel l'emoji est considéré comme un signe linguistique composé de quatre caractéristiques essentielles qui le définissent. Nous y présentons également les différentes fonctions que les emojis peuvent exercer dans une conversation (Section 1.2) ainsi que les travaux récents sur l'interprétation et l'usage des emojis (Sections 1.3 et 1.4).

Le second chapitre (Chapitre 2), présente l'état de l'art des travaux liés aux emojis. Nous abordons les travaux dédiés à la compréhension des emojis en séparant ceux dédiés à l'étude de leur similarité de ceux dédiés à l'étude de leur signification (Section 2.1.2), avant de présenter les diverses ressources utilisées dans l'état de l'art (Section 2.2). Finalement, les travaux récents de prédiction d'emojis sont détaillés (Section 2.3) en séparant ceux utilisant les emojis pour améliorer la détection d'émotion ou de sentiments, de ceux dont l'objectif est la prédiction d'emojis en elle-même.

Dans le chapitre 3, nous proposons une prédiction de plusieurs emojis simultanés pour un contexte conversationnel privé en distinguant deux approches. Une approche générative (Section 3.2) est présentée afin de prendre en compte les fonctions expressives et référentielles de l'emoji. Puis une approche discriminante (Section 3.3) par classification multi-étiquettes est mise en place pour prédire plusieurs emojis en comparant son application sur des emojis sentimentaux ou non. Les caractéristiques utilisées sont orientées autour du sentiment et de l'humeur de l'utilisateur.

Le chapitre 4 s'intéresse à la catégorisation automatique d'emojis selon leurs usages afin de permettre l'obtention d'un jeu de méta-étiquettes d'emojis. L'obtention automatique de catégories est appliquée par apprentissage non supervisé aux emojis faciaux représentant des expressions de l'émotion (Section 4.4) afin de pouvoir être validé par une catégorisation d'émotions basiques existante (Section 4.5).

Dans le chapitre 5, la recommandation d'emojis est mise en place en réutilisant les catégories d'emojis obtenues automatiquement, pour une classification monoétiquette de catégories d'emojis. La tâche de classification est mise en place par apprentissage profond et ainsi détournée pour servir une recommandation lais-

sant le choix final à l'utilisateur.

Finalement, le chapitre 6 présente deux plateformes utilisées pour évaluer les systèmes présentés lors des chapitres 3 et 5. Une démonstration mise en place dans un stand d'un salon dédié à la valorisation de la recherche combine les deux approches présentées au chapitre 3 pour recommander des emojis aux utilisateurs (Section 6.2). Une plateforme d'évaluation est ensuite proposée pour évaluer la recommandation associant les travaux des chapitres 4 et 5, tout en reproduisant un contexte conversationnel privé.

En conclusion, nous résumons les travaux de recherche présentés dans cette thèse. Nous présentons ensuite différentes perspectives amenant à de futurs travaux possibles.

# 1. Contexte théorique

## Plan du chapitre


| | | |
|---|---|---|
| 1.1 | Introduction | 31 |
| 1.2 | Les fonctions de l'emoji | 33 |
| 1.3 | Interprétation des emojis | 35 |
| 1.4 | Usage des emojis | 37 |
| 1.5 | Recommandation ou suggestion ? | 40 |
| 1.6 | Représentation des emojis dans une perspective de recommandation | 40 |
| 1.7 | Conclusion | 42 |
| 1.8 | Synthèse | 43 |


## 1.1. Introduction

Dans ce chapitre, l'objectif est de présenter le cadre théorique d'étude des émojis à travers leurs définitions et les fonctions qu'ils exercent dans la conversation.

De nouveaux emojis apparaissent régulièrement, augmentant leur popularité. Cependant, la popularité n'est pas le seul facteur de création de nouveaux emojis, l'est aussi la nécessité d'étendre les possibilités d'expression. En effet les emojis sont des signes, et en tant que tels, ils servent à représenter des idées diverses et parfois abstraites. Pour aborder la question de la nature d'un emoji, il convient de se rapporter dans un premier temps à l'étude des signes, la sémiotique, afin de comprendre les signes que sont les emojis. Ainsi, si l'on se réfère à la théorie des signes de Charles Sanders Peirce (PEIRCE, 1902), l'emoji 🙀 est un signe possédant un *representamen* par l'image de l'emoji, un objet par le chat, ses pattes et son expression, ainsi qu'un interprétant final par l'idée de surprise non restreinte aux chats. Cet interprétant final explique que cet emoji est souvent utilisé pour faire référence à l'objet de la surprise, plus qu'à l'objet du chat. D'un point de vue linguistique, nous pouvons distinguer au sein des emojis les trois composantes qui définissent un signe linguistique selon Saussure (DE SAUSSURE, 1916) : le signifiant étant l'emoji en tant qu'image, le signifié correspondant à l'idée à laquelle fait référence l'image de l'emoji et enfin l'interpréteur correspondant à l'idée que l'on se fait de la relation entre les deux concepts précédents. Si l'on prend un exemple concret, le signifiant pourrait être 🙀, le signifié serait "*chat surpris*" et l'interpréteur serait alors le concept de surprise non nécessairement associé à l'idée du chat, selon le contexte. Il apparaît alors évident que l'interpréteur est fortement orienté par la *doxa* du destinataire, l'interprétation des emojis

peut alors fortement varier en fonction de celle-ci. De plus, le signifiant se révèle d'une grande importance puisqu'il peut changer en fonction de l'émetteur du signe : le même emoji 😾 ou encore 😿 correspond à 😹 dans la version Facebook, entraînant alors un changement de stimulus à chaque fois, ainsi qu'un changement complet de signifiant pour l'emoji 😹.

Ces concepts de signifiants et signifiés provenant de la linguistique (De Saussure, 1916), il s'agit finalement bien d'une description des raisons de l'ambiguïté propre aux emojis, allant également dans le sens de leur prise en compte en tant que mots à part entière. Les emojis représentant un ou plusieurs concepts ne sont en effet pas si éloignés des mots, et plus exactement des termes ou d'un jargon particulier. Ils répondent la plupart du temps à des champs lexicaux précis auxquels ils essaient de plus ou moins de se conformer. Il est ainsi possible de trouver des emojis de végétaux, de sport, de sentiments, d'animaux et autres, autant d'emojis que nécessaire afin de pouvoir exprimer les différents éléments du champ lexical associé, à des granularités différentes. Et quand un élément ne correspond pas forcément à l'association qui lui est désirée, il voit alors son signifiant modifié, comme ce fut le cas pour l'emoji du danseur représenté différemment en 2016 chez Samsung 🕺, Google 💃 et Twitter 💃, et s'étant uniformisé en 2017 en deux emojis 💃 🕺 désormais visibles dans la liste de référence Unicode [1]. Il s'agit pourtant dans ce cas d'une normalisation de fait toujours susceptible de changer.

L'actuelle multiplication des emojis ne fait cependant pas toujours réponse à un besoin de signifiants pour des signifiés, des concepts et idées n'ayant pas leur emoji associé. Elle correspond à l'une des quatre caractéristiques essentielles pour qu'un ensemble de signes deviennent une langue (De Saussure, 1916) :

1. le caractère arbitraire du signe. Les emojis ont été décidés par un groupe de personnes et auraient très bien pu être totalement différents selon les personnes choisies ;
2. la multitude des signes. Il y a de nombreux emojis et plus à venir encore ;
3. la grande complexité du système. Les emojis peuvent suffirent à communiquer et leur association et utilisation est bien plus complexe qu'il n'y paraît ;
4. l'inertie collective. Les emojis sont la possession de tous et tendent à se normaliser.

Mais le besoin de davantage d'emojis est également sociologique puisqu'il peut correspondre à l'avènement de mouvements et tendances périodiques, tels peuvent l'être des élections nationales ou un événement sportif par exemple. De

1. http://unicode.org/emoji/charts/full-emoji-list.html

plus, comme dit précédemment, les emojis sont des objets d'une langue, et celle-ci, en tant que "système de signes exprimant des idées" est une institution sociale (DE SAUSSURE, 1916). Mais le véritable impact sociologique sur les emojis, et celui qui nous intéresse précisément ici, est la désambiguïsation sociale, c'est-à-dire le fait qu'un emoji est créé en prenant en compte l'attente d'une interprétation spécifique issue, dans certains cas, d'un ancrage fortement social, temporaire ou non. C'est par exemple le cas des récents emojis mis en place à ce jour : certains sont issus d'une *doxa* propre au web et aux réseaux sociaux comme 🤯 qui a pour objectif d'être interprété comme une incompréhension ou une incrédulité, là où sans *doxa* il serait simplement interprété comme une tête qui explose. De nombreux nouveaux emojis viennent de "mêmes" issus de réseaux sociaux ou forums populaires sur internet, comme le forum Reddit[2]. D'autres types d'emojis plus facilement interprétables, sans contexte social, émergent cependant en résultat d'une pression sociétale, politique ou non. C'est le cas des variantes de couleurs de peau (*skin tones*), des différentes représentations masculines et féminines de professions 👨 👩, ou encore de différentes places emblématiques pour certaines religions ⛪ 🕌 🏰.

## 1.2. Les fonctions de l'emoji

D'un point de vue linguistique, les emojis jouent différents rôles au sien d'une conversation. Nous avons déjà fait le lien entre emoji, sémiotique et linguistique en les définissant comme des signes et en montrant qu'ils possèdent toutes les caractéristiques du signe linguistique tel qu'il est traditionnellement défini. Il en va de même pour une conversation dans laquelle sont utilisés des emojis. En effet, une conversation n'est autre qu'une instance ordonnée d'une communication entre plusieurs individus. Et c'est cette communication que les emojis viennent enrichir en tentant d'y apporter des éléments propres à l'acte parlé et externes au contexte de la conversation écrite. Ainsi, l'application des fonctions du langage de Roman Jackobson (JAKOBSON, 1963) devient utile à l'analyse du rôle des emojis. Parmi les six fonctions de la communication, les emojis en remplissent plusieurs :

**L'expression.** L'emoji est avant tout un moyen d'expression d'une idée, comme l'est tout signe linguistique. Il permet en effet d'exprimer le sentiment du locuteur vis-à-vis d'un propos ou d'un objet, ou encore d'exprimer toute une sémantique à l'aide d'un enchaînement d'emojis. C'est par exemple le cas lors de messages dénués de mots et seulement constitués d'emojis pour transmettre une ou plusieurs informations.

2. https://www.reddit.com/

**La fonction phatique.** Les emojis sont également utilisés pour jouer un rôle phatique dans la conversation, c'est-à-dire permettre de garder le contact sans nécessairement avoir une information à transmettre. Là où cette fonction peut être quelquefois remplie à l'écrit par le biais d'interjections, elle est plus courante lors d'une conversation orale en face à face puisqu'un simple sourire, regard, ou hochement de tête peuvent servir à garder contact avec son interlocuteur et à s'assurer de sa toute pleine attention. Les emojis représentant ce genre d'expressions du visage sont certaines fois utilisés à cet égard, constituant généralement la totalité du message. À ce titre, certaines applications de communication sociales[3] intègrent désormais une fonctionnalité de réaction à un message textuel par le biais d'un unique emoji, cherchant ainsi à séparer certains cas d'utilisation de l'emoji à des fins de fonction phatique du flux de la conversation.

**Méta-linguistique et désambiguïsation.** De par son aspect différent d'un texte normal, l'emoji peut non seulement servir à exprimer une émotion ou une idée, mais cette expression en elle-même peut permettre de désambiguïser un contenu. C'est ainsi lorsqu'un emoji est volontairement utilisé pour exprimer un sens contraire à se qui transparaît dans le texte, signalant par la même occasion l'utilisation du sarcasme ou encore de l'ironie. Ce rôle méta-linguistique n'est pas exactement l'utilisation d'un emoji pour parler d'autres emojis, mais une déviante dans laquelle l'emoji se soustrait alors de sa qualité de simple signe linguistique pour s'appliquer comme un complément aux autres mots présents dans le texte. On pourrait alors parler d'emojis prenant une fonction de didascalie du genre théâtral.

**L'intégration du contexte par la fonction référentielle.** Lorsque les emojis sont utilisés pour désigner un objet non présent dans le contenu textuel il exerce une fonction de référence à cet objet, par exemple : "Je me suis rendu au travail rapidement 🚴" désigne le vélo comme moyen de transport et explication supplémentaire à l'information textuelle véhiculée dans la phrase, de la même manière que l'on désignerait le vélo en le pointant du doigt lors d'une conversation en face à face.

**Fonction conative et usage des influenceurs.** Les emojis peuvent être utilisés uniquement pour influencer le récepteur, et notamment pour cibler une tranche d'âge d'individus afin de leur faire considérer la marque proposée comme plus proche, amicale, ou tout simplement plus attrayante que les autres. Il a été montré par approche empiriste que l'usage des emojis dans les publicités permettait dans la majorité des cas de la rendre plus efficace, et ainsi d'augmenter les gains

3. Ceci est par exemple le cas de Discord (http://discordapp.com/) ou Slack (https://slack.com/)

qui en découlent (TEAM, 2015).

**Attrait du signifiant et fonction poétique.** L'emoji en lui-même peut être utilisé pour simplement faire référence à la particularité du signe, du dessin, de l'image qui constitue son signifiant. Ce cas, plus en retrait, est notamment présent lorsque de nouveaux emojis apparaissent, ces derniers sont alors l'objet du message et peuvent simplement être utilisés à des fins de démonstration.

L'emoji peut donc exercer différentes fonctions dans une conversation. Par ce qu'elles expliquent la motivation de l'usage d'un emoji, ces fonctions sont déterminantes pour identifier le ou les emojis à recommander. Ces fonctions sont difficilement prédictibles uniquement à partir du texte, ce qui en fait l'une des problématiques majeures dans la recommandation automatique d'emojis. Dans le cadre de cette thèse, nous nous sommes concentrés sur les fonctions d'expression, de référentiel, d'enrichissement méta-linguistique. Les deux dernières ont fait l'objet de deux approches différentes, tandis que les fonctions conatives et poétiques sont implicitement prises en compte par la nature du corpus d'apprentissage.

## 1.3. Interprétation des emojis

Avant de manipuler les emojis, que ce soit pour les recommander ou simplement les prédire, il est nécessaire de comprendre ce qu'ils sont : qu'est-ce qu'un emoji exactement ? Plusieurs travaux ont essayé de mieux comprendre les emojis en se basant notamment sur la place qu'ils prennent dans la société et dans l'interaction entre les personnes. Cet aspect sociologique a fait l'objet de plusieurs travaux, dont certains portés sur les émoticônes peuvent être croisés avec un sous-ensemble d'emojis actuels, les emojis représentant des expressions faciales. Ainsi, en 2013 Jibril *et al.* (JIBRIL et ABDULLAH, 2013) ont publié un état de l'art mettant en avant l'importance des émoticônes dans les conversations électroniques, auxquelles nous ferons référence par l'acronyme CMC (*Computer-Mediated Communication*). Cette étude représente également un premier pas vers la considération des émoticônes comme un objet linguistique à part entière, en plus d'un simple objet de reproduction du contexte de la discussion en face à face. Ce rôle est détaillé davantage par la suite dans les travaux de Skovholt (SKOVHOLT, GRØNNING et al., 2014) dédiés aux fonctions des émoticônes dans les communications par courriels dans lesquels ils identifient 12 fonctions possibles, regroupées par catégories :

1. Signature (Positive)
2. Jokes/irony (Humor)
3. Request (Softener)

4. Rejection (Softener)
5. Correction (Softener)
6. Complaint (Softener)
7. Thanks (Strenghteners)
8. Greetings (Strenghteners)
9. Wishes (Strenghteners)
10. Appraisal (Strenghteners)
11. Promises (Strenghteners)
12. Admissions (Strenghteners)

Ces catégories et fonctions sont issues uniquement de courriels professionnels, rendant ces conclusions difficilement applicables à d'autres contextes de communication. De plus, par ce contexte professionnel, ils remettent en avant l'impact des actes de langage (AUSTIN, 1975 ; SEARLE, 1969) et de la politesse (LEECH, 2016), avec notamment le fait qu'utiliser des émoticônes et "smileys" dans des courriels professionnels ait un impact sur le fait d'être ou de ne pas être pris(e) au sérieux par son interlocuteur.

L'interprétation des emojis n'a pas seulement été abordée par l'analyse du rôle qu'ils exercent dans la conversation mais également via leur sémantique. En 2016, plusieurs travaux de recherche ont montré le difficile consensus à trouver lorsqu'il s'agit de définir précisément le sens apporté à tel ou tel emoji (C. KELLY, 2015 ; GUIBON, OCHS et al., 2016). La prise en compte des différences entre contexte conversationnel formel et informel, ainsi que la manière d'aborder les emojis joue un rôle dans l'interprétation de ces derniers. En effet, un emoji se veut différent d'une émoticône en ce qu'il ne constitue pas un ensemble figé de représentations graphiques à combiner, mais une image à part entière qui peut alors potentiellement tout représenter. Dans leur enquête sous forme de questionnaire, Kelly *et al.* (C. KELLY, 2015) montrent par exemple qu'une simple phrase telle que "I miss you 😄" est considérée à 19% sarcastique, 16% amicale et 26% honnête et sincère. Aussi, 87% ont répondu favorable à l'affirmation selon laquelle les emojis possèdent plusieurs significations et ont estimé que dans 70% des cas les emojis facilitaient l'interprétation du message. Il a également été montré, à l'aide d'une enquête, que l'interprétation des emojis varie beaucoup en fonction de leur représentation graphique (H. MILLER, THEBAULT-SPIEKER et al., 2016), ce qui fut le cas notamment pour l'emoji de la danse dont nous montrons un exemple au tableau 1.2. Dans leur étude faite à l'aide d'*Amazon Mechanical Turk*, ils ont obtenu 5 000 interprétations pour 25 emojis ayant chacun 5 représentations différentes et en ont calculé les différences d'interprétations, dont les plus grandes sont visibles en figure 1.1.

| | Apple | Google | Microsoft | Samsung | LG |
|---|---|---|---|---|---|
| Top 3 | 3.64 | 3.26 | 4.40 | 3.69 | 2.59 |
| | 3.50 | 2.66 | 2.94 | 2.36 | 2.53 |
| | 2.72 | 2.61 | 2.35 | 2.29 | 2.51 |

Figure 1.1. – Emojis aux interprétations les plus différentes

| | **Apple** | **Samsung** |
|---|---|---|
| ***Face with rolling eyes emoji*** | | |
| **Polarité** | Légèrement négative | Positive |
| **Émotions** | Pas content / Triste / Hésitant | Content / Surpris |

Tableau 1.1. – L'emoji *face with rolling eyes* : des émotions et polarités opposées

Le tableau 1.1 montre un exemple de polarités opposées issues des différences de représentation graphique de l'emoji. Il s'agit donc bien du même emoji, si l'on considère son code de référence comme signifié de l'emoji, mais ayant un deux signifiants différents en fonction de la plateforme utilisée, un iPhone d'Apple ou un Galaxy de Samsung par exemple. Bien entendu, dans ce cas présent les différences de polarités sont grandes et dues aux différents indices faciaux représentés graphiquement, mais dans d'autres cas ces différences peuvent être minimes tout en entraînant une interprétation différente de l'emoji.

La polarité n'est pas la seule donnée dépendante de l'interprétation de l'emoji par l'utilisateur. De rares fois, l'emoji représente bien plus que le simple concept qui lui est attribué : son signifiant représente plusieurs signifiés possibles, lui conférant par là une forte ambiguïté. L'exemple que nous montrons est celui de l'emoji de la danse, ou du danseur et de la danseuse. Le tableau 1.2 montre trois versions de cet emoji restées actives plusieurs années jusqu'à leur normalisation en 2017 en séparant l'emoji danseuse de l'emoji danseur .

## 1.4. Usage des emojis

L'étude de l'usage des emojis peut être considéré comme remontant à celles sur les émoticônes avec notamment Tossel *et al.* (TOSSELL, KORTUM et al., 2012) qui

| | Google | Twitter | Samsung |
|---|---|---|---|
| ***Dancer* emoji** |  |  |  |

Tableau 1.2. – Emoji *dancer* : différentes représentations selon le contexte logiciel

ont collecté 6 mois de SMS issus d'iPhone pour étudier l'usage des émoticônes afin de savoir si différences il y a entre hommes et femmes ou encore la distribution de l'utilisation des émoticônes selon leurs types : **:) :( :D ;)** etc. Les travaux de Tossel *et al.* ont également un impact sur l'usage des emojis puisqu'ils sont parmi les premiers à étudier des messages privés, contrairement aux études habituelles sur les messages publics. Suivant cette même logique, plusieurs rapports ont été publiés en dehors de laboratoires de recherche publique sur l'usage des emojis dans les nouvelles technologies numériques. Ce fut le cas en 2015 avec deux rapports d'entreprises différentes : Microsoft avec SwiftKey[4] et Emogi[5]. Le premier, Swiftkey, est un clavier multi fonctions pour système android dont l'entreprise a publié un rapport (*SwiftKey Emoji Report* 2015) sur l'analyse d'un milliard d'emojis utilisés au travers de leur application. Ils ont montré que quel que soit le langage utilisé, les emojis positifs sont toujours les plus utilisés, notamment les visages heureux qui représentent la majorité des emojis comme le montre le graphique visible en figure 1.2.

Le second, Emogi, est un moteur de contenu additionnel pour des conversations privées dont l'équipe de recherche a montré que les emojis ont tendance à remplacer certains mots d'argot (TEAM, 2015). Surtout, ils ont cherché à savoir la raison derrière l'utilisation des emojis. Une enquête a donc été effectuée au près des utilisateurs, arrivant à la conclusion que la majorité des utilisateurs utilisent les emojis pour être mieux compris par leurs interlocuteurs. L'ordre décroissant des raisons d'utilisation des emojis démontré dans cette étude est le suivant :

1. Aide à la compréhension du message ( 70%)
2. Création d'une connexion personnelle ( 50%)
3. Facilité d'utilisation ( 40%)
4. Une façon comme une autre de communiquer ( 25%)

D'autres travaux de recherche se sont focalisés sur l'analyse de l'usage des emojis afin de savoir s'ils empiètent sur les émoticônes. Pavalanathan *et al.* (PAVALANATHAN et EISENSTEIN, 2015) ont ainsi collecté des tweets en anglais de

4. https://www.microsoft.com/en-us/swiftkey?rtc=1
5. https://www.emogi.com/

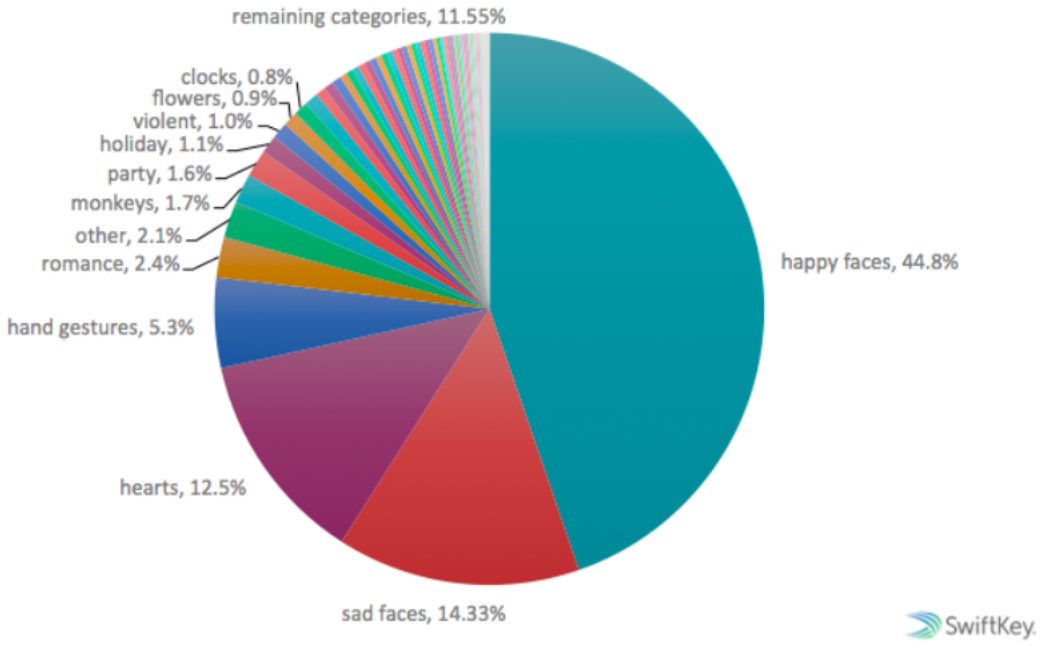


Figure 1.2. – Distribution des usages d'emojis par catégories. (*SwiftKey Emoji Report* 2015).

février 2014 à août 2015 et l'ont soumis à deux groupes : des utilisateurs confirmés d'emojis (groupe de traitement) et des utilisateurs n'ayant pas encore utilisé d'emojis (groupe de contrôle). Par leur approche orientée données, ils ont montré que les émoticônes positives telles que **:-)** et **:P** ont un taux de disparition plus grand que les émoticônes négatives comme **:(**. Ils accréditent ce taux de remplacement élevé au choix plus divers d'emojis positifs comparés aux emojis négatifs. Mais surtout, leur étude a confirmé l'hypothèse selon laquelle les emojis remplacent de plus en plus les émoticônes occidentales habituelles, les *kaomojis* n'étant pas pris en compte dans cette étude. Le tableau 1.3 montre quelques exemples d'expressions exclusives aux *kaomojis* et emojis.

| | **Emoji** | ***Kaomoji*** |
|---|---|---|
| **Sourire** |  | (°ʖ°) |
| **Expression *"Why ? !"*** | ✖ | Щ(°Д°Щ) |
| **Massage de tête** |  | ✖ |

Tableau 1.3. – Exemples de représentations exclusives entre emojis et *kaomojis*.

Les différents rôles des emojis à travers leur usage dans une conversation sont aussi un axe d'étude utilisé. Ryan Kelly et Leon Watts (R. KELLY et WATTS, 2015) ont ainsi mis en place des interviews avec des participants de huit pays (Royaume-Uni, États-Unis, Malaisie, Espagne, Italie, Allemagne, Inde et Singapour) afin de mieux connaître l'utilisation des emojis et les motivations derrière ces usages. Ils ont ainsi démontré que les emojis ne sont pas utilisés que pour transmettre des émotions et des humeurs, mais permettent aussi de contrôler l'intonation du message, une prosodie du message écrit, en plus de pouvoir permettre de cacher ou masquer certains sentiments. Les travaux de Kelly (R. KELLY et WATTS, 2015) montrent bien un usage des emojis qui va au-delà de l'émotion puisqu'ils définissent trois autres rôles : le maintien de la conversation (qui s'accorde avec la fonction phatique du chapitre 1), le jeu d'une interaction amusante et la création d'une atmosphère d'intimité dans la conversation.

## 1.5. Recommandation ou suggestion ?

L'objectif principal de ces travaux de thèse est de pouvoir proposer des emojis à l'utilisateur lors de la frappe d'un message. Toutefois, les méthodes proposées plus loin dans ce manuscrit ne correspondent pas toujours à l'idée de recommandation habituelle en TAL. En effet, le domaine de la recommandation est précis et possède déjà des approches et méthodes spécifiques, là où notre approche se situe davantage entre prédiction et suggestion ou recommandation à proprement parler. Il convient donc de préciser qu'au cours de ce manuscrit nous utiliserons le terme "recommandation d'emojis" dans un sens plus large et ne faisant pas nécessairement référence au domaine précis de la recommandation. En effet, un même contexte induit les mêmes emojis à recommander à l'utilisateur au fur et à mesure de sa rédaction, ce qui se distingue d'une recommandation de films par exemple, pour laquelle un film déjà vu ne doit pas être recommandé à nouveau. La recommandation des emojis entraîne donc un aspect redondant issu de leur utilisation.

## 1.6. Représentation des emojis dans une perspective de recommandation

Une des problématiques de recherche est la représentation des emojis. En tant qu'objets hybrides, les emojis véhiculent aussi bien linguistiquement que visuellement de l'information comme l'expression de sentiments ou d'émotions diverses et variées sur lesquels nous nous focalisons dans cette thèse. Cette problématique qui consiste à déterminer la représentation adéquate des emojis se voit adressée par le biais de leur représentation et leur recommandation en prenant en compte

les différentes fonctions des emojis dans une conversation.

Déterminer la manière idéale pour recommander des emojis nous permet également de mieux identifier comment les représenter. Ainsi, les emojis peuvent être considérés comme des objets uniques à part entière, autonomes et suffisants quant à l'expression de leur signification. Cette représentation peut se traduire par une recommandation différente pour chaque approche : les emojis peuvent ainsi être considérés comme faisant partie d'une hiérarchie fixe qu'il convient de recommander en choisissant une granularité plus ou moins précise, ou alors ils peuvent considérés comme des éléments d'un groupe, chaque groupe étant indépendant des autres. Aussi, la représentation comme objets autonomes et uniques entraînerait alors la prise en compte de chaque emoji comme un élément à recommander.

Adresser la recommandation comme une solution d'exploration du problème de représentation des emojis s'applique également dans l'approche technique : faut-il considérer les emojis comme de simples mot ou comme des expressions linguistiques figées servant à définir la ou les idées représentées par l'emoji ? (EISNER, ROCKTÄSCHEL et al., 2016) Chaque emoji doit-il être représenté par un vecteur d'agrégations d'informations contextuelles, ou par un simple encodage binaire de vocabulaire ? La bonne représentation technique découle ainsi de l'approche choisie pour représenter les emojis, et des traits mis en avant.

Trouver la bonne représentation des emojis passe également par l'impact qu'elle peut avoir sur la recommandation effectuée. Il est toutefois difficile d'évaluer cette recommandation, et d'autant plus difficile d'évaluer dans ce système quelle en a été la représentation adéquate des emojis. L'évaluation de la recommandation interroge notamment sur le bien fondé d'une évaluation objective, à l'aide de métriques quantitatives, de la recommandation. *A contrario*, peut-on considérer l'évaluation subjective comme utile et réalisable ? Quelle méthode de l'évaluation est la plus pertinente et reflète avec davantage de précision la qualité intrinsèque de la recommandation ?

La recommandation en prenant en compte les fonctions des emojis, ainsi que son évaluation, sont liées à la représentation des emojis. Ils influencent l'ordonnancement des travaux effectués dans le présent manuscrit en cherchant dans un premier temps à recommander correctement les emojis, puis à parfaire l'évaluation de cette recommandation.

## 1.7. Conclusion

Les emojis sont un objet simple au premier abord mais finalement de nature complexe. Ce chapitre présente le contexte théorique des emojis, leur nature et leur rôle dans la conversation à l'aide d'existants sémiotiques et linguistiques. Du point de vue de la sémiotique, un emoji est un signe composé de trois parties distinctes : son image (signifié), son concept (signifiant) et son interprétation (interpréteur). D'un point de vue linguistique en revanche, ces parties décrivent la difficulté d'interprétation de l'emoji. Les emojis sont par nature des objets ambigus pour lesquels un seul champ lexical ne suffit pas toujours à les définir.

Dans une conversation, ils ont plusieurs fonctions possibles en permettant d'exprimer une idée, garder le contact, préciser le message textuel, le remplacer, influencer l'interlocuteur ou encore s'auto-référencer. Chacune de ces fonctions ont un impact direct sur la conceptualisation d'une recommandation d'emojis : la fonction désirée guide l'usage de l'emoji et ainsi, l'emoji à recommander. C'est aussi pour cela, qu'il s'agit d'une recommandation et non d'une suggestion d'emojis.

Ce chapitre met en avant la problématique de recherche principale qui consiste tout d'abord à savoir comment représenter l'emoji pour mieux le recommander à l'utilisateur tout en prenant en compte ses fonctions possibles. Est-il nécessaire d'obtenir une approche différente pour chaque fonction ? La recommandation d'emojis consiste-t-elle à recommander chaque emoji ou plutôt des concepts, des signifiants communs à plusieurs signifiés ? Une fois une telle recommandation effectuée, comment faire pour l'évaluer ?

## 1.8. Synthèse

**Un emoji = un signe** défini en sémiotique par un signifié, un signifiant et un interpréteur

**6 fonctions d'un emoji dans une conversation** : exprimer, désambiguïser, référer, influencer, conserver la conversation vivante et utiliser l'emoji pour son apparence particulière.

**Interprétation des emojis.** Les emojis peuvent être interpétés différemment en fonction de leur signifiant ainsi que leurs fonctions. Plusieurs études par questionnaire ont émergé sur cette question.

**Usage des emojis.** La manière dont les emojis sont utilisés à fait l'objet d'études mettant en avant l'objectif de leur usage, les tendances par pays, ainsi que la domination des emojis joyeux.

**Recommandation** $\neq$ **suggestion**. La recommandation est un domaine précis mais recommander des emojis se situe dans un entre-deux.

**Problématiques** de recherche :

1. Représentation de l'emoji
2. Recommandation adaptée aux emojis et pour chaque fonction de l'emoji
3. Évaluer au mieux la recommandation de l'emoji

# 2. État de l'art : prédiction et recommandation d'emojis

## Plan du chapitre


2.1 De la compréhension des emojis aux ressources **45**
2.1.1 Similarité des emojis 45
2.1.2 Sémantique des emojis 46
2.2 Ressources disponibles **49**
2.2.1 Inventaires : normalisation et description 49
2.2.2 Claviers 50
2.3 Modèles prédictifs **50**
2.3.1 Emoji, sentiments et émotions 50
2.3.2 Prédiction d'emoji 51
2.4 Discussion **52**
2.5 Synthèse **55**


## 2.1. De la compréhension des emojis aux ressources

Plusieurs travaux de recherche en informatique explorent des méthodes d'apprentissage automatique pour identifier des typologies d'emojis et ainsi en extraire leur sens ou similarités sous-jacentes. Dans cette section, nous distinguons deux types de travaux : ceux dédiés à l'étude de leur similarité et ceux dédiés à étudier directement leur signification. Ceux deux types d'études ont pour but commun une meilleure compréhension des emojis et se traduisent parfois dans la création de ressources ou de méthodologies spécialisées préparant un terrain favorable à l'exploitation des emojis. Pour cela, la création de ressources ou de méthodologies spécialisées est nécessaire. Toutefois, la majorité des ressources construites sont obtenues suite à la recherche d'un but principal commun : la compréhension des emojis.

### 2.1.1. Similarité des emojis

La similarité des emojis a principalement été étudiée avec l'éclosion des plongements lexicaux (MIKOLOV, CHEN et al., 2013 ; MIKOLOV, SUTSKEVER et al., 2013), alors utilisés pour déterminer non pas la similarité des mots entre eux à partir de leur contexte, mais des emojis. C'est ainsi que Barbieri *et al.* (BARBIERI,

RONZANO et al., 2016) ont appris des plongements lexicaux en architecture *skip-gram* selon l'architecture proposée par Mikolov (MIKOLOV, CHEN et al., 2013) pour obtenir des vecteurs d'emojis en contexte à partir de 9 millions de tweets originaires des États-Unis. En variant le pré-traitement des tweets avec la suppression ou non des mots-vides, la conservation ou non des ponctuations, ils ont montré que de meilleurs résultats de calcul de similarité sont obtenus avec des données davantage pré-traitées. L'évaluation est effectuée en comparant les similarités obtenues avec 50 paires d'emojis conçues manuellement et étiquetées par similarité et parenté (*relatedness*).

Apprendre des représentations vectorielles d'emojis n'est pas l'apanage des études dédiées sur les emojis comme l'ont prouvé Soroush *et al.* (VOSOUGHI, VIJAYARAGHAVAN et al., 2016) en mettant en place des plongements lexicaux de caractères à partir d'un corpus de tweets, dont certains caractères représentaient des emojis et leurs codes. Pour cela ils ont mis en place un encodeur-décodeur constitué de réseaux de neurones récurrents *Long Short Term Memory* (HOCHREITER et SCHMIDHUBER, 1997) (LSTM) associés à plusieurs couches de convolution (LECUN, BENGIO et al., 1995) (CNN) afin d'obtenir une représentation des tweets, ce qui leur a permis d'obtenir des scores élevés en classification de tweets par sentiment et par association sémantique. Une autre tentative (DHINGRA, ZHOU et al., 2016) similaire de représentation des tweets par plongements lexicaux et l'usage d'un décodeur-encodeur fut mise en place à l'aide de réseaux récurrents à portes (*Gated Recurrent Units*) prenant en entrée du modèle une table de correspondance de caractères du tweet dans laquelle chaque emoji devient un charactère supplémentaire. Cette fois-ci le gain a été évalué par la prédiction de mots-dièses liés aux tweets.

L'étude de la similarité des emojis a aussi été effectuée dans le but d'améliorer la disposition des emojis sur les claviers virtuels dédiés disponibles sur la quasi totalité des smartphones (POHL, DOMIN et al., 2017). Pour cela, Pohl *et al.* (POHL, DOMIN et al., 2017) ont également mis en place des plongements lexicaux d'emojis sur un ensemble de tweets puis ont utilisé la mesure de similarité de Jaccard (JACCARD, 1912) à partir des mots clés de l'Unicode afin d'obtenir des paires d'emojis, qui sont alors évaluées à partir de 90 paires manuellement annotées.

### 2.1.2. Sémantique des emojis

La seconde principale approche utilisée pour la compréhension des emojis génératrice de ressources est l'étude du sens des emojis, non pas par la discipline qu'est la sémantique, mais par l'utilisation d'approches automatisées ou semi-automatisées. Ce fut par exemple le cas de Soroush *et al.* (VOSOUGHI, VIJAYARAGHAVAN et al., 2016) qui ont mis en place un encodeur-décodeur afin d'améliorer

leur système sur la tâche d'association sémantique, mais d'autres ressources sont davantage lexicales.

L'une des premières ressources lexicales dédiées aux emojis a été publiée en 2015 par Petra Kralj Novak *et al.* (KRALJ NOVAK, SMAILOVIĆ et al., 2015) avec un lexique de correspondance sentiment-emoji. Créé avec l'aide 83 annotateurs sur 1.6 million de tweets dans 13 langues de pays européens, leur but a été de définir la polarité des 751 emojis les plus fréquents sur Twitter à ce moment donné [1]. Il s'agit donc d'une approche différente de l'apprentissage d'un lexique de sentiment comme cela a pu être le cas en utilisant les émoticônes et le texte de tweets comme entrée de réseaux de neurones (TANG, WEI et al., 2014). En effet, leur approche fut manuelle et pour chaque emoji, trois scores sont étiquetés en contexte : les scores de positivité, négativité et neutralité. Le second apport de ce lexique est la création d'un score de sentiment qui est la moyenne des probabilités discrètes sur chaque polarité possible (positif, négatif, neutre), donc si $c$ est l'étiquette de polarité et $N$ les occurrences de cet emoji dans les tweets, ils définissent les occurrences par polarité comme suit :

$$N(c), \sum_c N(c) = N, c \in \{-1, 0, +1\} \tag{2.1}$$

Et définissent les probabilités discrètes de sentiment d'emoji ainsi :

$$\begin{aligned} (p_-, p_0, p_+), \sum_c p_c = 1 \\ \equiv p_- = p(-1), p_0 = p(0), p_+ = p(+1) \end{aligned} \tag{2.2}$$

Ce lexique fournit donc le score de sentiment de chaque emoji sous la forme suivante :

$$\begin{aligned} \bar{s} = \sum_c p_c \cdot c \\ \equiv \bar{s} = -1 \cdot p_- + 0 \cdot p_0 + 1 \cdot p_+ = p_+ - p_- \end{aligned} \tag{2.3}$$

D'autres ressources sont connexes à l'analyse de sentiment : elles ont été dans un premier temps pensées pour améliorer l'analyse de sentiment via l'utilisation d'emojis. En 2011, aoki *et al.* (AOKI et UCHIDA, 2011) ont proposé une méthode permettant de combiner plusieurs informations émotionnelles d'un emoji dans un seul et même vecteur. Pour ce faire ils se sont fondés sur le modèle de Plutchik (PLUTCHIK, 1960) en considérant un vecteur d'emoji de taille 14, chaque donnée correspondant à 8 émotions basiques (surprise, tristesse, dégoût, colère, joie, anticipation, confiance et peur) ainsi que 4 émotions mixtes (amour, admiration, désapprobation, remord, mépris et optimisme) du modèle de Plutchik, chacune de ces émotions possédant son sous-ensemble de mots émotionnels.

1. Lexique disponible : http://kt.ijs.si/data/Emoji_sentiment_ranking/

Cette approche précurseuse de l'intégration d'informations d'émotions ou de polarité dans un vecteur d'emoji a plus récemment été remplacée par des approches automatiques fondées notamment sur les plongements lexicaux. C'est ainsi le cas d'*emoji2vec* (EISNER, ROCKTÄSCHEL et al., 2016), l'un des premiers modèles de plongements lexicaux d'emojis. Ce modèle fut appris à partir de 6 088 descriptions courtes accompagnant les 1 661 emojis normalisés par Unicode [2] en remplaçant chaque terme des descriptions par leur vecteur extrait du modèle de Google News *word2vec* (MIKOLOV, CHEN et al., 2013 ; MIKOLOV, SUTSKEVER et al., 2013), puis en faisant la somme de ces vecteurs pour représenter la description $v_j = \sum_k = 1^N w_k$. Ce vecteur $v_i$ concaténant chaque représentation des termes de la description est ensuite associé à un vecteur d'emoji $x_i$ mis à jour pendant l'entraînement, leur objectif étant de retrouver les emojis à partir de leur description en utilisant la fonction sigmoïde du produit scalaire des représentations $x_i$ et $v_j$ comme suit : $\sigma(x_i^T v_j)$. Ils ont obtenu un taux d'exactitude de 85,5%.

L'étude du sens inhérent à chaque emoji a également fait l'objet d'un lexique sémantique nommé EmojiNet [3]. Wijeratne *et al.* ont agrégé plusieurs ressources libres d'accès (voir section 2.2) afin de les relier avec les sens donnés par *The Emoji Dictionary* [4] avant de les compléter par leurs définitions détaillées issues de *BabelNet* [5] (NAVIGLI et PONZETTO, 2010). Concrètement, chacun des 2 389 emojis y est représenté par 8 informations distinctes : le code Unicode, le nom court Unicode, la description Unicode, un ensemble de mots clés, un ensemble d'images, un ensemble d'emojis liés, un ensemble de catégories et un ensemble de sens différents pour lequel l'emoji peut être utilisé. Pour évaluer ce lexique, une annotation manuelle avec des juges annotateurs humains est utilisée sur la tâche de désambiguïsation, obtenant un coefficient Kappa (FLEISS, COHEN et al., 1969) de 73,55%. Ce lexique a ensuite été utilisé pour mettre en place une mesure de similarité sémantique dédiée aux emojis (WIJERATNE, BALASURIYA et al., 2017) dans laquelle ils distinguent la description de l'emoji ("*A gun emoji, more precisely a pistol. A weapon that...*") de ses définitions et étiquettes sémantiques ("*gun, weapon, pistol, violence, revolver, handgun...*"). À l'aide de 10 annotateurs humains et d'un ensemble de 508 similarité d'emojis [6] ils l'appliquent à l'analyse de sentiment à partir de plongements lexicaux de Twitter et Google News.

2. http://unicode.org/emoji/charts/full-emoji-list.html
3. http://emojinet.knoesis.org/api.php
4. 'http://emojidictionary.emojifoundation.com/home.php?learn
5. https://babelnet.org/
6. http://emojinet.knoesis.org/emosim508.php

## 2.2. Ressources disponibles

Les ressources disponibles liées aux emojis sont variées et, dû à la particularité de l'objet traité, ne sont pas uniquement issus des travaux de recherche. Il convient d'énumérer les différentes ressources par type ainsi que leurs particularités.

### 2.2.1. Inventaires : normalisation et description

— Unicode Emoji List [7] : une liste exhaustive et à jour des emojis "standards". Cette norme définit et choisit les emojis les plus communs sur chaque plateforme afin de leur attribuer une description, des images associées aux principaux constructeurs, mais surtout un code et un caractère unicode unique. Il est important de noter que certains emojis naissent d'une combinaison de caractères unicode existants, utilisés pour le teint de peau ou d'autres attraits : un homme 👨 mais professeur 👨‍🏫. Cette liste possède également une catégorisation d'emoji à gros grain évolutive en fonction des différents ajouts.

— Emojipedia [8] : un inventaire avec moteur de recherche intégré et différentes définitions. Il possède la particularité d'indiquer le sens conformément attendu pour chaque emoji ainsi que, quelques fois, son contexte d'utilisation. Par exemple, 😃 est accompagné de la description suivante :

> A yellow face with smiling eyes and a broad, open smile, showing upper teeth and tongue on some platforms. Often conveys general happiness and good-natured amusement.

— The Emoji Dictionary [9] : un dictionnaire d'emojis alimenté librement par les internautes. Il permet de regrouper pour chaque emoji des exemples d'utilisation, un regroupement de définitions, en plus des adjectifs, verbes et noms propres associés. La qualité obtenue est variable comme le montrent les noms propres associés à 😂 visibles en figure 2.1.

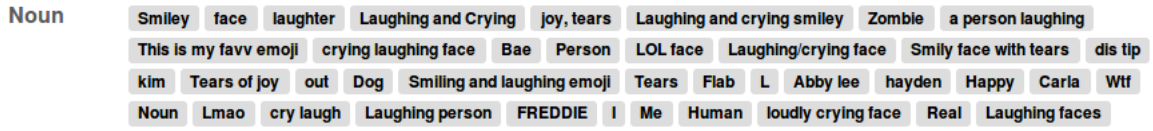


Figure 2.1. – Exemple de noms propres associés à l'emoji dans The Emoji Dictionary

7. http://unicode.org/emoji/charts/full-emoji-list.html
8. https://emojipedia.org/
9. https://emojidictionary.emojifoundation.com/

### 2.2.2. Claviers

— iEmoji [10] : un clavier d'emojis destiné aux ordinateurs et téléphones afin de comprendre comment ils sont utilisés dans les applications sociales. Ce système fournit une définition détaillée pour chaque emoji.
— Emoji Meanings [11] : un ensemble de définitions détaillées pour chaque emoji disponible sur les produits Facebook. Une application mobile en superposition de clavier est également disponible [12].

## 2.3. Modèles prédictifs

Des modèles prédictifs liés aux emojis ont vu le jour essentiellement à partir de 2017 et se distinguent par deux types d'approches : celles utilisant les emojis comme atout et celles dédiées aux emojis.

### 2.3.1. Emoji, sentiments et émotions

Les emojis les plus utilisés [13] sont liés aux sentiments que les utilisateurs veulent transmettre, c'est pourquoi un parallèle peut être fait entre leur utilisation dans les systèmes d'analyse de sentiment avec celles des émoticônes (HOGENBOOM, BAL et al., 2013), tous deux pouvant intensifier, exprimer ou contredire un sentiment actuel. Leur usage peut aussi s'inspirer des travaux de Jonathon Read (READ, 2005) sur la réduction de dépendance aux domaines dans l'analyse de sentiment à l'aide d'émoticônes. Mais le sentiment, ou plus précisément la polarité, n'est pas le seul cas d'utilisation des emojis pour améliorer le système de prédiction, Rao *et al.* (RAO, Q. LI et al., 2014) ont ainsi utilisé les emojis comme indicateurs de score lors d'une évaluation par utilisateurs, chaque emoji représentant ainsi une émotion parmi un ensemble de 8 émotions pré-définies : "touching", "empathy", "boredom", "anger", "amusement", "sadness", "suprise" and "warmness". Les émotions dérivées des emojis sont alors associées à un modèle de sentiment par domaine consistant, soit en une génération d'un jeu de sujets par mots suivant d'un échantillonage d'émotions par sujet, soit d'une génération de sujets à partir d'émotions.

Plus récemment, les emojis ont servi à l'amélioration de prédiction de mots-dièses (DHINGRA, ZHOU et al., 2016 ; VOSOUGHI, VIJAYARAGHAVAN et al., 2016), mots-dièses aussi liés aux domaines, sujets, qu'au sentiment inhérent au tweet.

10. https://www.iemoji.com/
11. https://www.emojimeanings.net/
12. https://play.google.com/store/apps/details?id=com.worldsapartsoftware.emojimeanings
13. http://emojitracker.com/

Les emojis ont également été utilisés pour améliorer la détection de sarcasme, l'analyse de sentiment et la classification par émotion (FELBO, MISLOVE et al., 2017) en prenant en compte les 68 emojis les plus courants dans plus de 1 246 millions de tweets. Pour ce faire, ils ont mis en place un réseau de neurones comprenant deux couches de Bi-LSTM (HUANG, W. XU et al., 2015), des couches récurrentes à double sens associées à un mécanisme d'attention. Ces travaux ont pu mettre en avant l'importance de la diversité des emojis pour chaque tâche évaluée.

### 2.3.2. Prédiction d'emoji

L'emoji est récemment passé du statut d'outil permettant d'améliorer une tâche connexe à objet principal de la prédiction. En 2016, Xie *et al.* ont ainsi cherché à recommander des emojis en réponse d'un flux de conversation. Ils ont donc classifié des conversations par 10 emojis possibles en représentant les dialogues par un encodeur utilisant un LSTM (HOCHREITER et SCHMIDHUBER, 1997) pour chaque réponse, suivi d'un LSTM hiérarchique (J. LI, LUONG et al., 2015) représentant l'ensemble de la conversation. Ce modèle appris sur 1 164 694 conversations de Weibo [14] a donné un taux de *Mean Reciprocal Rank* (MRR) (RADEV, QI et al., 2002) de 54,8% et une précision pour les 3 emojis les plus présents (P@3) de 65,7%. D'une manière similaire mais en se concentrant sur la prédiction d'emojis dans les tweets, Barbieri *et al.* (BARBIERI, BALLESTEROS et SAGGION, 2017) ont utilisé des Bi-LSTM (HUANG, W. XU et al., 2015) pour classer des tweets en 20 emojis possibles représentant les emojis les plus utilisés dans leur corpus de 40 millions de tweets. Ils ont obtenu une f-mesure de 34% pour les 20 emojis et 65% pour les 5 emojis les fréquents. En comparant les résultats de Xie (XIE, Z. LIU et al., 2016a) et Barbieri (BARBIERI, BALLESTEROS et SAGGION, 2017) avec les 48,8% d'exactitude obtenus par (FELBO, MISLOVE et al., 2017) pour prédire les 5 emojis les plus fréquents, la tâche de prédiction directe d'emojis n'est pas triviale. Ce qui s'est confirmé par une autre tentative de classification de tweets par 50 emojis les plus présents (ZHAO et ZENG, 2017) obtenant 45,8% en f-mesure et 40,3% en exactitude à partir de l'architecture CNN de Kim (KIM, 2014). Une autre approche de classification a été intentée par Li *et al.* (X. LI, YAN et al., 2017) en comparant les matrices représentant les emojis suites à des plongements lexicaux avec la représentation créée par la couche de CNN dans leur réseau, ils montrent que cette approche est efficace sans avoir recours à une dernière couche de softmax (BRIDLE, 1990). Ils ont obtenu 63,01% de précision pour les 5 emojis les plus fréquents (P@5) ainsi que 41,39% de MRR, allant dans le même sens que les autres études de prédiction d'emoji. La classification de tweets par emojis a également fait l'objet d'une tâche à SemEval2018 (BAR-

14. https://www.weibo.com/login.php/

BIERI, CAMACHO-COLLADOS et al., 2018) consistant à classer des tweets anglais et espagnols par 20 emojis et dont les f-mesures maximales ont été de 35,99% pour l'anglais et 22,36% pour l'espagnol.

La prédiction d'emoji a récemment évolué en prenant en compte plusieurs modalités, avec pour point de départ la mise en place de légende automatique d'images sous forme d'emojis (CAPPALLO, MENSINK et al., 2015) à l'aide de plongements sémantiques pour le texte et de CNN pour représenter l'image, les deux représentations étant alors comparées. D'une manière proche, Barbieri *et al.* (BARBIERI, BALLESTEROS, RONZANO et al., 2018) ont mis en place une prédiction de réaction à des images sous forme d'emojis en considérant les 20 emojis les plus fréquents. Les caractéristiques visuelles sont obtenues par réseaux de neurones résiduels (HE, X. ZHANG et al., 2016) tandis que les caractéristiques textuelles des emojis sont obtenues par FastText (JOULIN, GRAVE et al., 2016), obtenant une amélioration de 13,42% comparé à l'utilisation unique de caractéristiques textuelles ou visuelles. La multimodalité a également été utilisée pour résumer automatiquement une vidéo en un ensemble d'emojis (CAPPALLO, SVETLICHNAYA et al., 2019) ou encore pour prendre en compte les saisons pour influencer la prédiction d'emoji (BARBIERI, MARUJO et al., 2018).

## 2.4. Discussion

Dans ce chapitre, nous avons présenté les différents travaux sur les emojis. Cet état de l'art est constitué de travaux récents mettant en évidence l'émergence actuelle de recherches sur les emojis. Les premiers travaux ont principalement cherché à comprendre les emojis par le biais de travaux générateurs de ressources que nous séparons en deux parties : ceux dédiés à l'étude de similarité d'emojis, bien souvent à l'aide de plongements lexicaux, et ceux dédiés à l'obtention d'une meilleur définition sémantique pour chaque emojis à l'aide notamment d'auto-encodeurs, de ressources lexicales ou de plongements lexicaux. La prise en compte d'emojis dans des modèles de classification a ensuite été effectuée d'abord dans l'objectif d'une meilleure analyse du sentiment ou des émotions, avant d'avoir pour objectif unique la prédiction d'emojis en elle-même.

Les travaux actuels présentent plusieurs limites. Tout d'abord la recommandation des emojis n'est pas considérée puisque seule la prédiction d'emojis l'est par le biais d'une classification mono-étiquette. Les emojis prédits sont ensuite restreints aux emojis les plus fréquents, perdant alors tout sens dans la prédiction qui en suit pour une recommandation ultérieure, les emojis les plus fréquents étant souvent uniquement à polarité positive. Ces deux principales limites sont celles auxquelles nous souhaitons pallier.

Le travail présenté dans ce manuscrit propose une contribution multiple au traitement des emojis. La première contribution réside tout d'abord dans l'approche méthodologique de par les paramètres de plongements lexicaux des mots et des emojis, mais également de par la portée considérée pour la recommandation. L'objectif de construire une recommandation d'emojis n'a été abordé que par (Xie, Z. Liu et al., 2016a) et dans le travail présenté l'objectif n'est pas de classer les messages par emoji mais bien de proposer un panel d'emojis comme autant de choix à l'utilisateur. La seconde concerne la représentation des emojis, aussi bien dans leur considération comme de simples mots à part entière que dans la manière dont ils sont pris en compte techniquement : en tant que simple indice, représentation vectorielle de leur contexte compressé, ou méta-catégories. Enfin, la dernière principale contribution revient à la nature même du système de recommandation établi qui se veut être un système hybride entre prédiction et recommandation, et finalement assez éloigné des systèmes de recommandations habituels. Il s'agit d'un système sur-mesure répondant aux spécificités de la tâche.

## 2.5. Synthèse

**Domaine récent** : les travaux de recherche sur les emojis ont émergé en 2015.

**Compréhension des emojis** :
— plusieurs travaux se sont attelés à comprendre les emojis par l'étude de leurs usages et de leur impact social et conversationnel ;
— certains travaux essaient de comprendre les emojis en trouvant leur sémantique adéquate.

**Ressources sur les emojis** :
— issues de travaux sur la compréhension des emojis
— issues de modèles appris pour leur prédiction
— issues de sources diverses (sur-couche clavier, listes de standardisation, etc.)

**Prédiction d'emojis** :
— utilisée pour aider à d'autres tâches connexes (analyse de sentiment, etc) ;
— utilisée en tant qu'objectif en soi : nouvelle tâche émergente pour tester différentes approches.

**Limites actuelles** :
— il n'y a pas de véritables travaux sur la recommandation d'emojis. Uniquement la prédiction d'un seul emoji pour une phrase ;
— les travaux actuels sont souvent déconnectés de la réalité applicative (les tweets sont des données d'un type différent, aucune évaluation par l'utilisateur final, etc.) ;
— la prédiction d'emojis est souvent restreinte aux emojis les plus fréquents, créant un biais quant à la difficulté de la tâche et la constitution du jeu d'étiquettes.

**Contributions et différenciations** :
— la présente thèse se concentre sur la recommandation pour l'utilisateur ;
— la prédiction n'est pas restreinte à un seul emoji par message. Plusieurs emojis possibles sont considérés, individuellement ou par groupes ;
— l'évaluation du système se fait également par l'utilisateur réel et non uniquement par le fait de retrouver les emojis utilisés dans un corpus.

# 3. Prédiction automatique d'emojis

## Plan du chapitre


3.1 Introduction **57**
3.2 Substitution lexicale par approche générative **58**
3.2.1 Évaluation 64
3.2.2 Discussion 66
3.3 Enrichissement extra-linguistique par approche discriminante **67**
3.3.1 Corpus de messagerie sociale privée 67
3.3.2 Ensemble des caractéristiques 69
3.3.3 Classification multi-étiquettes pour la prédiction d'emojis 72
3.4 Limites de la prédiction directe d'emojis **81**
3.5 Conclusion **82**
3.6 Synthèse **84**


Les systèmes prédictifs peuvent être utilisés comme source de recommandation ou comme un des éléments permettant d'aller vers une recommandation automatique. Dans ce chapitre nous étudions des systèmes prédictifs d'emojis afin de comprendre s'ils sont intrinsèquement suffisants à une bonne recommandation d'emojis.

## 3.1. Introduction

La prédiction automatique d'emojis peut s'aborder de plusieurs manières : par une tâche de prédiction à l'aide de modèles génératifs (YE, X. LIU et al., 2012 ; CAI, C. ZHANG et al., 2007), ou par une classification à l'aide de modèles discriminants (voir section 2.3). Les modèles génératifs consistent à générer du contenu supplémentaire à partir du contenu existant. Appliqués sur du texte, de tels modèles peuvent être utiles pour prédire la fin d'un mot en cours de frappe ou le mot suivant en se basant sur le contenu existant de la séquence en cours. Ils peuvent ainsi être utilisés pour prédire l'emoji à un emplacement donné, toujours en fonction du contenu précédent et suivant. Ces modèles génératifs ont l'avantage d'être plus adaptés à une prédiction d'emojis effectuée constamment à chaque nouveau caractère inséré dans la séquence, puisqu'ils peuvent plus aisément prendre en compte des mots non finalisés, c'est-à-dire des mots en cours d'écriture.

En revanche, générer du nouveau contenu ne sied pas toujours à la tâche de prédiction d'emojis puisqu'il ne s'agit pas toujours de générer un nouveau contenu à l'emplacement précis, mais également de prédire un emoji en fonction

de la globalité de la séquence. C'est dans ce cas que se situe l'apport de l'utilisation de modèles discriminants qui, contrairement aux modèles génératifs, sont principalement utilisés pour prendre en compte toute la séquence afin de l'annoter, cette annotation prenant la forme d'un emoji. L'attribution d'emojis pour des séquences textuelles peut alors être utilisée pour prédire des emojis aussi bien en considérant l'ordre des mots ou non.

Le choix de l'approche est ainsi dépendant de la tâche que l'on cherche à accomplir mais, comme précisé précédemment, la prédiction d'emojis peut se faire de différentes manières. Nous décidons de considérer une approche différente en fonction du type d'usage de l'emoji que l'on souhaite capturer pour ensuite le prédire. Ces usages sont en lien avec les différentes fonctions qu'exercent les emojis dans une conversation (voir section 1.2 du chapitre 1), bien qu'elles ne les traitent pas toutes. Ainsi, il nous apparaît important de différencier l'approche utilisée afin de trouver la plus adéquate à chaque utilisation ciblée, les emojis remplaçant un mot servant souvent simplement de référentiel, nous utilisons l'approche générative pour cela (chapitre 3.2), là où un emoji final étiquetant le texte entier appartenant davantage au méta-linguistique sera alors traité par une approche discriminante (chapitre 3.3).

## 3.2. Substitution lexicale par approche générative

Les modèles génératifs pour la prédiction d'emojis peuvent être utilisés de manières différentes. Nous considérons ces modèles génératifs comme utiles à la prise en compte de l'utilisation de l'emoji pour sa fonction référentielle, c'est-à-dire lorsque l'emoji est utilisé à des fins de substitution lexicale. C'est par exemple le cas de la phrase "*J'ai une nouvelle* 📌", qui représente une utilisation de l'emoji à des fins de substitution lexicale, l'emoji 📌 remplaçant directement le mot "*écharpe*". Un tel usage ne prend souvent en compte que le contexte restreint du mot en cours, celui qui sera alors remplacé par l'emoji adéquat. De tels systèmes existent dans les applications de messageries existantes telles que Mood Messenger, Samsung, iMessage d'Apple, ou d'autres. Toutefois, ils se limitent bien souvent à l'utilisation d'un lexique de coréférences entre un vocabulaire donné et un emoji, sachant que plusieurs mots différents sont associés à un seul emoji. Dans un premier temps, notre objectif est d'obtenir une amélioration de tels systèmes par règles (voir chapitre 2) sans trop augmenter la complexité du processus de prédiction ; les modèles génératifs nous paraissent appropriés pour cela, et plus précisément les modèles de markov cachés, *Hidden Markov Models* (HMM). Ces derniers permettent, par leur système de transition entre états, de prendre en compte la génération du mot ou des caractères suivants dans un texte pour la complétion automatique de mots, en partant du principe que les états probables sont les mots de la phrase en cours d'écriture.

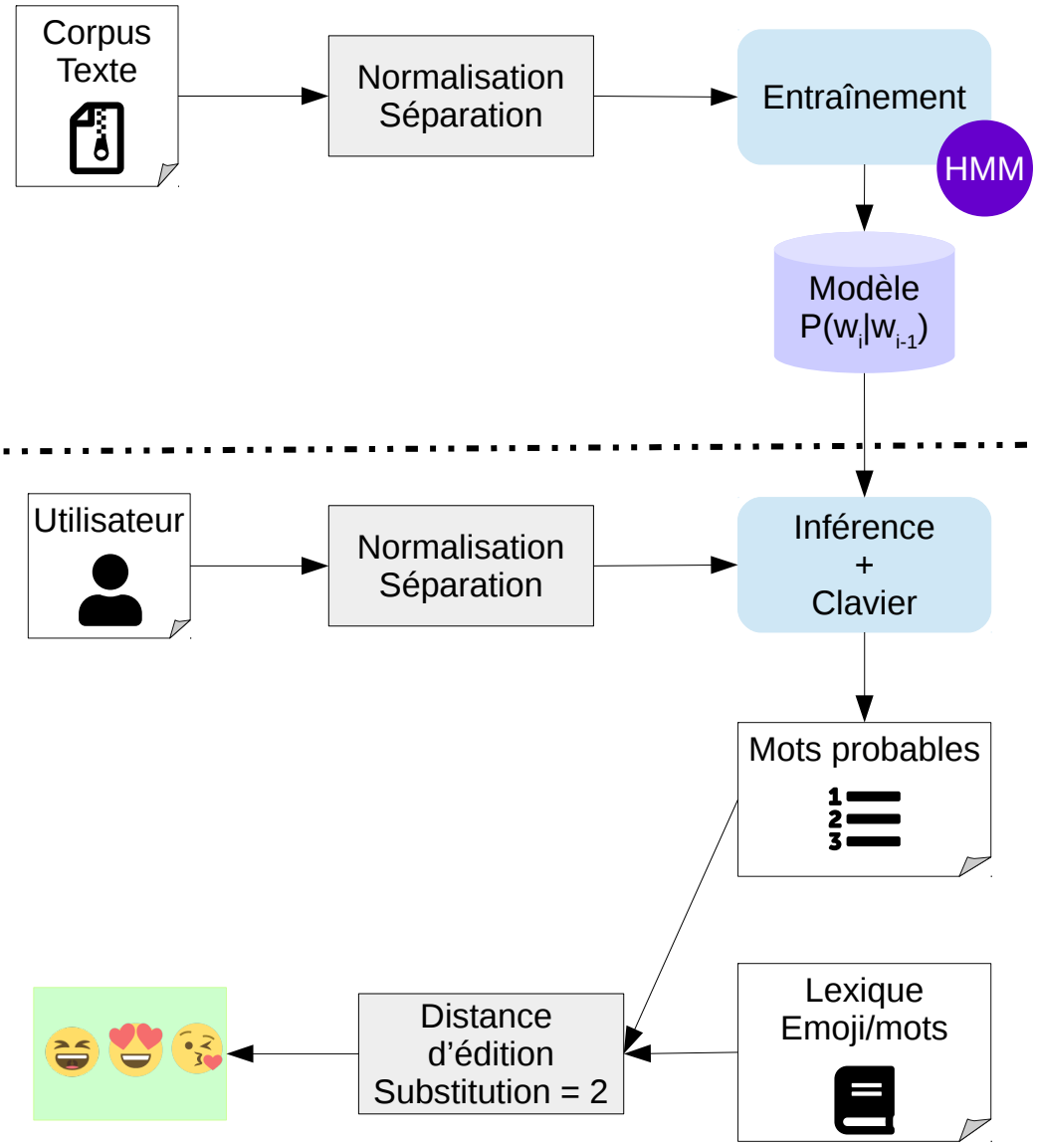


Figure 3.1. – Étapes de prédiction d'emojis selon le mot en cours et le mot précédent

Pour démarrer à partir de l'approche existante de Mood Messenger afin de l'améliorer, nous utilisons les HMM pour appliquer un système d'auto-complétion de mots, soit générer le mot suivant ou compléter le mot en cours. Ce modèle

suit les étapes visibles en Figure 3.1 et est appris à partir d'un corpus de textes [1] en anglais issus du domaine journalistique ou romanesque afin d'essayer de réaliser un apprentissage sur de la langue bien formée. Cela peut être considéré comme problématique étant donné que l'on traite dans le cadre de cette thèse des données de messageries instantanées privées dans lesquelles la formulation standard de la langue n'est pas toujours respectée. Toutefois, ce texte ne sert que de référent pour un lexique, il convient donc qu'il soit bien formé pour ne pas entraîner d'erreurs possibles dans l'auto-complétion de mots, et pour rendre son efficacité plus générique. Les caractéristiques de ce corpus sont présentées au Tableau 3.1.

| | |
|---|---|
| **Mots** | 1 105 834 |
| **Mots uniques** | 29 136 |
| **Lignes** | 128 457 |
| **Sources** | Littéraire, juridique, lexique, dictionnaire |
| **Langue** | Anglais |

Tableau 3.1. – Détails du corpus de texte utilisé pour entraîner les HMM

### Pré-traitements

Le corpus du tableau 3.1 est pré-traité en ignorant les mots vides issus de la liste de mots vides de Scikit-Learn, avant d'y effectuer une normalisation du texte en minuscules ainsi qu'une séparation des termes du corpus. Il convient de noter que la normalisation du texte n'est pas une lemmatisation puisque nous souhaitons également prendre en compte la variation morphologique des termes, leurs variantes se trouvant bien souvent dans des contextes différents. Aussi, la séparation des termes du corpus se fait par une expression régulière simple et ne saurait donc constituer une tokenisation. En voici la règle :

```
[a-z]+
```

### Entraînement

Pour obtenir un modèle de probabilité pour chaque transition possible observée dans le corpus d'entraînement, l'apprentissage du modèle se fait à partir du décompte du nombre d'occurrences de chaque terme associé à l'obtention du terme en cours et des termes suivants observés, ainsi que de leurs occurrences respectives. Le résultat de ces étapes est un dictionnaire ayant chaque terme comme clé et comme valeur tous les mots pouvant le suivre juste après ($N + 1$), ainsi que leur nombre d'apparitions observées dans ce contexte. Le but étant d'en

1. Le corpus d'entraînement utilisé est disponible ici : http://norvig.com/big.txt

inférer une probabilité par la suite lors de l'application du modèle, la phase de test. L'auto-complétion est effectuée en reprenant le principe de Rodrigo Palacios [2] combiné avec l'implémentation des chaînes de markov cachées (HMM) de Stephen Marsland [3] et de l'algorithme de Viterbi (FORNEY, 1973) pour définir le chemin le plus court à partir des séquences de mots observées dans le corpus d'entraînement du tableau 3.1.

### Prédiction

Lors de cette phase de test, le système d'auto-complétion prend uniquement en compte le contexte du mot précédent $N-1$ pour la prédiction du mot actuel $N$. De plus, la disposition des touches du clavier en fonction de la langue et du pays, soit QWERTY pour les États-Unis et AZERTY pour la France, est utilisée afin de prendre des mots possibles même en substituant le dernier caractère écrit du mot en cours par l'une des touches voisines du clavier. Finalement, la prédiction de mot par le HMM se résume ainsi à :

1. **Vérifier des préfixes de mots probables suivant le premier**. Le préfixe est fonction du dernier caractère du mot en cours substitué par une des lettres voisines dans la disposition du clavier. En un exemple concret, si l'on considère le mot précédent comme étant "*is*" et le mot actuel comme étant "*bad*", alors selon le clavier AZERTY, les lettres voisines sont *e, r, f, s, c* et *x*. Ceci permet alors de rechercher des mots possibles commençant par "*bad*", "*bae*", "*bar*", "*baf*", "*bas*" et "*bac*".
2. **Comparer les mots probables et leurs occurrences** dans ce contexte. Ainsi, si l'on conserve l'exemple précédent, on peut obtenir les mots probables "*badly*", "*bacterial*", "*barred*", "*back*", "*bad*", "*based*", "*bacteria*" ainsi que leurs nombres d'occurrences respectifs issus du corpus d'entraînement *'badly' (1), 'bacterial' (1), 'barred' (1), 'back' (1), 'bad' (10), 'based' (11), 'bacteria' (1)*.
3. **Sélectionner les $n$ mots les plus probables :** dans notre exemple cela reviendrait à sélectionner les mots ayant été aperçus le plus dans son contexte. Donc, pour $n = 2$, seuls les mots "*based*" et "*bad*" seraient choisis comme sortie de prédiction du HMM.

Le mot prédit est ensuite utilisé pour trouver l'emoji associé dans le lexique. Nous avons également pris en compte les deux mots précédents $N-1$ et $N-2$ par effet de cascade lors de la prédiction du mot actuel considérant son contexte précédent sans y voir une quelconque amélioration du système à partir de l'approche d'évaluation présentée ci-après (section 3.2.1). La dernière étape de la

2. https://github.com/rodricios/autocomplete
3. https://github.com/alexsosn/MarslandMLAlgo/blob/master/Ch16/HMM.py

prédiction de l'emoji pour le mot en cours consiste à utiliser la table de coréférences entre emoji et vocabulaire afin de transformer ces mots possibles en emojis.

Une fois la prédiction du mot suivant effectuée, nous appliquons une mesure de distance d'édition avec les entrées du lexique, les clés étant le vocabulaire, et la valeur étant l'emoji associé. Associée à la probabilité du mot en cours, cette distance d'édition permet de prendre en compte quelques défauts d'orthographe ou de variations, toutefois elle ne saurait être réellement efficace sans une lemmatisation du mot prédit. En effet, l'argot des réseaux sociaux, les abbréviations, ou encore les fautes de conjugaisons sont des obstacles à l'utilisation d'une distance d'édition. C'est pourquoi nous effectuons une lemmatisation sur le mot généré par les HMM afin d'éviter les problèmes de variations morphologiques dus, par exemple, à la conjugaison. À noter que cette lemmatisation ne peut être effectuée dans tous les cas, en effet, bien que la conjugaison soit sans impact, le pluriel se révèle être un élément déterminant pour certaines entrées du lexique. C'est par exemple le cas pour l'emoji 👭 pour lequel l'une des clés pourrait être "*girls*". Le résultat de cette lemmatisation sert alors de référence à l'utilisation de la distance d'édition sur le lexique.

| **Type** | **Code** | **Iso Code** | **TextEN** | **Image** |
|---|---|---|---|---|
| Common | Code propriétaire | 1f37a | beer, drink, alcohol, booze, etc. | 🍺 |

Tableau 3.2. – Format indicatif du lexique utilisé

Le lexique de correspondance entre vocabulaire et emojis utilisés suit le format visible en Table 3.2 auquel il convient d'ajouter une colonne de vocabulaire différente pour chaque langue, mais notre travail se concentre principalement sur l'anglais et le français. Ce lexique est issu d'une combinaison de plusieurs sources telles que des listes d'éléments d'un champ lexical donné, *etc.*, toutes collectées en amont par l'entreprise. Un tel lexique figé ne saurait constituer une base parfaitement exhaustive vu la diversité des termes possibles et la création constante de nouveaux termes, un groupe d'utilisateurs pouvant décider de se créer un lexique commun comme signe d'appartenance ou de distinction vis-à-vis d'une communauté tierce. Nous proposons d'autres d'améliorations en perspectives (section 3.2.2).

Pour exploiter ce lexique à l'aide d'une distance d'édition nous avons, dans un premier temps, considéré l'algorithme de Levenshtein (LEVENSHTEIN, 1966) qui

permet de calculer la distance entre deux chaînes de caractères à l'aide de la formule récursive suivante :

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{si } \min(i,j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{sinon.} \end{cases} \tag{3.1}$$

Cette formule calcule le nombre d'étapes minimales nécessaires pour transformer la chaîne de caractères $a$ en $b$ en itérant parallèlement sur chacun de leurs caractères, et pour ce faire inclut l'insertion $lev_{a,b}(i, j-1)+1$, la suppression $lev_{a,b}(i-1, j)+1$, et la substitution $lev_{a,b}(i-1, j-1)+1_{(a_i \neq b_j)}$ d'un élément dans la séquence de caractères, chacune de ces opérations ayant une valeur de 1. $(a_i \neq b_j)$ représente la valeur du coût de la substitution de caractères. Avec cette formule nous nous sommes rendu compte du manque de restriction au niveau de la substitution ; une substitution ayant un coût à 1 se révélait trop permissive lors de l'utilisation du lexique.

Pour illustrer ce constat, prenons pour exemple un lexique témoin avec le terme "*fun*" lié à l'emoji 😄 et le terme "*gun*" lié à l'emoji 🔫. Si l'on se réfère à l'algorithme initial de Levenshtein (LEVENSHTEIN, 1966), la distance entre les deux chaînes de caractères est égale à 1 par simple substitution du premier caractère "*f*" en "*g*", nous donnant ainsi une prédiction de l'emoji 🔫 pour le terme "*fun*", ce qui est loin du résultat escompté. C'est pour cette raison que nous avons décidé de modifier la valeur de la fonction de coût de substitution $(a_i \neq b_j) = 1$, en l'augmentant à 2, voire à 3, de telle sorte que la substitution devienne $(a_i \neq b_j) = 2$. En effet, lors de la comparaison du mot prédit par les HMM et le lexique de mots et d'emojis, nous imposons une limite stricte en permettant au lexique d'être utilisé uniquement en cas de différence minime entre un élément du vocabulaire et le mot prédit, c'est-à-dire une distance d'édition de 1. De cette manière nous évitons l'un des principaux inconvénients de l'utilisation d'un lexique qui est présent dans les modèles de prédictions d'emojis existants : la prédiction prématurée d'emojis non pertinents. En effet, cet inconvénient est réduit par l'apport de l'auto-complétion des mots effectuée en amont, et qui permet alors de prédire plus efficacement un emoji pour un mot en cours d'écriture en prenant en compte le contexte précédent. Associée à cette correspondance plus stricte du mot au lexique, la prédiction de l'emoji à l'emplacement immédiat en devient plus efficace et pertinente.

Pour ces raisons, la distance d'édition utilisée est donc celle de Levenshtein dans laquelle la pondération de la substitution est modifiée. Cela a pour effet

d'augmenter la distance entre deux chaînes de caractères liées par une substitution, contrairement à l'insertion et à la suppression.

### 3.2.1. Évaluation

La question de l'évaluation d'une telle approche n'est pas simple et l'idéal serait d'avoir une métrique permettant d'établir avec précision si l'emoji proposé à chaque moment de la frappe est pertinent ou non, cela afin d'obtenir une évaluation quantitative. Nous ne pouvons pas nous permettre d'utiliser n'importe quel emoji d'un corpus réel pour savoir si notre modèle le retrouve comme ce serait le cas dans l'exemple suivant :

> Vous avez fait du bon travail aujourd'hui 👍, merci.

Dans cet exemple, l'emoji n'est pas utilisé à des fins de remplacement de mot et sort donc du périmètre de ce que nous voulons traiter dans ce chapitre : la prédiction de l'emoji lors de son utilisation comme référentiel. Une évaluation quantitative partielle pourrait être obtenue à l'aide d'un indicateur informant si oui ou non l'emoji en question a été utilisé pour remplacer un mot complet ou en cours d'écriture mais, à notre connaissance, il n'existe pas de tel corpus disponible publiquement. Pour pallier ce problème, nous avons utilisé un corpus propriétaire et confidentiel de Caléa Solutions contenant un indicatif du remplacement d'un mot par emoji indiqué sous la forme visible dans l'exemple suivant :

> Dis-moi, nous allons prendre une {biè}_[emojiCode], tu te joins à nous ?

Le format indique donc les lettres tapées avant d'avoir choisi un emoji pour les remplacer ce qui nous permet, si l'on reprend cet exemple, d'obtenir les données de comparaison suivantes :

| Classe | Texte |
|---|---|
| 🍺 | Dis-moi, nous allons prendre une biè |

Tableau 3.3. – Format du corpus d'évaluation pour le modèle génératif

Le tableau 3.3 présente le format de base du corpus de test qui tend, de par la suppression du contenu présent après le mot initialement remplacé, à reproduire le contexte que l'utilisateur avait à l'instant T de son choix de sélection de l'emoji. Avec cette approche nous pouvons obtenir une évaluation quantitative du système de prédiction proposé et ainsi obtenir une indication de la qualité globale de ce dernier. Cependant, il convient de noter que cette méthode ne saurait donner une évaluation complètement exhaustive puis qu'elle possède quelques biais à prendre en compte :

1. L'évaluation se fait à partir des actions de l'utilisateur dans un contexte défini où ce dernier se voit déjà proposer une prédiction d'emojis à base de lexique, comme indiqué précédemment dans l'existant, même s'il peut également les sélectionner manuellement. Il est aussi important de noter qu'une telle évaluation ne prend pas en compte tous les cas où la recommandation de l'emoji est pertinente mais uniquement lorsque l'emoji a été utilisé, or un emoji non utilisé ne signifie pas nécessairement une mauvaise recommandation.
2. L'indication de remplacement n'est valable que pour le premier emoji choisi en remplacement, les emojis sélectionnés successivement pour remplacer un seul mot ne sont donc pas indiqués et ainsi ignorés. Cependant, il peut y avoir plusieurs remplacements au sein du message comme par exemple :

   Tu vas au {trava}_[🏢] en {vél}_[🚲] ?
3. Pour pouvoir au mieux évaluer le système, il est important d'utiliser le même lexique de correspondance vocabulaire-emojis.
4. La taille du corpus est restreinte aux données que nous avons pu obtenir (Table 3.4).

| | |
|---|---|
| **Mots** | 898 476 |
| **Contextes** | 220 048 |
| **Emojis possibles** | 1 368 |
| **Emojis différents présents** | 1 215 |

Tableau 3.4. – Données du corpus d'évaluation par reproduction du contexte précédant un emoji de remplacement

Finalement, nous avons obtenu des résultats mitigés d'un point de vue quantitatif. Le tableau 3.5 présente les moyennes des résultats obtenus :

| **Filtre** | **Précision** | **Rappel** | **F-mesure** |
|---|---|---|---|
| **1215 emojis** | 0.1439 | 0.2142 | 0.1721 |
| **Emojis présents min 1000x** | 0,2549 | 0,2523 | 0,2536 |
| **20 emojis les plus utilisés** | 0,2175 | 0,2420 | 0,2291 |
| **20 emojis les mieux prédits** | 0,8325 | 0,7180 | 0,7710 |

Tableau 3.5. – Résultats globaux en macro selon différents filtres

Dans le tableau 3.5, la première ligne ("1215 emojis") indique les scores macro bruts obtenus sur l'ensemble des 1 215 emojis présents dans le corpus d'évaluation des HMM. Ces scores sont très faibles et il en est de même pour les autres filtres : l'ensemble des emojis, les 20 emojis les plus utilisés ou encore les emojis présents au moins 1 000 fois. Ceci peut s'expliquer par plusieurs facteurs :

1. Beaucoup d'emojis (582 emojis sur 1 215) ont un f1-score de 0, ce qui pénalise fortement le calcul de macro f-mesure pour l'ensemble de la prédiction. Ceci représente 35 851 contextes (voir tableau 3.4) sur un total de 220 048.
2. La pluralité des contextes d'apparition possibles est trop grande.
3. Le corpus d'entraînement est générique et non spécialisé pour une utilisation dans le cadre de messages instantanés privés.

En regardant les 20 emojis les mieux prédits, on remarque que cette approche est en réalité sélective et fonctionne relativement bien sur un panel restreint d'emojis avec 39 emojis dépassant 50% de f-mesure pondérée par nombre d'occurrences. Par exemple, l'emoji 🏰 obtient 58% de f-mesure pour 138 occurrences. Parmi ces emojis bien prédits certains représentent tout de même des sentiments comme l'emoji propriétaire de Mood Messenger signifiant un acquiescement par hochement de tête animé  avec 83% de f-mesure pour 1 907 occurrences, ou encore 😾 avec 65% de f-mesure pour 1 173 occurrences. Ce constat confirme l'hypothèse de la nécessité d'une combinaison d'approches spécialisées par nature d'utilisation de l'emoji, par la fonction que l'emoji va appliquer. Cette approche générative semble convenir aux quelques cas de fonction référentielle du discours.

### 3.2.2. Discussion

Dans cette section, nous avons proposé un modèle de prédiction d'emojis fondée sur une approche générative à partir de HMM en combinant génération de mot avec un lexique agrégé de correspondances emojis-termes, la correspondance étant appliquée sur le mot prédit par distance d'édition. Cette approche sert de base de référence pour pouvoir comparer notre système avec ce qui existe actuellement dans les applications pour la suggestion d'emojis (BOJJA, KARUPPUSAMY et al., 2017), mais permet également de traiter différentes fonctions possibles de l'emoji (voir chapitre 1). La contribution apportée est donc l'application de modèles génératifs à la prédiction d'emojis par remplacement du contenu textuel prédit en emoji (*emojification*). La prédiction d'emojis par approche générative telle que montrée précédemment présente toutefois quelques limites. Le lexique tout d'abord, qui, bien qu'il soit le résultat de l'agrégation de plusieurs sources, se trouve limité vis-à-vis de l'ambition de la tâche. En effet, un lexique bien formé et conçu par agrégation de lexiques manuels ne peut

être complètement représentatif du champ lexical associé à un emoji, puisque certains emojis peuvent être utilisés à des fins dépassant le but initial de leur conception, c'est d'ailleurs de plus en plus fréquent avec les nouveaux emojis plus nuancés tels que 😳 et 🤯, *i.e.* "*mind blown*". La question se pose alors de savoir comment obtenir un lexique idéal pour la recommandation d'emojis de manière automatique, ce qui constitue la principale limite de notre approche.

La prédiction de l'emoji remplaçant un mot étant encore perfectible, nous nous concentrons désormais sur l'enrichissement extra-linguistique qu'il apporte à l'aide de modèles discriminants, expliqué en section suivante.

## 3.3. Enrichissement extra-linguistique par approche discriminante

Lorsqu'un message est envoyé accompagné d'un emoji, une des raisons d'utilisation de cet emoji est d'enrichir le contenu textuel par une information externe (voir chapitre 1). L'emoji vient donc ajouter de l'information, contradictoire ou non, vis à vis du message qui, dans des cas simples hors ironie, sarcasme ou information externe, peut se suffire à lui-même. Un tel message pourrait être "Il y a des macarons en vente demain." enrichi par l'emoji 😊 qui y ajoute alors un avis venant se greffer au contenu textuel, un enrichissement extra-linguistique. Dans cette partie, cet enrichissement est abordé au travers d'une approche discriminante, c'est-à-dire d'une approche visant à dissocier, classifier les contenus textuels contrairement à l'approche générative utilisée précédemment. Pour cela, la tâche de classification par apprentissage supervisé est envisagée et, plus précisément, une classification multi-étiquettes considérant plusieurs emojis possibles par contenu textuel (message).

### 3.3.1. Corpus de messagerie sociale privée

Avant d'aborder la prédiction d'emojis à l'aide de classification supervisée, il convient de définir le corpus que nous utilisons pour construire et appliquer ces modèles de classification. Ces corpus sont constitués de messages informels privés de langue anglaise dont certaines caractéristiques sont illustrées dans le tableau 3.7. Ils proviennent initialement d'un ensemble de 1 300 000 messages confidentiels de l'entreprise Caléa Solutions dont nous avons extrait uniquement les messages contenant des emojis pour permettre au classifieur supervisé d'apprendre les corrélations entre les emojis et les caractéristiques de chaque phrase. Ainsi pour le premier corpus ("corpus étendu" dans le Tableau 3.7) nous ne récupérons que les phrases qui contiennent des emojis, et pour le second unique-

ment les phrases qui contiennent des emojis sentimentaux (“corpus dédié” dans le Tableau 3.7). Par le terme “emojis sentimentaux” nous désignons les emojis représentant des sentiments (amour, joie, tristesse, *etc.*), et que nous distinguons des emojis d’objets, de concepts ou d’idées tels qu’une voiture, un drapeau ou un café par exemple. Le corpus est segmenté en phrases à l’aide du modèle anglais d’OpenNLP [4] (BALDRIDGE, 2005). Notez que cette approche peut présenter quelques limites sur des corpus de données informelles habituellement peu respectueuses de la ponctuation.

Le pré-traitement utilisé pour chaque données textuelle du corpus est constitué de plusieurs étapes :

— Suppression des mots vides à l’aide de la liste de mots vide de Scikit-Learn
— Lemmatisation et tokenisation : en utilisant les données de WordNet et NLTK. Chaque mot est lemmatisé en utilisant la fonction de variation morphologique inclue dans WordNet [5]).
— Vectorisation du texte à l’aide du TfIdfVectorizer de Scikit-learn dans lequel nous varions ensuite la portée des n-gram (1 à 5) ainsi que l’analyseur prenant en compte les mots ou les caractères
— Ajout des caractéristiques externes à chaque matrice représentant une phrase : identifiant de l’humeur, scores de polarité, nombre de mots, etc. (détaillés par la suite en section 3.3.2)
— Transformation des classes en matrices binaires représentant la présence ou l’absence de chaque classe pour chaque phrase afin de pouvoir mettre en place un étiquetage multi-étiquettes

Dans notre corpus, nous représentons chaque phrase par une paire {emojis | texte} permettant ainsi d’obtenir le texte sans les emojis et les emojis associés au message. Le tableau 3.6 montre un exemple d’une phrase du corpus et de ses emojis associés.

| Classe | Texte |
|---|---|
| 😢😔 | I heard about the news, it actually is quite depressing |

Tableau 3.6. – Exemple factice d’une phrase représentée par la paire emojis|texte

Dans le corpus, nous avons identifié 169 emojis sentimentaux [6] à partir de leur représentation (*i.e. son triplet de scores de polarité décrit ci-après*), calculés à par-

4. Modèle d’OpenNLP de découpage en phrases disponible ici : http://opennlp.sourceforge.net/models-1.5/

5. Code du lemmatiseur utilisé : https://www.nltk.org/_modules/nltk/stem/wordnet.html

6. https://gguibon.github.io/coria2017_data.html

tir de l'*Emoji Sentiment Ranking* (ESR) (KRALJ NOVAK, SMAILOVIĆ et al., 2015). L'ESR fournit les scores de polarité négative, neutre et positive pour 751 emojis à partir d'une annotation manuelle par 83 annotateurs de 1,6 million de tweets en contexte effectué pour 13 langues européennes. Ces emojis sentimentaux ont identifiés à l'aide des valeurs de sentiment (*sentiment score*) et de polarité associées à 751 emojis dans l'*Emoji Sentiment Ranking* (ESR) (KRALJ NOVAK, SMAILOVIĆ et al., 2015). Ainsi, l'emoji 😡 qui est représenté par le triplet {négatif ; neutre ; positif} suivant {0,532 ; 0,108 ; 0,360}, est porteur de sentiment. Ce qui n'est pas le cas pour l'emoji ✨ ({0,052 ; 0,545 ; 0,403}) dont la valeur neutre est supérieure aux autres. Bien entendu cet emoji ✨ pourrait être porteur de sentiment dans certains contextes, mais l'utilisation de l'ESR permet d'obtenir la valeur globale moyenne issue de nombreux contextes d'apparition pour chaque emoji.

| | **Corpus étendu** | **Corpus dédié** |
|---|---|---|
| **Nombre de phrases** | 8 8882 | 9 700 |
| **Mots** | 607 776 | 69 930 |
| **Emojis** | 148 928 | 18 384 |
| **Emojis distincts** | 1 070 | 164 |
| **Taux d'emojis sentimentaux** | 43.34% | 100% |
| **Nombre moyen d'emojis par phrase** | 1,68 | 1,90 |
| **Longueur moyenne des phrases** | 6 mots | 7 mots |
| **Phrases positives ssth*** | 5 832 | 1 014 |
| **Phrases négatives ssth*** | 0 | 0 |
| **Phrases positives echo**** | / | 1 532 |
| **Phrases neutres echo**** | / | 7 040 |
| **Phrases négatives echo**** | / | 1 128 |
| **Humeurs distinctes utilisées** | 38 | 38 |

Tableau 3.7. – Caractéristiques des deux corpus utilisés (l'un dédié aux emojis sentimentaux, l'autre étendu à tous les emojis). *valeurs prédites avec SentiStrength (Thelwall, Buckley et al., 2010). **valeurs prédites avec Echo (Hamdan, Bellot et al., 2015) uniquement pour le corpus dédié.

### 3.3.2. Ensemble des caractéristiques

Nous avons utilisé une représentation vectorielle des phrases du corpus. Cette représentation vectorielle peut varier de dimension en fonction des caractéris-

tiques considérées, nous avons donc évalué les performances des classifieurs en testant plusieurs combinaisons de caractéristiques. L'ensemble des caractéristiques disponibles est le suivant :

**Sac de mots/caractères et nombre de mots.** Le contenu textuel peut être représenté d'au moins deux façons différentes : par un sac de mots ou par un sac de caractères. Le nombre de mots contenus dans une phrase est également ajouté comme caractéristique. Ainsi la phrase "I love you" sera représentée comme {I} {love} {you} en sac de mots, et {I} {l} {o} {v} {e} {y} {o} {u} en sac de caractères.

**N-grammes.** Aux sacs de mots sont ajoutés les associations de mots/caractères. Des bi-grammes de lettres ou des bi-grammes de mots de la phrase "I love you" donneront ainsi {I+love} {love+you} en sac de mots, et {I+l} {l+o} {o+v} {v+e} {e+y} {y+o} {o+u} en sac de caractères. Nous utilisons un maximum de 5-grammes puisque au-delà l'amélioration des performances stagne (Tableau 3.10). Ces n-grammes permettent notamment de prendre en compte une syntaxe légère dans la phrase, *i.e.* une syntaxe limitée à l'association de mots. Ces associations de mots sont susceptibles d'être discriminantes pour un emoji donné.

**L'humeur.** Nous avons accès à l'humeur de l'utilisateur. L'utilisateur peut en effet sélectionner un état d'humeur et le changer quand il le souhaite via une interface de sélection sous forme d'emojis propriétaires visibles en Figure 3.2. Ainsi pour chaque message, et donc pour chaque phrase, nous avons accès à l'humeur affichée par l'utilisateur lors de la rédaction du message. Il y a 38 humeurs, chacune étant représentée par un emoji dans l'interface de sélection de l'humeur (bien, très bien, seul, triste, et ainsi de suite). Bien que cette donnée soit inédite, elle n'en reste pas nécessairement fiable puisque l'utilisateur peut très bien ne jamais modifier son humeur du moment. Toutefois une telle donnée étant liée au sentiment général de l'utilisateur, nous l'utilisons par la suite, d'autant plus que notre corpus est restreint aux utilisateurs assidus de l'application, c'est-à-dire ceux ayant enregistré un compte alors qu'il s'agit d'une étape facultative. L'impact de cette caractéristique est étudiée par la suite.

**Polarité de la phrase.** Pour chaque phrase, nous prédisons sa polarité afin d'obtenir de nouvelles caractéristiques représentatives du sentiment que peut traduire le ou les emojis à intégrer dans la représentation vectorielle d'une phrase donnée. Pour ce faire nous avons utilisé le logiciel SentiStrength[7] (THELWALL, BUCKLEY et al., 2010) avec le modèle pré-entraîné sur des commentaires de MySpace et des tweets. Il exploite des ressources lexicales propres à la grammaire, à l'orthographe de la communication en ligne (*slang words*, répétitions de

7. http://sentistrength.wlv.ac.uk/

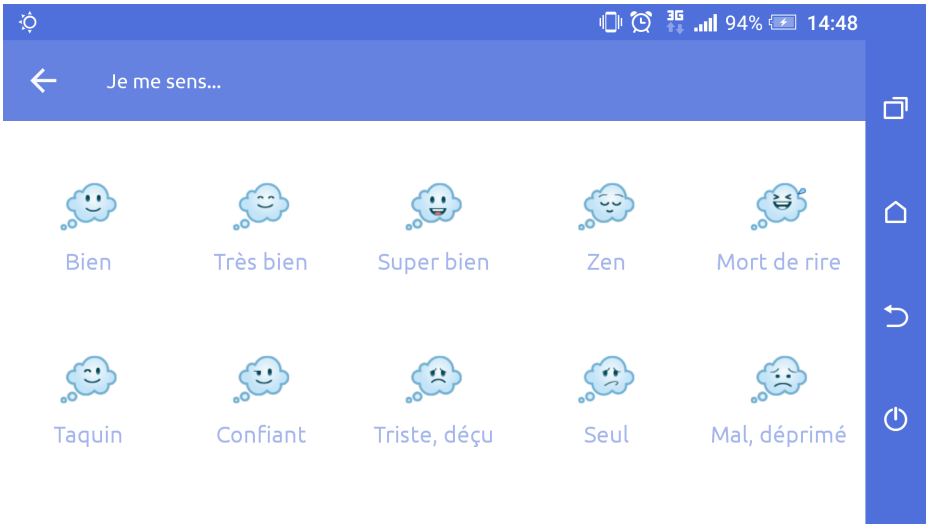


Figure 3.2. – Sélection de l'humeur (*mood*) par le biais de 38 emojis

caractères), et à la polarité pour prédire deux scores de polarité dans un texte. Un score positif, et un score négatif, tous deux allant d'une échelle de 1 (neutre) à 5 et de -1 (neutre) à -5. SentiStrength (THELWALL, BUCKLEY et al., 2010) est un des modèles d'analyse de sentiment les plus adaptés aux données que nous utilisons puisqu'il a été appris sur des messages informels courts non standardisés. A notre connaissance, il n'existe pas de modèle appris sur des messages courts privés, les tweets et commentaires publiques constituant actuellement les données d'apprentissage habituelles pour cette tâche. Pour le corpus dédié, nous avons également pris en compte Echo (HAMDAN, BELLOT et al., 2015) un analyseur de sentiment appris pour les tweets et qui a donné 1 532 phrases positives, 7 040 phrases neutres et 1 128 phrases négatives. L'utilisation combinée de SentiStrength (THELWALL, BUCKLEY et al., 2010) et Echo (HAMDAN, BELLOT et al., 2015) n'a cependant pas apporté d'améliorations dans la prédiction des emojis. Ainsi la prédiction de la polarité pourrait également être effectuée par d'autres outils comme l'analyseur de sentiments de Stanford (SOCHER, PERELYGIN et al., 2013), mais ce dernier nécessiterait de transformer le corpus en corpus arboré décrivant les informations syntaxiques. Ce modèle n'est toutefois pas spécifiquement adapté au jeu de données que nous avons puisque le micro-blogging est souvent limité en taille (Twitter), et destiné à être compris par tous donc plus à même de suivre un certain standard qu'un message privé.

**Présence de points d'interrogation ou d'exclamation.** Ces deux éléments de ponctuation sont souvent utilisés à des fins d'accentuation ("*Vraiment ? ? ?*"), d'autant plus dans les messages instantanés puisque leur utilisation n'est en rien obligatoire. Nous avons explicitement ajouté une caractéristique binaire permettant d'indiquer la présence d'un point d'interrogation dans la phrase ou d'un

point d'exclamation. Il convient toutefois de noter que nous n'avons pas pris en compte la répétition de ces signes, cela étant déjà pris en compte dans les modèles d'analyse de sentiment utilisés.

Dans cet ensemble de caractéristiques, l'usage de l'humeur et des scores d'intensité de sentiment de la phrase sont les plus importants vis-à-vis de notre hypothèse de départ selon laquelle les emojis sentimentaux seront mieux prédits avec des caractéristiques dédiées. Les autres caractéristiques servent principalement à représenter le contenu textuel.

### 3.3.3. Classification multi-étiquettes pour la prédiction d'emojis

Pour l'apprentissage des modèles nous avons utilisé les *ML-RandomForest* implémentés dans SciKit-Learn [8] avec un total de 20 arbres de décisions sans limiter leur profondeur, un nombre de caractéristiques maximum de $\sqrt{n_features}$ et une séparation des nœuds quand il y a au moins deux échantillons. Bien que différents en certains points, chaque modèle a été construit selon le protocole suivant :

1. Mélange aléatoire des phrases du corpus en question (corpus étendu ou dédié - Tableau 3.7 -)
2. Extraction des caractéristiques (*features*)
3. Création d'un jeu de test (30%) et d'entraînement (70%)
4. Entraînement d'un classifieur multi-étiquettes (*ML-RandomForest*)
5. Prédiction des étiquettes (*i.e.* emojis) sur le jeu de test
6. Évaluation par étiquette (*i.e.* emoji) et évaluation globale du classifieur

Ce protocole est répété 3 fois, lors d'une validation à trois plis stratifiés (*Stratified 3 fold Cross Validation*), pour chaque configuration afin d'éviter toute coïncidence de résultats et pour permettre de généraliser les performances du modèle autant que faire se peut. Les emojis présents sont ignorés lors de la prédiction sur un corpus de test.

#### 3.3.3.1. Prédiction des emojis sentimentaux dans un corpus dédié

Dans un premier temps nous essayons de prédire les emojis sentimentaux dans les phrases à l'aide d'un corpus réduit en classes possibles, *i.e.* un corpus dédié aux 169 emojis sentimentaux ("corpus dédié" au Tableau 3.7). Ces 169 emojis sentimentaux sont les classes possibles, les autres classes sont donc ignorées.

8. http://scikit-learn.org/

Ainsi, à partir du corpus d'origine nous créons un sous-corpus composé uniquement de phrases contenant des emojis sentimentaux parmi les 169 sélectionnés (voir section 3.3.1), et nous ne nous limitons donc pas aux dix emojis les plus utilisés comme Xie *et al.* (Xie, Z. Liu et al., 2016b). Les caractéristiques de ce corpus sont indiquées dans le tableau 3.7.

| | **Exactitude** | **Précision** | **Rappel** | **Moyenne harmonique** |
|---|---|---|---|---|
| **Caractéristiques** | *Lemmatisation*, 1 à 5-grammes, tf-idf* | | | |
| **Sac de mots** | 48,17 | 89,65 | 50,93 | 64,04 |
| **Sac de caractères** | 55,12 | 92,19 | 57,63 | 70,13 |
| **Caractéristiques** | *Lemmatisation*, 1 à 5-grammes, nombre de mots, tf-idf, score négatif, score positif, exclamation, interrogation* | | | |
| **Sac de mots** | 49,94 | 90,65 | 52,49 | 65,63 |
| **Sac de caractères** | 55,31 | 92,3 | 57,83 | 70,31 |
| **Caractéristiques** | *Lemmatisation*, 1 à 5-grammes, humeur, nombre de mots, tf-idf, score négatif, score positif, exclamation, interrogation* | | | |
| **Sac de mots** | 54,31 | 88,63 | 58,13 | 69,56 |
| **Sac de caractères** | **60,28** | **94,30** | **62,52** | **74,49** |
| **Caractéristiques** | *Lemmatisation*, 5-grammes, humeur, nombre de mots, tf-idf, score négatif, score positif, exclamation, interrogation* | | | |
| **Sac de mots** | 23,97 | 79,44 | 25,74 | 37,24 |
| **Sac de caractères** | 57,22 | 93,95 | 59,37 | 72,00 |

Tableau 3.8. – Moyennes des performances de prédiction d'emojis sentimentaux avec *ML-Random Forest*. *La lemmatisation est ignorée pour les sacs de caractères.

En appliquant le protocole détaillé en début de section 3.3.3 nous obtenons les résultats visibles dans le tableau 3.8. L'entraînement est effectué sur 70% du corpus dédié aux emojis sentimentaux (6790 phrases) et le test effectué sur les 30% restants (2910 phrases), le tout en 3 itérations afin d'obtenir les moyennes de scores visibles au tableau 3.8.

**Méthode d'évaluation.** Dans ces résultats il convient de noter que l'exactitude (*accuracy*) correspond ici à la pertinence moyenne des emojis et non à celle du jeu d'étiquettes complet (*Powerset*). Ainsi, si une phrase est étiquetée 😍 et 😘, chaque emoji est considéré séparément. Nous ne faisons donc pas d'évaluation par *PowerSet*, c'est-à-dire l'ensemble des sous-ensembles de combinaisons possibles pour un jeu d'étiquettes donné. Par exemple, puisque nous n'utilisons une évaluation discriminante, la phrase donnée précédemment (Tableau 3.6) pour laquelle seul l'emoji 😓 a été prédit alors que deux emojis étaient attendus, ne verra que l'exactitude de l'emoji correctement prédit augmenter.

L'exactitude affichée dans nos résultats correspond donc à la moyenne des exactitudes de chaque emoji, pondérées par leur fréquence dans le corpus de test. Ainsi, la classe 😓 verra son exactitude augmenter, mais pas la classe 😔. L'exactitude globale étant ensuite la somme des exactitudes de chaque classe, pondérées par leur fréquence d'apparition. Cette évaluation permet d'obtenir une évaluation par classe, généralisée par la suite, et est également appliquée aux métriques de précision et de rappel.

Le tableau 3.8 suit une logique d'incrémentation des caractéristiques, et représente l'approche de classification multi-étiquettes par adaptation, la classification multi-étiquettes pouvant être abordée en plusieurs problèmes mono-étiquette consistant à utiliser un classifieur pour chaque classe possible (approche par transformation) ou bien en modifiant directement l'algorithme de classification utilisé pour l'adapter à une classification multi-étiquettes (approche par adaptation). L'approche par transformation est ici trop coûteuse puisqu'elle reviendrait à mettre en place 164 classifieurs de *Random Forest*, multiplié par le nombre d'estimateurs choisis au sein de chaque classifieur, entraînant une augmentation du temps de calcul. De plus, cette approche est moins performante qu'une approche par adaptation (*ML-RandomForest*) : avec normalisation tf-idf, et 1 à 5 grammes de caractères, les résultats étant de 91,83% contre 92,19% en précision, 57,37% contre 57,63% en rappel, et 69,8% contre 70,13 en f-mesure, soit légèrement moins efficace que ceux obtenus avec un multi-étiquetage par adaptation montrés dans le tableau 3.8 pour les mêmes données et les mêmes caractéristiques utilisées.

Les modèles de multi-étiquetage présentés dans le tableau 3.8 permettent de prédire les emojis d'une phrase avec une f-mesure maximale de 74,49%, pour une bonne précision de 94,30%, mais un rappel de seulement 62,52%. Cette meilleure prédiction est obtenue en considérant les sacs de caractères avec des n-grammes allant de 1 à 5 caractères, l'humeur de l'utilisateur lors de l'écriture du message, le nombre de mots total de la phrase, les scores de polarité négatifs et positifs prédits, la présence d'un point d'exclamation et celle d'un point d'in-

terrogation. Le tout est lemmatisé en amont, les composants des vecteurs sont ensuite pondérés par tf-idf.

**Analyse de l'impact de la polarité.** Les résultats montrent qu'une fois les scores de polarités négatives et positives ajoutés comme caractéristiques, les résultats du classifieur n'obtiennent pas un gain notable ; et ce bien qu'elles soient accompagnées du nombre de mots et de la présence ou non d'un point d'exclamation et d'interrogation. Ce résultat tend à montrer que l'analyse de sentiment en utilisant SentiStrength[7] (THELWALL, BUCKLEY et al., 2010) n'est pas pertinente pour prédire les emojis sentimentaux. Dans la section 3.3.3.2, nous proposons d'explorer la prédiction d'emojis dans le corpus étendu (Tableau 3.7), en nous basant sur l'hypothèse selon laquelle la prédiction d'emojis sentimentaux serait plus efficace en utilisant une approche spécifique.

**Analyse de l'impact des sacs de caractères/mots.** Le tableau 3.8 réfute l'idée d'associer systématiquement un mot avec un emoji. Avec les sacs de mots, les mots présents dans les représentations vectorielles de deux phrases sont comparés, ce qui nécessite qu'ils soient identiques, même après lemmatisation. La lemmatisation est effectuée à l'aide du lemmatiseur appris sur les synsets de *WordNet* disponible dans NLTK[9]. Avec l'approche par sac de caractères plusieurs facteurs entrent en compte, la langue non standardisée et les abréviations en sont un premier facteur qui sera traité de mnaière implicite. Par exemple, le "you" aura au moins un point commun avec son abbréviation "u", ce qui n'est pas le cas dans un sac de mots. Bien entendu, d'autres mots auront ainsi un point commun avec "you", mais ceci explique probablement en partie la nette hausse de performance lors de l'utilisation de sacs de caractères (Tableau 3.8), d'autant plus que la langue est potentiellement moins standardisée dans des corpus de messages privés (messagerie instantanée) que publics (tweets). Le second facteur est le possible bi-linguisme, des mots en espagnol, français, arabe romanisé peuvent s'y glisser et induire l'utilisation ou non d'un emoji. En effet, bien que la langue de chaque phrase soit détectée, s'agissant de messages courants, il est par exemple possible de retrouver de l'espagnol en début de message tel que "*hola, I was wondering about what you did yesterday...*". Enfin, le troisième facteur - la longueur des phrases - permet de comprendre pourquoi dans certains cas il est impensable d'associer un emoji à un sac de mots. En effet, même si la longueur moyenne des phrases détectées est de 7 mots (tableau 3.7), un message entier peut représenter un paragraphe sans aucun signe de ponctuation, ou à l'inverse n'être composé que d'une seule lettre (exemple : "😍 u"). Sachant que la segmentation par phrase est effectuée automatiquement, toutes ces variations peuvent la rendre plus ou moins fiable, et ainsi remettre en question cette donnée.

9. http://www.nltk.org/

**Analyse de l'impact de l'humeur.** L'humeur a un impact important sur la qualité de la prédiction avec un gain constant d'environ 4% en f-mesure sur le meilleur score du tableau 3.8. L'humeur est propre au corpus utilisé, et semble particulièrement représentative du sentiment de l'utilisateur qui correspond à des emojis. Bien qu'elle soit très influente sur la qualité de la prédiction, l'humeur à elle seule ne peut suffire à prédire les emojis, le contenu textuel et sa représentation (*i.e.* sac de caractères/mots) doivent y être associés. Les corpus utilisés montrent ainsi que les emojis varient principalement selon le texte et l'humeur, il apparaît donc nécessaire de quantifier précisément son influence, ce à quoi nous nous attelons par la suite.

| **Ensemble** | **1 à 5-grams de caractères, nombre de mots tf-idf, score positif, score négatif, exclamation, interrogation** | | | | |
|---|---|---|---|---|---|
| Rang d'importance | **1er** | **2ème** | **3ème** | **4ème** | **5ème** |
| Caractéristique | Espace | Nombre mots | "f h" | "o" | Score positif |
| Score d'importance | 0.0063 | 0.0053 | 0.0034 | 0.0032 | 0.0032 |
| **Ensemble** | **1 à 5-grams de caractères, humeur, nombre de mots tf-idf, score positif, score négatif, exclamation, interrogation** | | | | |
| Rang d'importance | **1er** | **2ème** | **3ème** | **4ème** | **5ème** |
| Caractéristique | Humeur | Nombre mots | Espace | "ways " | "u" |
| Score d'importance | 0,0602 | 0,0051 | 0,0045 | 0,0031 | 0,0029 |
| **Ensemble** | **Lemmatisation, 1 à 5-grams de mots, humeur, nombre de mots tf-idf, score positif, score négatif, exclamation, interrogation** | | | | |
| Rang d'importance | **1er** | **2ème** | **3ème** | **4ème** | **5ème** |
| Caractéristique | Humeur | Nombre mots | "u" | "love" | "." |
| Score d'importance | 0,0997 | 0,0461 | 0,0141 | 0,01368 | 0,01136 |
| **Ensemble** | **1-grams de caractères, humeur, nombre de mots, tf-idf, score négatif, score positif, exclamation, interrogation** | | | | |
| Rang d'importance | **1er** | **2ème** | **3ème** | **4ème** | **5ème** |
| Caractéristique | Humeur | Espace | "o" | Nombre mots | "i" |
| Score d'importance | 0.0943 | 0.0514 | 0.0453 | 0.0430 | 0.0411 |

Tableau 3.9. – Les cinq caractéristiques discriminantes selon les scores d'importance des *RandomForest*. Avec différents ensembles de caractéristiques utilisées.

**Caractéristiques discriminantes.** Pour analyser l'importance de chaque caractéristique, nous comparons le score d'importance calculé par le *ML-RandomForest* pour chaque test (Tableau 3.9). Ces scores conférés après entraînement du classifieur confirment l'importance de l'humeur dans la prédiction des emojis. Seules cette caractéristique et celle du nombre de mots semblent avoir des scores élevés. Étant fondées sur des sacs de caractères, les caractéristiques peuvent paraître souvent difficilement intelligibles (ex :"ways" en 4ème position dans le tableau 3.9). Pour une approche par sacs de mots avec humeur et polarités, les 5 caractéristiques les plus importantes sont l'humeur, le nombre de mots, le mot "u" (you), le mot "love" et le point (Tableau 3.8). Ces scores d'importance sont fonction du nombre total de caractéristiques retenues, d'où leur score relativement bas. En effet il convient de préciser qu'en raison de l'utilisation de différents n-grammes, le nombre de caractéristiques peut être très élevé et ainsi aller de 123 caractéristiques avec 1-gramme à 61268 caractéristiques en utilisant des n-grammes de 1 à 5 grammes. Ceci impacte directement la taille de la représentation vectorielle de la phrase ainsi que la valeur d'importance attribuée, sans pour autant changer la hiérarchie des caractéristiques selon leurs scores d'importance du tableau 3.9.

| Caractéristique | Exactitude | Précision | Rappel | F-mesure |
|---|---|---|---|---|
| ***Baseline*** | 53,64 | 92,26 | 56,02 | 68,93 |
| **1 à 2 grammes** | +1,05 | +0,003 | +1,12 | +0,87 |
| **1 à 3 grammes** | +1,10 | -0,16 | +1,3 | +0,93 |
| **1 à 4 grammes** | +1,21 | -0,27 | +1,41 | +0,98 |
| **1 à 5 grammes** | +1,02 | +0,04 | +1,1 | +0,82 |
| **1 à 6 grammes** | +0,99 | -0,14 | +1,12 | +0,80 |
| **Humeur** | **+5,59** | **+2,79** | **+5,04** | **+4,74** |
| **Nombre de mots** | +0,0 | +0,10 | -0,04 | +0,0 |
| **Score positif*** | +0,34 | +0,08 | +0,37 | +0,30 |
| **Score négatif*** | +0,22 | +0,16 | +0,18 | +0,19 |
| **Scores positif et négatif** | +0,2 | +0,01 | +0,25 | +0,19 |
| **Exclamation** | +0,02 | -0,06 | +0,04 | +0,01 |
| **Interrogation** | +0,03 | -0,06 | +0,04 | +0,02 |

Tableau 3.10. – Impact empirique de chaque caractéristique par rapport à la "*baseline*" constituée d'un sac de caractères (moyennes obtenues pour 3 exécutions). *Scores prédits avec SentiStrength

La grande majorité de ces caractéristiques sont peu discriminantes si on en croit le tableau 3.10 mais influencent les résultats une fois regroupées comme c'est le cas au tableau 3.8. Bien que les résultats visibles dans les meilleurs scores du tableau 3.8 montrent que celles-ci ne sont pas déterminantes, nous avons cherché à vérifier cette hypothèse ainsi que les scores d'importance par une quantification de l'impact de chacune d'elles.

Dans le tableau 3.9, nous avons analysé le score d'importance des caractéristiques en considérant les caractéristiques ensemble. Nous complétons cette analyse par une étude de l'impact de chaque caractéristique prise isolément (Tableau 3.10). Par exemple, la ligne de l'humeur représente les performances de classification en considérant uniquement le sac de caractères et l'humeur. Les résultats sur cette ligne confirment l'importance de cette caractéristique et l'influence qu'elle peut avoir sur la performance de la prédiction. Mais également que les utilisateurs en font une utilisation pertinente pour le classifieur. Le tableau 3.10 permet ainsi de voir l'impact des caractéristiques isolées. Il s'agit donc d'une approche empirique dans laquelle la *baseline* est uniquement composée d'un sac de caractères.

Nous observons qu'augmenter les n-grammes semble apporter un peu de performance mais augmente énormément le nombre de caractéristiques créées. Ces n-grammes sont appliqués à l'aide du *TfidfVectorizer* de Scikit, et concernent la totalité des caractères y compris les espaces et ponctuations, ce qui explique pourquoi le point se retrouve présent dans le tableau 3.9. Aussi, encore une fois l'humeur est confirmée comme étant la caractéristique la plus discriminante.

Enfin, il est intéressant de constater que contrairement aux scores d'importance attribués par le classifieur aux caractéristiques utilisées, le nombre de mots ne semble ici n'avoir aucun importance. Le score de polarité positive a seulement un impact minime mais tout de même plus élevé que le score de polarité négative, ce qui s'explique par le fait que ce dernier ne varie jamais. Ceci constitue une limite de l'utilisation de SentiStrength (THELWALL, BUCKLEY et al., 2010) sur notre corpus.

Dans cette section, nous avons construit un modèle de classification multi-étiquettes pour prédire les emojis sentimentaux. Dans la section suivante, nous proposons d'explorer cette même méthode pour construire un modèle capable de prédire tout type d'emoji.

#### 3.3.3.2. Prédiction des emojis dans un corpus étendu

Pour construire un modèle de prédiction d'emojis de tout type, nous avons utilisé la même approche sur un corpus uniquement composé de phrases avec emojis (Tableau 3.7), qu'ils soient sentimentaux ou non. Ce corpus est de taille plus grande mais le nombre de classes possibles augmente également. Les 88882 phrases peuvent désormais appartenir à 1070 classes (emojis) différentes. Notre corpus ne contenant pas tous les emojis standards, et se limitant à ceux réellement utilisés et présents dans notre jeu de données, nous avons 1070 emojis au total et non les 2389 emojis d'Unicode. Aussi, les 164 emojis sentimentaux

identifiés représentent 43.34% des emojis utilisés, les 906 emojis restants représentant 56.66% des utilisations d'emojis. La représentation des classes n'est donc pas équilibrée dans le corpus. Compte tenu des limites d'obtention de données réelles, et du déséquilibre de l'utilisation des emojis (PAVALANATHAN et EISENSTEIN, 2015) déjà visible sur Twitter par le biais de l'EmojiTracker[10], il est très difficile d'équilibrer les classes utilisées. Le sur-échantillonage possède l'inconvénient majeur de rendre la distribution des données non réaliste puisqu'il ne peut tenir en compte des différents mots possibles associés aux emojis, en se limitant à reproduire une même association. Le vocabulaire concerné en devient donc plus important sans pour autant être certain de la corrélation entre les termes utilisés et l'emoji, puisque l'usage de l'emoji peut provenir d'un contexte externe. D'un autre côté, le sous-échantillonage nous prive d'informations importantes relatives aux différentes associations possibles entre mots et emojis.

| | **Exactitude** | **Précision** | **Rappel** | **F-mesure** |
|---|---|---|---|---|
| **Caractéristiques** | *1 à 5-grammes, tf-idf* | | | |
| **Sac de caractères** | 60,44 | 94,78 | 62,07 | 73,76 |
| **Caractéristiques** | *1 à 5-grammes, nombre de mots, tf-idf, score négatif score positif, exclamation, interrogation* | | | |
| **Sac de caractères** | 60,48 | 94,76 | 62,11 | 73,79 |
| **Caractéristiques** | *1 à 5-grammes, humeur, nombre de mots, tf-idf score négatif, score positif, exclamation, interrogation* | | | |
| **Sac de caractères** | 63,91 | 94,66 | 65,86 | 76,63 |
| **Caractéristiques** | *humeur, nombre de mots, tf-idf, score négatif score positif, exclamation, interrogation* | | | |
| **Sac de caractères** | 61,31 | 94,28 | 63,28 | 74,79 |
| **Caractéristiques** | *5-grams de lettres, humeur, nombre de mots, tf-idf score négatif, score positif, exclamation, interrogation* | | | |
| **Sac de caractères** | 62,30 | 95,94 | 63,62 | 75,32 |

Tableau 3.11. – Moyennes des prédictions d'emojis sentimentaux avec Random Forest et sac de caractères

Comme précédemment le tableau 3.11 montre les moyennes de trois exécutions aléatoires entrainées sur 70% du corpus (soit 62217 phrases) et testées sur les 30% restant (soit 26665 phrases). Toutes les étiquettes d'humeur sont présentes en entraînement et en test. Les résultats globaux présentent une légère

10. http://www.emojitracker.com/

amélioration mais restent similaires à ceux obtenus sur le corpus dédié aux emojis sentimentaux. Il y a toujours un fort impact de l'humeur dans la prédiction des emojis. En effet, si l'on en croit les travaux réalisés sur les emojis de Twitter, avec notamment l'EmojiTracker[10] et l'EmojiSentimentRanking (KRALJ NOVAK, SMAILOVIĆ et al., 2015), ainsi que l'importance (43.34%) des emojis sentimentaux dans notre corpus, les emojis les plus utilisés sont ceux porteurs de sentiment. Il n'est donc pas étonnant de retrouver un impact fort de l'humeur ; impact confirmé par les scores des caractéristiques visibles dans le tableau 3.12.

| **1 à 5-grams de caractères, humeur, nombre de mots tf-idf, score positif, score négatif, exclamation, interrogation** | | | | | |
|---|---|---|---|---|---|
| Rang d'importance | **1er** | **2ème** | **3ème** | **4ème** | **5ème** |
| Caractéristique | Humeur | "lex " | "alex" | " al" | "x " |
| Score | 0,0515 | 0,0059 | 0,0058 | 0,0058 | 0,0058 |

| **Lemmatisation, 1 à 5-grams de mots, humeur, nombre de mots tf-idf, score positif, score négatif, exclamation, interrogation** | | | | | |
|---|---|---|---|---|---|
| Rang d'importance | **1er** | **2ème** | **3ème** | **4ème** | **5ème** |
| Caractéristique | Humeur | "alex" | Nombre mots | "simply" | "pugla*" |
| Score | 0,0688 | 0,0408 | 0,0267 | 0,01481 | 0,0049 |

Tableau 3.12. – Les cinq caractéristiques discriminantes calculées par les *RandomForest*, avec sac de mots ou de caractères. Pour ce dernier, les scores stagnent à 0.0058 au delà de la 5e position.

Pour confirmer cela nous comparons les performances du même modèle sur les deux ensembles d'emojis : ceux porteurs de sentiment et les autres (Tableau 3.13). Que ce soit avec ou sans l'intégration de l'humeur de l'utilisateur, la prédiction est plus performante sur les autres emojis, ceux qui ne font pas partie des 169 emojis sentimentaux.

Tous ces résultats sur le corpus étendu et l'écart ténu entre les performances infirment l'hypothèse selon laquelle, parce qu'ils sont vecteurs d'émotions, les emojis sentimentaux seraient mieux prédits à l'aide d'un modèle de classification dédié. Par exemple, l'emoji neutre (selon les polarités de l'ESR) 😴 obtient une f-mesure de 81% et un taux d'exactitude de 68%. Le corpus étant de plus grande taille, les résultats sont légèrement plus élevés, mais lorsque l'humeur est prise en compte les performances sont meilleures pour les emojis non sentimentaux, notamment en précision. Toutefois, elles restent similaires et ne justifient pas une prise en compte spécifique des emojis porteurs de sentiment lors de leur prédiction.

| | Exactitude | Précision | Rappel | F-mesure |
|---|---|---|---|---|
| **Caractéristiques** | *1 à 5-grammes, nombre de mots, tf-idf, score négatif, score positif, exclamation, interrogation* | | | |
| **Emojis sentimentaux** | 59,69 | 94,12 | 61,71 | 73,69 |
| **Emojis autres** | 61,69 | 95,49 | 62,96 | 74,32 |
| **Caractéristiques** | *1 à 5-grammes, humeur, nombre de mots, tf-idf, score négatif, score positif, exclamation, interrogation* | | | |
| **Emojis sentimentaux** | 63,22 | 93,33 | 65,93 | 76,60 |
| **Emojis autres** | 64,70 | 95,31 | 66,26 | 76,84 |

Tableau 3.13. – Comparaison des performances d'un même modèle général sur les deux jeux d'étiquettes (moyennes de 3 exécutions aux découpages aléatoires mais comparables)

## 3.4. Limites de la prédiction directe d'emojis

L'approche proposée dans ce chapitre consiste à prédire des emojis par classification multi-étiquettes à l'aide de caractéristiques discriminantes orientées vers le sentiment et l'humeur. Cette approche possède toutefois plusieurs limites :

**Déséquilibre.** Il n'est pas possible de prendre en compte efficacement les emojis rares du corpus d'entraînement. Ceci entraîne alors une sur-représentation des emojis les plus utilisés qui représentent un petit échantillon compte tenu du nombre important d'emojis disponibles. Ainsi, nous pourrions nous retrouver avec un système qui peut bien fonctionner mais qui, soit fera du sur-apprentissage sur les éléments rares amenant ainsi une mauvaise prédiction de ces derniers, soit possédera un score de rappel tellement faible que ces éléments seraient tout aussi bien ignorés. La gestion du déséquilibre dans la représentation des emojis est ainsi cruciale si l'on veut obtenir une bonne recommandation, tandis que pour une simple prédiction elle devient moins importante, la recommandation consistant aussi à proposer d'autres emojis liés même si leur utilisation est moins importante.

**Jeu d'étiquettes changeant.** La liste des emojis est toujours amenée à être modifiée et enrichie, comment le modèle de prédiction pourrait-il alors prendre en compte des emojis qu'il n'a pas vu lors de la phase d'entraînement ? Cette limite est très importante dans le cadre d'un système de recommandation à usage réel puisqu'elle provient plus du contexte sociétal où les utilisateurs et les entreprises créent régulièrement de nouveaux emojis, que d'une problématique de

performance pure. En effet, non seulement de nouveaux emojis apparaissent régulièrement, mais chaque application possède son lot d'emojis, ou d'autocollants (*stickers*, des images plus détaillées représentant une attitude émotionnelle). Prenons pour exemple l'emoji licorne, comment promouvoir cet emoji s'il vient juste d'être intégré et donc qu'il n'y a aucun cas d'exemples d'utilisation ? Cette adaptabilité du système de recommandation désiré est importante à prendre en compte, les systèmes de prédiction ou suggestion pure ne suffisant pas pour ce cas.

**Interprétabilité.** Prédire des emojis est une première étape nécessaire mais ne donne pas d'informations sur la raison de cette prédiction. Proposer ainsi des emojis un à un sans donner un sens à la prédiction en structurant la réponse n'entraîne pas toujours une confiance de l'utilisateur vis-à-vis du résultat proposé et de la recommandation effectuée. Si l'utilisateur ne peut pas interpréter les résultats, alors il aura l'impression que le système se trompe, ou alors ne saura pas comprendre pourquoi le système se trompe dans tel ou tel cas. Un bon exemple de ce cas de figure est le cas du sarcasme et de l'ironie : le système de prédiction peut prédire à la fois des emojis parce qu'ils correspondent à l'émotion présente dans le message mais aussi parce qu'il les a vu apparaître dans ce contexte alors qu'ils étaient utilisés pour du sarcasme, c'est-à-dire montrer une émotion ou une polarité contraire au contenu du message.

> This cake is just wonderful ! 🙄

## 3.5. Conclusion

Dans ce chapitre, la prédiction d'emoji a été abordée en ayant pour objectif principal de prédire directement chaque emoji. L'hypothèse des différentes fonctions des emojis dans une conversation et dans l'interprétation d'un message (voir chapitre 1) a fait l'objet de deux approches différentes : une prédiction d'emojis par génération de contenu pour remplacer le mot en cours (fonctions référentielles et d'expression) et une prédiction d'emojis par classification supervisée pour enrichir le contenu textuel (fonctions expressives, conatives et métalinguistiques).

L'approche générative mélangeant chaînes de Markov cachées et lexique de référence agrégé s'est montrée déséquilibrée puisque très efficace sur un certain panel d'emojis et beaucoup moins dans l'ensemble. La prise en compte de 1 215 emojis s'est révélée être une des sources de la difficulté de la tâche, qui montre également des limites par la constitution du lexique de référence.

L'approche discriminante utilisant une classification supervisée multi-étiquettes pour proposer plusieurs emojis possibles pour chaque phrase a donné des ré-

sultats encourageants. Elle est comparable avec l'état de l'art sur la prédiction directe d'emojis (BARBIERI, BALLESTEROS et SAGGION, 2017 ; XIE, Z. LIU et al., 2016a ; X. LI, YAN et al., 2017) et se différencie principalement par une approche orientée caractéristiques discriminantes de sentiment et d'humeur avec des arbres de décision, mais aussi par le corpus de messages instantanés privés utilisé. Cette approche s'est montrée efficace aussi bien sur un corpus limité à 164 emojis sentimentaux qu'à 1 070 emojis divers et variés.

L'information relative à l'humeur de l'utilisateur lors de l'écriture du message s'est montrée déterminante pour la prédiction d'emojis dans un corpus instantané privé. Toutefois l'approche générale, aussi bien générative que discriminante, souffre du fort déséquilibre du jeu d'étiquettes : quelques emojis sont sur-représentés tandis que d'autres sont très peu utilisés. Aussi, la prédiction de l'emoji ou des emojis n'est pas toujours interprétable puisqu'il peut s'agir de plusieurs emojis représentant chacun des idées contradictoires par exemple. De plus, le jeu d'emojis à prédire ne peut pas directement être étoffé sans obtenir des exemples d'utilisation du nouvel emoji. Cette prédiction d'emojis est donc limitée et ne saurait être suffisante à une véritable recommandation d'emojis. Pour pallier à cette problématique, nous proposons dans le chapitre suivant de ne plus prendre en compte les emojis directement, mais leurs catégories dans le chapitre 4.

## 3.6. Synthèse

**Approches méthodologiques** : générative et discriminante

**Approche générative**
— **Objectifs** : améliorer les systèmes existants présent et prendre en compte deux fonctions des emojis (référentielle et expression).
— **Méthode** : utilisation de HMM pour générer le mot à suivre à partir du mot en cours ou du mot précédent. Distance d'édition avec un lexique agrégé pour transformer ce mot en un emoji.
— **Résultats** : efficace pour un ensemble restreint d'emojis de type objet, mais inefficace dans l'ensemble.
— **Limites** : biais de constitution du lexique.
— **Perspectives** : constituer automatiquement le lexique et remplacer la génération du mot.

**Approche discriminante**
— **Objectifs** : prédire plusieurs emojis par phrase et prendre en compte trois fonctions des emojis (conative, expressive et méta-linguistique).
— **Méthode** : classification supervisée multi-étiquettes par *RandomForest* avec prise en compte de l'humeur et de l'analyse de sentiment sur des messages privés.
— **Résultats** : classification efficace avec des sacs de caractères, aussi bien sur les emojis sentimentaux que sur l'ensemble. Apport non négligeable de l'humeur et du sentiment dans la classification.
— **Limites** : déséquilibre des données non pris en compte, impossibilité d'adapter à de nouveaux emojis et difficile interprétabilité de la prédiction finale lorsque plusieurs emojis se contredisent.
— **Perspectives** : ne plus considérer les emojis comme les objets directs de la prédiction, permettre l'ajout de nouveaux emojis.

**Contribution** :
— Prise en compte de cinq des six fonctions de l'emoji dans la conversation.
— Modèle de prédiction par approche générative et correspondance lexicale pour les emojis utilisés en remplacement d'un mot.
— Modèle de prédiction par classification multi-étiquettes pour les emojis utilisés pour ajouter de l'information contextuelle, orienter l'information textuelle ou garder la conversation en cours. Prédiction d'emojis pour proposer plusieurs emojis (jusqu'à 1070) et non pour classer les messages par emoji.
— Application à un corpus de messages instantanés privés avec prise en compte de l'humeur actuelle de l'utilisateur.

# 4. Catégorisation d'emojis émotionnels

## Plan du chapitre


4.1 Introduction **85**
4.2 Approche méthodologique **86**
4.3 Catégorisation d'emojis par plongements lexicaux existants **88**
4.4 De la nécessité de plongements spécifiques pour une catégorisation d'emojis **90**
4.4.1 Corpus de tweets 91
4.4.2 Plongements lexicaux adéquats 93
4.5 Évaluation des plongements lexicaux par comparaison à une typologie des expressions faciales émotionnelles humaines **97**
4.5.1 Partitionnement de référence 98
4.5.2 Métriques d'évaluations quantitatives des partitionnements automatiques 99
4.5.3 Évaluation quantitative des partitionnements automatiques 99
4.5.4 Analyse des erreurs 101
4.6 Conclusion **101**
4.7 Synthèse **103**


## 4.1. Introduction

Dans le chapitre précédent, nous avons abordé différentes approches de prédiction d'emojis. Cependant l'objectif de nos travaux n'est pas la prédiction en soi, mais la recommandation des emojis à l'utilisateur, qui se différencie par l'orientation du choix final plutôt qu'une attribution d'étiquette définitive. Dans notre cas, la recommandation est le fait de proposer proposer différentes émotions identifiées pour enrichir l'information textuelle du message par le biais d'emojis. C'est un des éléments qui nous différencie de l'état de l'art actuel dans lequel le centre d'intérêt est la prédiction d'emojis avec pour objectif l'amélioration des systèmes prédictifs. L'objectif de nos travaux est de laisser à l'utilisateur le libre choix final des emojis utilisés, ce choix devant simplement être guidé et facilité par un système de recommandation. En ce sens, la classification multi-étiquettes abordée dans le chapitre précédent est un premier pas vers cet objectif puisque la recommandation doit permettre de proposer plusieurs emojis.

Un autre objectif est de pouvoir considérer une organisation sous-jacente aux emojis lors de leur recommandation afin de faciliter la compréhension de la recommandation effectuée. Pour ces objectifs, nous proposons dans ce chapitre une méthodologie de catégorisation d'emojis.

Nous partons de l'hypothèse selon laquelle l'usage des emojis représentant des expressions faciales d'émotions reflète une organisation latente. C'est pourquoi nous appliquons notre méthode aux emojis émotionnels, en faisant l'hypothèse que leur usage reflète les nuances de leur organisation.

## 4.2. Approche méthodologique

L'approche méthodologique proposée repose sur la définition de groupes d'emojis obtenue automatiquement. L'objectif est de recommander à l'utilisateur des groupes d'emojis plutôt qu'un emoji unique ou plusieurs emojis indépendants comme ce fut le cas au chapitre 3. Pour répondre à ces objectifs, nous proposons une méthode automatique de recommandation d'emojis mélangeant un modèle de langue pour représenter les emojis, un apprentissage non-supervisé pour la constitution de groupes d'emojis et un apprentissage supervisé pour la prédiction de ces groupes d'emojis obtenus automatiquement. Pour définir automatiquement ces groupes d'emojis nous appliquons la méthodologie suivante dont chaque point fait référence à une des sections détaillées ci-après :

1. Obtention d'une représentation vectorielle des emojis issue de leurs contextes textuels d'utilisation (section 4.3).
2. Partitionnement automatique appliqué sur cet espace vectoriel d'emojis (section 4.4).
3. Évaluation du résultat à l'aide de comparaison avec une théorie de psychologie sociale sur les expressions faciales d'émotion (EKMAN, 1999). (section 4.5).

La majorité des algorithmes d'apprentissage automatique nécessite des données numériques pour fonctionner puisqu'il s'agit d'inférences statistiques et d'algorithmes probabilistes ne pouvant pas directement traiter du texte. Ainsi pour traiter du texte, l'une des approches courantes est de le transformer en un ensemble de vecteurs, c'est-à-dire de vectoriser le texte. Nous avons déjà abordé plusieurs approches pour concevoir des modèles de langue dans le chapitre 2. Dans ce chapitre, nous explorons plus particulièrement les plongements lexicaux devenus une norme ayant remplacé l'utilisation de *bi-grams*, *tri-grams* et autres *n-grams* car ils permettent de condenser l'information contextuelle là où les *n-grams* ne sont qu'une séquence. Mais bien entendu, on pourrait envisager des

plongements lexicaux de *n-grams*.

Lors de la prédiction d'emojis décrite dans le chapitre précédent, nous avons représenté le texte en utilisant des *n-grams* pour pouvoir le traiter par le biais d'une forêt aléatoire d'arbres de décisions. Cependant, cela n'est pas suffisant pour obtenir une représentation précise des emojis. Dans ce cas, un emoji est uniquement un unigramme, en ce sens qu'il n'est représenté que par son mot ou son code associé, ignorant alors les différents contextes d'apparition. Il convient alors d'obtenir une représentation de la somme de ses contextes d'apparition. Pour ce faire, nous utilisons les plongements lexicaux (*word embeddings*) qui permettent d'obtenir une représentation vectorielle en faible dimension issue de l'ensemble des contextes d'apparition du mot, et donc ici, de l'emoji qui nous intéresse. Nous considérons en effet l'emoji comme un mot à part entière représenté par ses différents contextes d'apparition.

L'objectif est de mettre en place des plongements lexicaux d'emojis dans des phrases afin d'obtenir une représentation des emojis un vecteur contenant des informations contextuelles. Une version de cette représentation vectorielle des emojis en contexte d'apparition étant disponible (POHL, DOMIN et al., 2017) (voir chapitre 2), nous décidons donc de l'utiliser afin de vérifier s'il est possible d'obtenir une répartition logique des emojis émotionnels. Plus précisément, nous vérifions si la proximité des vecteurs d'emojis correspond à une catégorisation latente d'émotions qu'ils transmettent ou représentent. Pour ce faire, nous avons sélectionné les emojis représentant des expressions du visage en nous limitant aux catégories précédemment définies par la norme Unicode, en l'occurrence les trois catégories "*face positive*", "*face neutral*" et "*face negative*". Pour être exact, ces catégories correspondent à la version 5.0 de la liste Unicode publiée le 20 juin 2017 et non à l'actuelle version 11.0 publiée le 21 mai 2018. Ces versions changeant une à deux fois par an, nous avons dû nous en tenir à une seule correspondant au début de nos expérimentations, afin de pouvoir comparer par la suite. Pour ce faire, nous observons l'usage de 63 emojis faciaux se rapprochant du visage humain, excluant ainsi les chats 😸, démons 😈, aliens 👽 ou autres 🙈. La liste d'Unicode contenait 70 emojis dans les catégories "face positive", "face negative" et "face neutral", mais seuls 63 d'entre eux sont présents dans notre corpus. Comme expliqué précédemment, cela est dû au déséquilibre dans leur utilisation réelle ; l'emoji tracker pour Twitter[1] en est un exemple flagrant, 😂 étant utilisé 2 298 972 052 fois contre 12 044 088 fois pour l'emoji 😟. L'emoji 😂 est donc utilisé environ 190 fois plus que 😟.

Cette méthode alliant plongements lexicaux (sections 4.3 et 4.4) et sur partitionnement pour définir des groupes d'emojis, est appliquée sur les emojis repré-

1. http://www.emojitracker.com/

sentant des expressions d'émotion par le visage. Elle est ensuite validée avec une théorie existante d'une typologie d'émotions fondées sur les expressions faciales humaines (section 4.5).

Notre premier objectif est donc de vérifier s'il est possible d'obtenir automatiquement, et à partir de leur usage réel, des catégories d'emojis (Section 4.4). Dans les deux sections suivantes, nous appliquons les étapes 1 et 2 de la méthodologie présentée ci-dessus en utilisant différents types de plongements lexicaux et d'algorithmes de partitionnement.

## 4.3. Catégorisation d'emojis par plongements lexicaux existants

Considérant l'état de l'art et les modèles de plongements lexicaux d'emojis déjà disponibles, nous utilisons dans un premier temps le modèle de Pohl (POHL, DOMIN et al., 2017) appris sur 21 millions de tweets et pour lequel un partitionnement hiérarchique (*hierarchical clustering*) était appliqué. En utilisant ce modèle et en ne récupérant que les vecteurs des emojis concernés, nous obtenons un espace vectoriel uniquement constitué d'emojis visage représentant l'expression d'une émotion. Cet espace vectoriel est visible dans la figure 4.1.

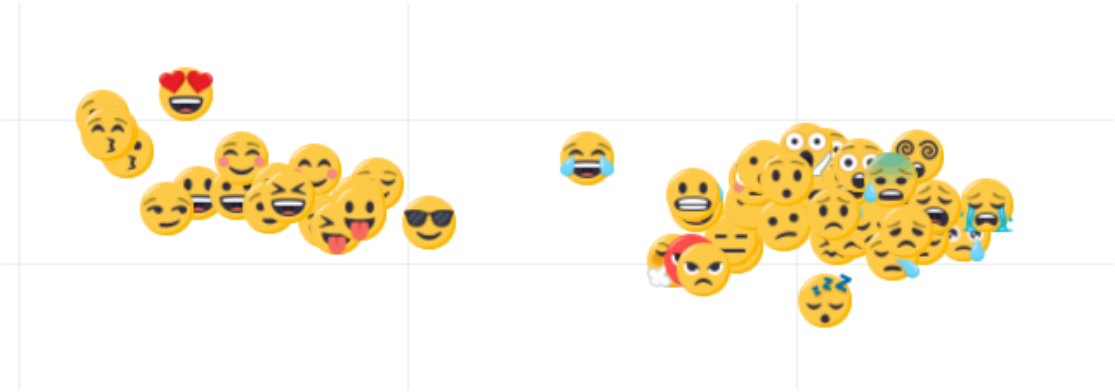

Figure 4.1. – Espace vectoriel d'emojis réduit à 2 dimensions

La figure 4.1 représente l'espace vectoriel des 63 emojis, espace obtenu à partir du modèle de Pohl (POHL, DOMIN et al., 2017). Cet espace vectoriel est initialement fait de 300 dimensions, mais cette figure représente cet espace vectoriel réduit à seulement 2 dimensions. La réduction en 2 dimensions est ici effectuée à l'aide du *T-distributed Stochastic Neighbor Embedding* (TSNE[2]) (MAATEN et HIN-

2. https://lvdmaaten.github.io/tsne/

TON, 2008), offrant une bonne alternative à l'analyse par composante principale (*Principal Component Analysis* ou PCA). Toutefois, les paramètres du TSNE, tels que le taux de perplexité ou le pas d'apprentissage (*learning rate*), peuvent faire fortement varier le résultat final. Dans la figure 4.1, les paramètres utilisés sont un taux d'apprentissage à 100, une perplexité à 30 et une *early exageration* de 2. Ces paramètres ont donc été définis afin d'obtenir une visualisation pouvant être aisément insérée dans ce manuscrit, un autre ensemble de paramètres permet une visualisation plus fine mais techniquement impossible à retranscrire ici[3]. Les autres valeurs de paramètres non détaillées ici sont ceux présents par défaut dans l'implémentation de Scikit-Learn[4]. Cette méthode de visualisation ne peut donc servir qu'à se faire une idée de la proximité que possèdent les éléments entre eux. Ainsi, dans la figure 4.1, on remarque une séparation globale en deux groupes d'emojis. Cette séparation semble indiquer uniquement une polarité positive, neutre ou négative.

L'espace vectoriel existant est désormais utilisé pour un sous-ensemble d'emojis précis, les 63 emojis visibles en figure 4.1. Nous récupérons ensuite uniquement les vecteurs d'emojis pour les utiliser comme données d'entrée pour un algorithme de partitionnement non supervisé. Ces deux étapes sont conformes à la méthodologie présentée précédemment. Afin de comparer nos résultats nous décidons de ne pas modifier l'algorithme de partitionnement, seuls les vecteurs en données d'entrée varient, qu'il s'agissent de ceux de l'existant (POHL, DOMIN et al., 2017) ou d'autres plongements lexicaux que nous appliquons par la suite (section 4.4). Nous appliquons donc les K-moyennes (*k-means*) (MACQUEEN et al., 1967) avec les valeurs des hyper-paramètres visibles dans la Table 4.1.

| | |
|---|---|
| **Clusters** | 63 |
| **Initialisations** | 1000 |
| **Itérations Maximum** | 500 |

Tableau 4.1. – Paramètres fixes des K-moyennes.

Les paramètres utilisés dans la Table 4.1 sont assez standards à ceci près que nous décidons de fixer le nombre de groupes à 63. Ce choix est dû au nombre d'emojis disponibles, le nombre de groupes (*clusters*) est donc égal au nombre d'emojis, permettant à l'algorithme de ne pas forcer l'attribution d'un groupe à un des éléments. Un emoji doit pouvoir être isolé, et il doit pouvoir y avoir un

3. L'ensemble des emojis est représenté plus en détails, entraînant le besoin d'une interaction avec la visualisation pour s'orienter dans l'espace en deux dimensions, ce qui est impossible dans un manuscript.

4. http://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html

nombre variable de groupes finaux. De cette façon nous ne présupposons pas du nombre de catégories d'émotions à obtenir par cette méthode.

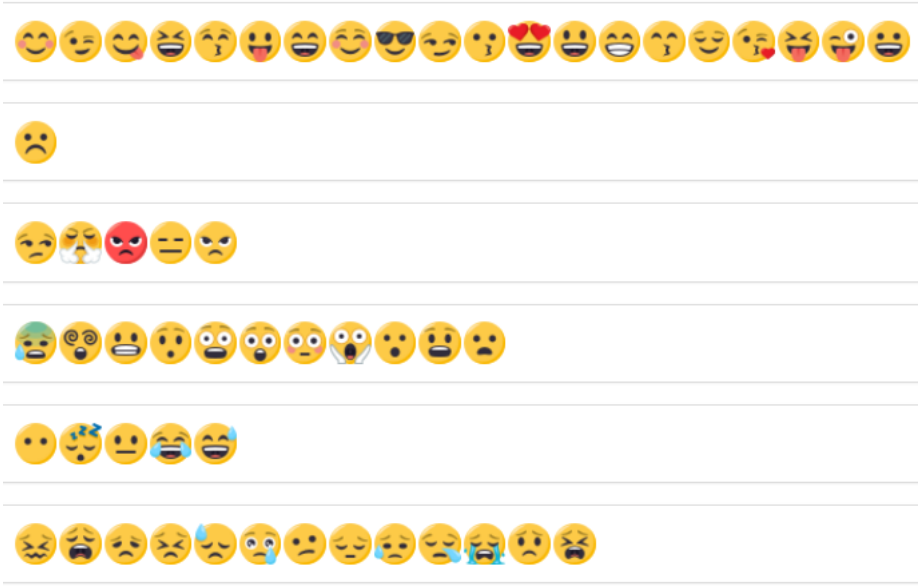

Figure 4.2. – Groupes obtenus à l'aide de l'espace vectoriel existant

Le résultat en est un partitionnement en groupes, visibles dans la figure 4.2. Parmi les 63 groupes possibles nous éliminons les groupes vides pour obtenir 6 groupes d'emojis. Ces 6 groupes sont trop généraux pour servir plus tard de point de départ pour une recommandation d'emojis, leur granularité n'est pas assez fine, les nuances intrinsèques au premier groupe (joie) ne sont pas exploitées. Il apparaît alors nécessaire d'essayer d'autres approches de plongements lexicaux pour arriver à nos fins.

## 4.4. De la nécessité de plongements spécifiques pour une catégorisation d'emojis

Constatant les limites de l'usage d'un tel modèle de plongements lexicaux, nous proposons de mettre en place plusieurs types afin d'en comparer les résultats. Étant donné que nous n'avons pas accès aux données d'entraînement du modèle précédent, il convient de récupérer nos propres données similaires et de tester différentes architectures de plongements lexicaux, et non uniquement l'utilisation de *skip-grams*. Pour cela nous utilisons les données présentées au tableau 4.2.

### 4.4.1. Corpus de tweets

Le tableau 4.2 présente le corpus final utilisé pour apprendre de nouveaux plongements lexicaux après nettoyage et filtre par langue et par détention d'emojis. Ce corpus est initialement, c'est-à-dire avant pré-traitements, constitué de plus d'un million de tweets émis sur le continent nord-américain, soit plus exactement les États-Unis. Nous supprimons les mots-dièse ainsi que les liens hypertextes (*URL*). Notre objectif n'étant pas *in fine* de recommander des emojis dans le cadre de Twitter, nous ignorons volontairement des spécificités majeures de cette plateforme, comme le mot-dièse, l'abondance de liens externes ou encore l'indication de ReTweet. Le nettoyage et le filtrage uniquement sur les tweets contenant des emojis, se résume en plusieurs étapes que voici et que nous détaillons ensuite :

1. Suppression des liens hypertextes
2. Normalisation du texte en minuscule
3. Suppression des mots-dièse
4. Identification des langues utilisées
5. Limitation de la répétition typographique
6. Tokenization
7. Lemmatisation

| | |
|---|---|
| **Tweets** | 695 031 |
| **Emojis** | 901 669 |
| **Mots par tweet** | 10,81 mots |
| **Emojis distincts** | 844 |
| **Emojis par tweets** | 1,30 |

Tableau 4.2. – Corpus de tweets contenant des emojis

**Normalisation du texte.** La normalisation effectuée est très limitée : nous nous contentons de mettre le texte en minuscules afin de mieux gérer la correspondance des mots par la suite. Cette normalisation aurait pu être plus avancée comme pour Tian (TIAN, DINARELLI et al., 2015) en prenant en compte les fautes d'orthographes, les fautes typographiques habituelles ou encore la normalisation des interjections. Cependant nous décidons de ne pas procéder ainsi afin d'obtenir le contexte d'utilisation réel dans lequel la recommandation sera effectuée. Les fautes d'orthographes peuvent être volontaires ou relever d'un vocabulaire propre à un groupe social. De plus, la normalisation des interjections pose problème selon les langues (*haha* en français, *jaja* en espagnol, *etc.*). Étant donné

l'objectif final de pouvoir adapter l'approche à de nouveaux messages et donc à de nouveaux mots d'argot, il convient de ne pas appliquer une normalisation de texte avancée ayant pour but de transformer le langage courant en communément admis comme correct.

**Identification de la langue.** Ces tweets sont récupérés via le filtre de l'interface de programmation (API) twitter [5] pour ne garder que les tweets écrits en anglais. Par souci de vérification et de possibles polyglottes, nous y avons ajouté une détection automatique de la langue employée majoritairement. Plus exactement, tous ces tweets ont été préalablement filtrés par un détecteur de langue utilisant la liste des mots vides de NLTK [6]. Leur ratio d'apparition dans le texte est analysé après intersection entre les deux ensembles de mots vides, ceux du texte et ceux des différentes langues, le tout afin d'obtenir un corpus quasi totalement monolingue puisque nous ne pouvons éviter l'insertion d'un mot d'une autre langue. Cette insertion de mots d'une autre langue correspond souvent à une diglossie dépendant du contexte conversationnel : le sujet de la conversation ou les capacités multilingues du correspondant. Nous conservons donc uniquement les tweets dont la langue principale détectée automatiquement est l'anglais.

**Répétition typographique.** La répétition typographique donne des indices d'accentuation dans le texte ("*Pretty pleeeaaaase ! ! ! !*") mais peut poser problème lors du traitement automatique du texte. Nous limitons la répétition typographique à deux lettres à l'aide d'expressions régulières afin d'obtenir de meilleures performances pour notre recommandation dans son identification des contextes similaires, limitant ainsi le nombre de vocabulaire différent. Ce choix est motivé par la taille du corpus : pour que les vecteurs obtenus trouvent des similarités plus facilement dans un corpus ayant des répétitions typographiques non limitées, il faudrait que ce corpus soit très large, ce qui n'est pas le cas du nôtre. Aussi, le choix de se limiter à deux lettres répétées est motivé par la langue anglaise qui ne contient pas de répétitions de trois lettres. Il s'agit donc d'un choix dépendant du langage cible et qu'il conviendrait de modifier lorsque la méthode est appliquée à une autre langue.

**Tokenisation et lemmatisation.** L'obtention des formes canoniques des mots (lemmatisation) et la séparation des éléments du texte (tokenisation) est effectuée à l'aide de WordNet (G. A. MILLER, 1995) et des modèles associés disponibles dans NLTK [7]. Cette étape joue le même rôle que la limitation de la répétition typographique mais ne saurait fonctionner parfaitement étant donné l'absence volontaire de normalisation du texte (voir paragraphe précédent).

5. https://dev.twitter.com/streaming/overview
6. http://www.nltk.org/
7. http://www.nltk.org/

### 4.4.2. Plongements lexicaux adéquats

Le corpus étant nettoyé et pré-traité, nous entraînons des plongements lexicaux à l'aide de *Word2Vec* (REHUREK et SOJKA, 2010 ; MIKOLOV, SUTSKEVER et al., 2013) implémenté dans Gensim[8] en variant deux architectures : (1) *Skip-Gram* pour prédire le contexte à partir de l'emoji avec l'algorithme *softmax* hiérarchique (MIKOLOV, CHEN et al., 2013), et (2) sac de mots continus (*CBOW*) pour prédire l'emoji à partir de son contexte. Ces deux architectures sont décrites dans les sections suivantes.

#### 4.4.2.1. Plongements lexicaux en skip-gram

Nous reprenons dans un premier temps l'architecture skip-gram pour entraîner ce modèle, c'est-à-dire pour prédire l'emoji à partir de son contexte, en conservant uniquement les mots apparaissant au moins 5 fois dans le corpus d'entraînement, le tout afin d'obtenir des vecteurs de dimension 300. Il s'agit donc de la même approche utilisée précédemment par Pohl (POHL, DOMIN et al., 2017) à ceci près que les données utilisées et leur pré-traitements diffèrent. En entraînant ces plongements lexicaux nous obtenons un espace vectoriel d'emojis (figure 4.3).

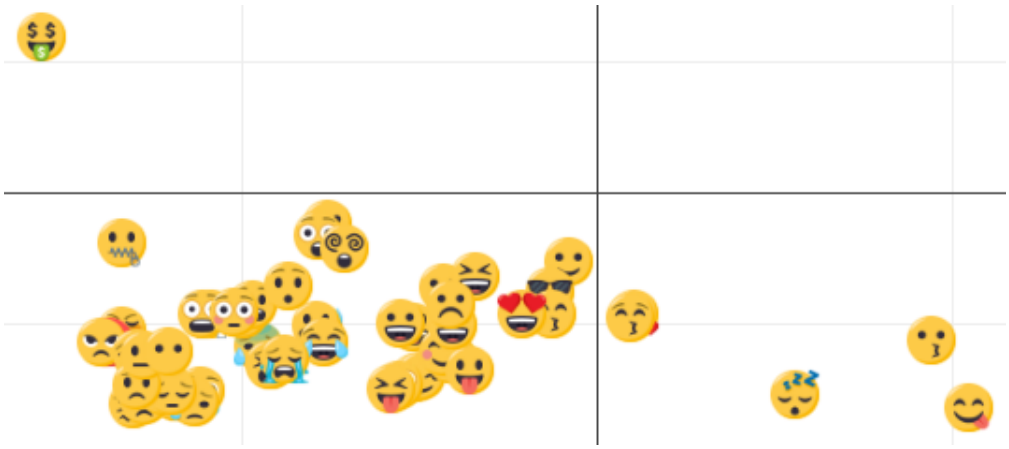

Figure 4.3. – Représentation en deux dimensions par TSNE de l'espace vectoriel d'emojis obtenu par architecture skip-gram.

L'espace vectoriel de la Figure 4.3 représente la distribution des vecteurs de chaque emoji obtenus par plongements lexicaux, entraînés sur tout le texte. Nous

8. https://radimrehurek.com/gensim/models/word2vec.html

considérons donc les emojis comme des éléments de la phrase à part entière. C'est ce qui nous permet d'obtenir leurs contextes textuels d'utilisation. Encore une fois, cette figure est obtenue avec un TSNE configuré pour rapprocher davantage (par l'*early exageration*) les éléments afin qu'ils apparaissent tous dans cette image. Nous pouvons constater une répartition proche de celle obtenue précédemment en figure 4.1.

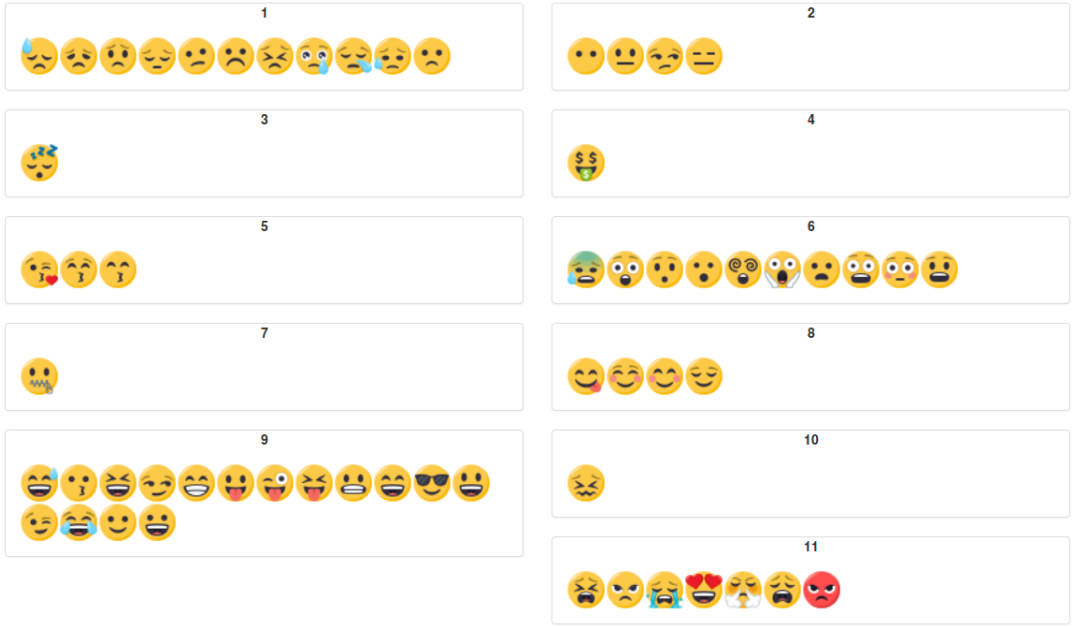


Figure 4.4. – Groupes obtenus en partitionnant automatiquement les vecteurs d'emojis de la figure 4.3 à l'aide de K-moyennes.

En appliquant les K-moyennes (MacQueen et al., 1967) sur les vecteurs d'emojis, nous obtenons 11 groupes non-vides sur les 63 groupes possibles dus aux 63 emojis considérés. Ces groupes sont visibles dans la Figure 4.4. Ces groupes ont une granularité plus fine que ceux obtenus précédemment (Figure 4.2), ce qui peut s'expliquer par plusieurs facteurs. Le changement de corpus tout d'abord, mais surtout le pré-traitement effectué et ciblé pour nos besoins de prise en compte du texte en supprimant au maximum les attributs propres à Twitter tels que les mots-dièse (cf. Section 4.4.1). Parmi ces groupes, certains peuvent se référer aux émotions basiques telles que le plaisir sensoriel 😊 😊 😋 😌, ou encore la tristesse par le biais du premier groupe (groupe 1). Ces 11 groupes présentent certaines limites et, si l'on exclut le bon isolement des emojis particuliers 🥳 🤐 🤑, certaines catégories restent trop généralistes ou ambiguës. C'est le cas de la catégorie 11 de la figure 4.4 qui mélange colère, excitation et tristesse.

Dans la section suivante nous présentons l'usage d'une structure prenant moins

en compte les éléments rares afin d'obtenir des groupes plus uniformes.

#### 4.4.2.2. Plongements lexicaux en sac de mots continu

Afin de pouvoir comparer avec les résultats précédents, nous modifions la manière dont sont obtenus les plongements lexicaux en passant de l'architecture skip-gram à celle des sacs de mots continus (*Continuous Bag-Of-Words* ou *CBOW*). Les résultats obtenus sont visibles en Figure 4.5.

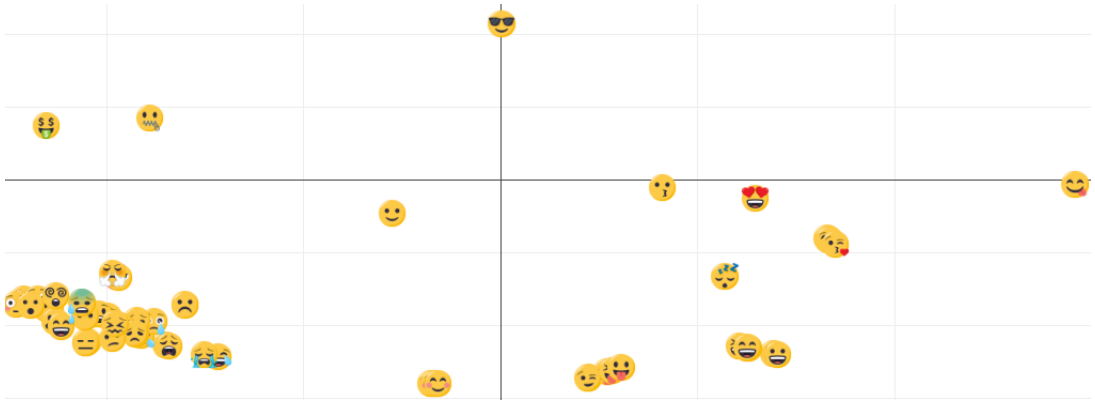

Figure 4.5. – Espace vectoriel d'emojis obtenu par sac de mots continus, réduit à deux dimensions.

L'espace vectoriel visible en Figure 4.5 montre plusieurs groupes et s'éloigne davantage de la constitution de deux groupes principaux relatifs à la polarité des emojis. Ces groupes obtenus sont plus fins et plus nombreux que précédemment. Ce constat est confirmé par l'application de K-moyennes sur ces vecteurs d'emojis. La figure 4.6 montre le résultat du partitionnement obtenu.

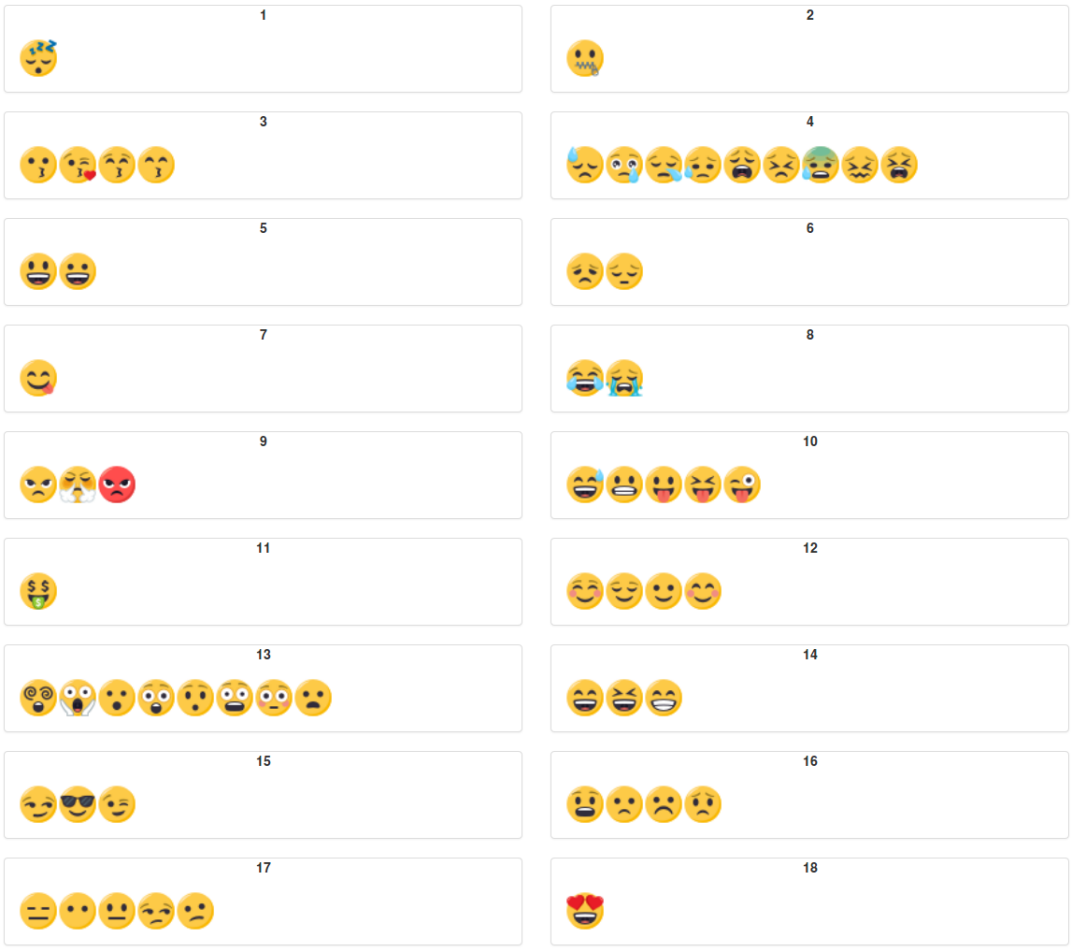


Figure 4.6. – Groupes obtenus en partitionnant automatiquement les vecteurs d'emojis de la figure 4.5 à l'aide de K-moyennes.

L'application du même partitionnement sur les vecteurs obtenus via la méthode de sacs de mots continus donne 16 groupes d'emojis. Toutefois certaines divisions semblent quelque peu étranges : les emojis représentant un bisou sont séparés en trois groupes, dont deux groupes semblent mélanger colère et mécontentement. Les groupes obtenus ne sont donc pas suffisamment précis malgré un meilleur résultat global.

Comment expliquer une telle différence de qualité entre les deux représentations vectorielles ? L'architecture skip-gram est très utilisée pour une représentation vectorielle de texte en vue d'une classification supervisée par la suite puisqu'elle possède l'avantage de bien prendre en compte les éléments rares (GUTHRIE, ALLISON et al., 2006). Cependant notre objectif diffère. Nous désirons obtenir une représentation moins précise des tendances d'utilisation des emojis en contexte en vue d'un partitionnement non supervisé. L'architecture en sac de mots continus est plus appropriée puisqu'elle reste plus en surface en considé-

rant des tendances générales en omettant les éléments rares, les éléments trop rares étant ensuite éludés par le paramétrage des plongements lexicaux pour lequel nous ignorons les mots apparaissant moins de 5 fois dans le corpus.

Il convient d'évaluer les partitionnements obtenus. Pour cela, nous proposons de comparer les groupes obtenus à une catégorisation proposée en psychologie à partir des expressions faciales d'émotions humaines (EKMAN, 1999) vis-à-vis de la catégorisation théorique existante.

## 4.5. Évaluation des plongements lexicaux par comparaison à une typologie des expressions faciales émotionnelles humaines

Les groupes obtenus précédemment nous ont permis de déterminer la meilleure approche pour la représentation vectorielle d'emojis à des fins de catégorisation. Il convient toutefois d'évaluer ces différents résultats et pour cela un fondement théorique est nécessaire. Ce fondement théorique nécessaire concerne les catégories d'émotions représentées par des expressions faciales. Dans nos travaux, nous proposons d'utiliser l'une des catégorisations les plus connues : la classification des 16 émotions de base représentables par le biais d'expressions du visage établie par (EKMAN, 1999). Cette classification se découpe en 16 catégories :

1. La joie (joy)
2. L'amusement (amusement)
3. La colère (anger)
4. Le mépris (contempt)
5. La satisfaction (contentment)
6. Le dégoût (disgust)
7. La gêne (embarrassment).
8. L'excitation (excitement)
9. La peur (fear)
10. La culpabilité (guilt)
11. La fierté dans la réussite (pride in achievement)
12. Le soulagement (relief)
13. La tristesse ou détresse (sadness/distress)
14. La satisfaction (satisfaction)
15. Le plaisir sensoriel (sensory pleasure)
16. La honte (shame)

### 4.5.1. Partitionnement de référence

L'évaluation se fait en fonction de la conformité des groupes obtenus avec la classification d'Ekman (EKMAN, 1999), plus simple à prendre en compte que la roue des émotions de Plutchik (PLUTCHIK, 2001), puisque cette dernière nécessiterait de prendre en compte les relations entre les différents concepts d'émotions. Pour cela il est nécessaire d'établir une correspondance entre les 63 emojis considérés et la classification en question. Nous partons donc des indices visuels des emojis pour manuellement attribuer à chaque emoji une étiquette parmi les 16 catégories. Cette base comparative, établie manuellement, sera comparée avec nos partitionnements automatiques. Elle est visible au tableau 4.3.

| Étiquette | Emojis |
|---|---|
| surprise | 😳 😮 😳 😳 🤯 |
| satisfaction | 😏 😉 😎 |
| amusement | 😅 😃 😜 😝 😬 |
| peur | 😱 |
| excitation | 😋 😂 |
| colère | 😤 👺 😒 |
| zip | 🤐 |
| joie | 😙 😀 🙂 😁 😌 😆 😊 😊 😙 😃 😄 😗 😍 😘 |
| tristesse | 😪 😖 😦 😣 😫 😃 😞 😔 😭 😅 😟 😪 😲 😨 😓 |
| mépris | 😐 😑 😶 😔 😕 😒 |
| sleep | 😴 |
| pas content | ☹ 🙁 |
| money | 🤑 |

Tableau 4.3. – Partitionnement de référence établi manuellement pour le calcul des métriques.

Le tableau 4.3 montre la présence de 9 catégories d'Ekman parmi 13 catégories définies manuellement. Les 16 catégories d'émotions proposées par Ekman ne sont donc pas toutes présentes dans le jeu d'étiquettes (*i.e.* les 63 emojis). De plus, malgré le fait que nous ayons volontairement pris les catégories d'Unicode relatives aux expressions d'émotions par le visage, trois emojis sont un peu à part : 😴 🤑 et 🤐. Ces trois emojis ne correspondent pas à des expressions faciales d'émotion et sont donc isolés dans leurs propres groupes.

### 4.5.2. Métriques d'évaluations quantitatives des partitionnements automatiques

En utilisant les groupes de référence du tableau 4.3, nous calculons quatre métriques afin d'obtenir une évaluation quantitative des divers partitionnements obtenus.

***Adjusted Mutual Information* (AMI).** L'AMI, définie par (VINH, EPPS et al., 2010), est l'information mutuelle entre deux partitionnements ajustée en fonction de la chance d'avoir plus d'information mutuelle. Cette chance est fondée sur le nombre de groupes : plus le nombre de groupes est élevé, plus le score d'information mutuelle (MI) sera ajusté.

**Homogénéité.** Le score d'homogénéité (ROSENBERG et HIRSCHBERG, 2007) indique la capacité pour les groupes obtenus de ne contenir que des étiquettes homogènes, c'est-à-dire uniquement des emojis appartenant à une même catégorie dans le résultat de référence (Tableau 4.3).

**Complétude.** La complétude représente la capacité du partitionnement à regrouper ensemble tous les éléments d'un même groupe de référence. Le groupe de référence doit donc être retrouvé au complet dans les résultats de la prédiction, ce qui en fait une métrique comparable au rappel.

**V-mesure.** Il s'agit de la moyenne harmonique entre les scores d'homogénéité et de complétude. Cette moyenne harmonique a été introduite pour la première fois par Rosenberg *et al.* (ROSENBERG et HIRSCHBERG, 2007) qui montrent que pour une F-mesure similaire, la V-mesure peut significativement changer. Elle permet d'obtenir des informations plus précises sur la qualité finale du partitionnement puisque, contrairement à la F-mesure par exemple, elle ne cherche pas seulement à prendre en compte la bonne attribution des étiquettes à une classe.

### 4.5.3. Évaluation quantitative des partitionnements automatiques

À l'aide de ces quatre métriques d'évaluation et du partitionnement de référence, nous obtenons différents scores en variant les algorithmes de partitionnement. Deux algorithmes nous intéressent principalement : les K-moyennes utilisées précédemment et le partitionnement spectral (*spectral clustering*) (NG, JORDAN et al., 2002). Ces deux algorithmes sont régulièrement comparées en mettant en avant la simplicité des K-moyennes la rendant applicable à des corpus de très grande taille, contre le fait que le partitionnement spectral n'assume pas de forme de groupes tandis que les K-moyennes orientent vers des ensembles

convexes (Von Luxburg, 2007). Nous configurons le partitionnement spectral[9] en définissant 63 groupes possibles, un noyau gaussien au coefficient $\gamma 0.7$ et une discrétisation pour se démarquer des k-moyennes

| | **AMI** | **Homogénéité** | **Complétude** | **V-mesure** |
|---|---|---|---|---|
| **K-means** | 35,08 | 68,44 | 59,39 | 63,59 |
| **Spectral Clustering** | 53,04 | 85,63 | 69,45 | 76,70 |

Tableau 4.4. – Résultats comparatifs en variant uniquement le partitionnement appliqué.

Le tableau 4.4 est obtenu en variant uniquement les algorithmes de partitionnement appliqués sur l'espace vectoriel obtenu par sac de mots (Figure 4.5). Ces résultats montrent un gain net du partitionnement spectral comparé aux K-moyennes, et ce, selon toutes les métriques utilisées. De plus, comme ils sont fondés sur la catégorisation d'Ekman, ils nous indiquent dans quelle mesure notre approche est capable de se rapprocher de cette catégorisation à partir de l'usage réel des emojis.

| | | | |
|---|---|---|---|
| Aucun 🥳 | Aucun 🤐 | Joie 😙 😘 😗 😚 | Tristesse 😓😢😪😔😩😣😭😖😫 |
| Joie 😆 😃 | Honte 😞 😔 | Excitation 😋 | Aucun (pas clair) 😂 😭 |
| Colère 😒 😤 👺 | Amusement 😅 😬 😃 😆 😜 | Aucun 🤑 | Plaisir sensoriel 😊 😌 🙂 😊 |
| Peur / Surprise 🤯😱😮😳😲😬😳😦 | Joie / Amusement 😄 😆 😁 | Satisfaction / Fierté 😏 😎 😉 | Aucun (pas content) 😫 😕 ☹ 😟 |
| Mépris 😑 😶 😐 😒 😕 | | Excitation 😍 | |

Tableau 4.5. – Groupes d'emojis obtenus par regroupement spectral sur des plongements lexicaux en sac de mots continus. Les étiquettes proviennent des catégories d'expression de l'émotion d'Ekman.

Dans le tableau 4.5 le résultat présenté est celui du meilleur partitionnement, les étiquettes faisant référence aux émotions basiques d'Ekman. Bien que ces étiquettes soient attribuées manuellement *a posteriori* à des fins d'interprétation,

9. L'implémentation de Sckikit Learn est utilisée

nous pouvons constater plusieurs aspects. Tout d'abord, les emojis non réalistes cités précédemment (🥳🤑🤐) sont isolés dans leur propre groupe, leur différence étant alors mise en avant. De plus, le partitionnement montre une différence de granularité entre ces groupes, comme en atteste la catégorie "joie" qui se retrouve divisée en 3 groupes faisant sens : le regroupement des emojis liés au bisou 😙😘😗😚, la joie d'intensité moyenne 😆 😃 et la joie de forte intensité 😄 😆 😁. Ce n'est pas le cas des catégories "tristesse" et "surprise" qui regroupent tous leurs éléments en un seul groupe d'emojis.

### 4.5.4. Analyse des erreurs

Cette catégorisation n'est pas exempt d'erreurs comme l'indiquent les scores d'AMI et de complétude (Tableau 4.4). La première erreur est l'insertion d'éléments appartenant à une autre catégorie. C'est par exemple le cas de la catégorie 😂😭 qui mélange deux emojis supposés opposés : l'un est hilare tandis que l'autre pleure à chaudes larmes. Toutefois, ces erreurs au niveau de l'évaluation quantitative n'en restent pas toujours des erreurs une fois évaluées qualitativement. En effet, la granularité fine présente pour uniquement certains groupes n'est pas prise en compte dans l'évaluation par métriques, et non présente dans le partitionnement de référence du tableau 4.3. Ainsi le résultat se veut bien meilleur et plus intéressant que lorsqu'il est simplement jugé par une métrique d'évaluation. Aussi, cette limite apparaît clairement pour certains groupes tels que le groupe de la "honte" et celui du "plaisir sensoriel" qui sont tous deux non prévus initialement dans le partitionnement de référence, mais qui font tout de même sens lorsque l'on souhaite interpréter les groupes obtenus. L'interprétation des emojis et donc des groupes obtenus automatiquement, empêche de pouvoir uniquement se fonder sur une évaluation quantitative de la méthode utilisée. Ces interprétations multiples possibles compliquent l'évaluation d'une bonne catégorisation. Ainsi, si l'on se fonde uniquement sur une comparaison avec le partitionnement de référence, des éléments qui auraient dû se trouver dans la catégorie "mépris" se retrouvent dans la catégorie "honte".

## 4.6. Conclusion

Dans ce chapitre, nous avons montré une méthodologie permettant d'obtenir automatiquement une catégorisation d'emojis en fonction de leur usage réel. Nous avons appliqué cette méthodologie pour les emojis représentant une expression faciale d'émotion en suivant la norme Unicode et ce, afin de pouvoir vérifier la conformité des résultats issus de l'usage réel, avec la typologie existante (Ekman, 1999). Cette méthodologie repose sur un pré-traitement de tweets, une

création de plongements lexicaux d'emojis et un partitionnement automatique par apprentissage non supervisé. L'évaluation repose sur une comparaison du partitionnement avec une typologie existante des expressions faciales humaines.

Nous avons montré que l'obtention de vecteurs d'emojis adéquats par plongements lexicaux est corrélée à l'utilisation finale de ces vecteurs. La plupart des modèles de plongements lexicaux sont à des fins de représentation d'une phrase pour une classification et mettent l'accent sur l'importance des éléments rares et discriminants. Nous avons également montré qu'afin d'obtenir une représentation globale pour un partitionnement, il est préférable d'obtenir des vecteurs issus de plongements lexicaux par architecture de sac de mots continus, car ils donnent moins d'importance aux éléments rares, ce qui donne des groupes plus homogènes. En particulier, ce changement de vecteurs a un plus grand impact sur le résultat final que la variation de méthode de partitionnement automatique.

Les résultats valident l'hypothèse selon laquelle les emojis sont utilisés à des fins de reproduction de comportements humains non verbalisés en langue naturelle, dont les rôles dans la communication sont abordés en section 1.2. Le système permet de montrer une granularité différente selon les catégories qui s'explique par un taux d'usage déséquilibré : la plupart des emojis de la catégorie "joie" sont les emojis les plus utilisés sur twitter [10]. Nous avons montré que l'évaluation de notre approche se doit d'être double, l'évaluation quantitative par métriques ne suffit pas à affirmer avec certitude de la qualité du partitionnement.

Ces résultats étant obtenus sur un jeu d'emojis restreint issu de la version 5 de la catégorisation Unicode, il conviendrait de reproduire les expériences sur un corpus de tweets plus récents possédant assez d'exemples pour les nouveaux emojis parus. Aussi, comme cette méthode possède l'avantage d'être applicable sur d'autres jeux d'emojis, il serait intéressant d'analyser les résultats sur des emojis de nature différente, tels que des emojis représentant des objets ou des concepts.

10. http://www.emojitracker.com/

## 4.7. Synthèse

**Objectifs** : obtenir automatiquement une catégorisation d'emojis par leurs usage qui soit utilisable pour de la recommandation.

**Approche méthodologique** : plongements lexicaux et partitionnement appliqués à un sous-ensemble d'emojis qu'il est possible d'évaluer : les emojis représentant des expressions du visage. Le partitionnement ne définit pas le nombre de catégories en amont.

**Résultats** : obtention de catégories d'emojis larges et peu fines en modèle *skip-gram* tandis que des plongements lexicaux plus en surface donnent une meilleure granularité. Obtention de catégories plus proches de celles de la référence théorique en utilisant un partitionnement spectral discriminant.

**Limites** : quelques erreurs et ambiguïtés dans les catégories obtenues. Méthodologie comparée à une théorie dominante mais ne faisant pas l'hunanimité. Granularité déséquilibrée influencée par le déséquilibre des données.

**Contribution** : obtention automatique de groupes d'emojis à granularité fine pouvant être recommandés. Utilisation d'apprentissage automatique comme outil de compréhension des emojis au travers de leur utilisation et de la comparaison avec une théorie existante. Confirmation de l'usage de l'emoji comme substitut des indices faciaux lors d'une conversation en face-à-face.

# 5. Recommandation d'emojis par prédiction de leur catégorie

## Plan du chapitre


5.1 Introduction **105**
5.2 Corpus d'apprentissage pour la recommandation d'emojis **107**
5.2.1 Données du corpus 107
5.2.2 Pré-traitement des données 107
5.3 Encodage des données **108**
5.4 Élagage et remplissage des messages **109**
5.5 Prédiction de catégories d'emojis **109**
5.6 Impact de la temporalité **116**
5.7 Utilisation du système prédictif pour une recommandation **119**
5.8 Limites **121**
5.8.1 Multiple usages des emojis dans le corpus 121
5.8.2 De la différence entre tweets et messages informels privés 123
5.9 Conclusion **124**
5.10 Synthèse **126**


## 5.1. Introduction

La recommandation est le fait de proposer des choix à l'utilisateur en fonction de critères objectifs ou subjectifs. Ainsi la recommandation, que ce soit sur Netflix, YouTube ou autres, est fondée sur les préférences et les avis que l'utilisateur a effectué mais aussi sur son historique de visualisation et les catégories de films. Ces exemples mélangent deux approches principales de la recommandation : la recommandation fondée sur la confiance (*Trust-based recommendation*) (O'DONOVAN et SMYTH, 2005) et celle fondée sur le contenu (*Content-based recommendation*) (PAZZANI et BILLSUS, 2007). La première utilise les indications de confiance telles que l'évaluation par note ou nombre d'étoiles, tandis que la seconde utilise le contenu directement généré par l'utilisateur. Dans ce chapitre, nous proposons une recommandation fondée sur le contenu (*i.e.* les emojis utilisés) pour proposer un ensemble de choix possibles à l'utilisateur, ce qui la distingue d'une prédiction directe d'emojis (présentée au chapitre 3).

Le recommandation proposée ici est fondée sur un système de classification appris à partir du contenu utilisateur. Plusieurs travaux de recherche très récents

portent sur la classification de messages par emoji (BARBIERI, BALLESTEROS et SAGGION, 2017 ; FELBO, MISLOVE et al., 2017 ; ZHAO et ZENG, 2017) (voir section 2.3). Toutefois, les modèles proposés ne sont validés, testés et conçus que dans le but de pouvoir retrouver l'emoji utilisé par l'utilisateur dans un contexte précis. Ainsi, la plupart des travaux sont orientés vers la prédiction d'emoji et non de la *recommandation* (XIE, Z. LIU et al., 2016a). Dans nos travaux présentés au chapitre 3, nous nous différencions des travaux existants en supposant que plusieurs emojis doivent être proposés via une prédiction multi-étiquettes en évitant par conséquent une prédiction mono-étiquette. Malgré cela, notre système requiert tout de même des exemples d'usages d'emojis liés à leur contexte, en guise de corpus d'entraînement. Il est donc nécessaire d'obtenir un système qui s'en détache tout en permettant de conserver l'usage de métriques classiques d'évaluation de classification. De plus, il convient de prendre en compte les cas où les emojis sont utilisés pour représenter le sentiment ou l'émotion inverse du message textuel afin, par exemple, de transmettre un message sarcastique. Il est en effet difficile de connaître le sentiment ou l'émotion que l'utilisateur souhaite transmettre uniquement à partir du contenu textuel.

Dans cette perspective, notre objectif est de proposer une recommandation d'emojis efficace sans utiliser d'approche classique de systèmes de recommandation (voir section 1.5). L'objectif n'est pas de recommander uniquement de nouveaux éléments, mais d'identifier à partir du contexte plusieurs emojis représentatifs de l'émotion que l'utilisateur veut transmettre. Cette particularité s'explique par la nature de l'objet recommandé : l'emoji est un élément persistant et redondant contrairement aux éléments généralement recommandés comme les livres ou les films.

Plusieurs hypothèses sont ainsi posées. Dans un premier temps, nous considérons la recommandation d'emojis comme un cas à part qui se situe entre prédiction et recommandation. Ensuite, nous considérons qu'une recommandation d'emojis doit pouvoir se détacher du simple système prédictif, en adaptant son utilisation et son évaluation. Par le biais de la prédiction des catégories obtenues, nous souhaitons pouvoir prédire la catégorie d'emojis à partir d'un message instantané, chaque catégorie étant liée à une émotion basique. Ce chapitre tend donc à illustrer le principe de recommandation hybride que nous proposons : une prédiction de panels de choix possibles pour l'utilisateur, regroupés par similarités sémantiques.

Dans ce chapitre, le corpus utilisé est décrit dans la section 5.2 et le traitement des données est détaillé en sections 5.3 et 5.4. La méthode utilisée est ensuite présentée en section 5.5. Enfin, l'analyse de cette méthode et de son utilisation pour la recommandation finale (sections 5.6 et 5.7) précède ensuite l'étude des limites de cette approche (section 5.8).

## 5.2. Corpus d'apprentissage pour la recommandation d'emojis

### 5.2.1. Données du corpus

Pour apprendre un modèle de recommandation d'emojis sentimentaux, nous avons utilisé un corpus de messagerie instantanée privée recueilli dans le cadre de l'application propriétaire de la société Caléa Solutions [1]. Ce choix est motivé par le contexte applicatif, l'objectif étant de développer un système de recommandation pour une messagerie sociale instantanée, il est nécessaire de constituer un corpus le plus proche possible du contexte d'utilisation final. Ce corpus correspond à celui utilisé pour la prédiction d'emoji (Section 3.3.1), enrichi de messages en anglais issus du Royaume-Uni et de l'Australie, passant de 1 118 124 messages à 1 304 698 messages avec un taux de 1,36 catégorie par message. Afin de rester au plus près possible du cadre applicatif désiré, une application instantanée privée, nous excluons les tweets comme un possible corpus (Section 5.8.2).

### 5.2.2. Pré-traitement des données

Le pré-traitement du corpus s'apparente à celui effectué précédemment (chapitre 3) : chaque message est d'abord séparé en deux champs, l'un pour l'étiquette et l'autre pour le texte associé (Figure 5.1). Ensuite, nous supprimons les liens hypertextes avant de normaliser le texte en minuscule. La seule information géographique ne constituant pas une source assez fiable d'identification de la langue, nous appliquons également une identification automatique des langues présentes dans un texte à l'aide de la proportion de mots vides propres à chaque langue. Enfin, là aussi chaque message est limité à deux répétitions typographiques maximum, tandis que la tokenisation et la lemmatisation sont les mêmes que précédemment.

L'objectif étant de prédire les catégories d'emojis obtenues automatiquement (chapitre 4), nous appliquons ensuite une correspondance entre emojis présents et catégories afin de ne garder que l'information relative à la catégorie. De cette manière, si un message contient plusieurs emojis appartenant à la même catégorie, il ne lui sera alors attribué qu'une seule étiquette relative à cette catégorie. Concrètement, le message suivant :

> 🧐😳😲 Wonderfuuuuul ! ! !

devient alors ceci :

1. Mood Messenger : http://www.moodmessenger.com/

| Catégorie | Texte |
|---|---|
| Surprise | wonderfuul!! |

Tableau 5.1. – Exemple de donnée après pré-traitements énumérés ci-dessus (répétition typographique, etc.).

Pour cette recommandation nous considérons une classification supervisée mono-étiquette, également appelée classification multi-classe, motivée par plusieurs facteurs. Tout d'abord, nous avons peu de messages ayant des emojis de plusieurs catégories. En moyenne, chaque message contient 1,36 catégories. Cela peut s'expliquer par le fait que les catégories sont déjà obtenues par le biais d'une proximité contextuelle contenue dans les plongements lexicaux. Pour prendre également en compte les cas où plusieurs catégories sont présentes pour un seul et même message, ces messages possédant plusieurs étiquettes sont répétés en plusieurs instances ayant chacune une de ces étiquettes. Ainsi, si un message possède des emojis de deux catégories différentes A et B, il sera répété deux fois, une fois avec la classe A, une autre fois avec la classe B.

## 5.3. Encodage des données

Une fois les données pré-traitées et nettoyées, il convient de les encoder afin d'obtenir une représentation numérique du texte.

Dans cette recommandation, nous souhaitons conserver l'ordre des mots pour plusieurs raisons. Tout d'abord, nous prenons en compte le message complet et n'essayons pas de séparer les phrases du corpus, ce qui diffère de l'approche utilisée pour la prédiction décrite au chapitre 3. Ce choix est motivé par la difficile séparation en phrases dans un corpus de conversations informelles, non standardisés. Il devient important de traiter différemment les termes apparaissant à la fin d'un message. Les sacs de mots et les sacs de caractères ne sont pas privilégiés dans la représentation des données choisie car il ne permettent pas de conserver l'ordre du texte en se concentrant sur des matrices de décompte de mots. L'empilement de caractéristiques globales calculées est également ignoré, nous considérons que les plongements lexicaux obtiennent de telles caractéristiques à partir du texte. De plus, les algorithmes utilisés précédemment ne prenaient pas en compte l'ordre des mots avec bien souvent l'usage de sacs de mots, ou de caractères. Ici nous varions les algorithmes de prédiction afin de pouvoir, le cas échéant, prendre en compte l'ordonnancement quand cela est possible. Une méthode permettant de représenter une suite de mots ordonnée est la mise en place de plongements lexicaux (*Word2Vec*) (MIKOLOV, CHEN et al., 2013), combinée à la vectorisation du texte en indices de mots.

Les plongements lexicaux sont abordés de deux manières : en amont pour l'ensemble des algorithmes, sauf pour les réseaux de neurones profonds pour lesquels les plongements lexicaux sont effectués via une couche cachée lors de l'entraînement. Le format pivot commun aux deux approches est la vectorisation du texte en indices de mots. Ces indices (nombres entiers positifs) sont alors transformés en vecteurs en utilisant le modèle skip-gram *Word2Vec* (REHUREK et SOJKA, 2010 ; MIKOLOV, SUTSKEVER et al., 2013), ou dans une couche cachée de plongements lexicaux transformant chaque index en vecteur de taille fixe. Les plongements lexicaux sont appliqués sans seuil de fréquence minimale contrairement à ceux du chapitre 4, afin de prendre en compte la totalité du corpus relativement petit de messages instantanés privés.

## 5.4. Élagage et remplissage des messages

La taille des vecteurs doit être fixe et uniforme pour pouvoir servir d'entrée d'un réseau. Devant cette nécessité, il convient de prendre en compte deux facteurs : la longueur des messages et la taille des vecteurs de chaque mots, définie dans les hyper-paramètres de plongements lexicaux. Le premier facteur, la longueur des messages, est fixé à 15 mots/termes qui correspond à la moyenne de longueur des messages du corpus comme l'ont fait Nguyen *et al.* (NGUYEN et GRISHMAN, 2015). Ainsi l'élagage (*trim*) permet de couper les messages ayant une taille supérieure à 15. Il reste cependant à définir quelle partie du message enlever. Il existe plusieurs méthodes d'élagage en coupant la bordure gauche du message (*left trim*) ou droite (*right trim*). Puisque nous voulons pouvoir mettre en avant la proximité dans l'ordonnancement des mots du message, nous appliquons un élagage gauche qui supprime le contenu le plus ancien, c'est-à-dire le texte le plus à gauche pour une langue orientée de gauche à droite comme l'anglais. Le même principe est appliqué pour les messages ayant une longueur inférieure à 15 : un élément unique et factice sera répété jusqu'à l'obtention de la longueur adéquate, aussi appelée remplissage égal (*same padding*). Ce procédé se résume à travers l'algorithme 1.

## 5.5. Prédiction de catégories d'emojis

Afin de recommander à l'utilisateur des emojis par catégories, nous mettons en place un système prédictif à partir d'une classification supervisée mais, cette fois-ci, mono-étiquette. L'objectif de ce classifieur est de classer chaque message dans l'une des 18 catégories définies au chapitre 4. Plus précisément, nous ne

**Algorithme 1** Élagage et remplissage utilisés pour normaliser la taille des vecteurs.

**Prérequis:** $m$ = message , $L$ = longueur du message, $M = \frac{1}{n}\sum_{i=1}^{n} L_i$
  **Pour chaque** $M$ **faire**
    **Si** $L < M$ **alors**
      **Tant que** $L < M$ **faire**
        Supprimer premier élément
    **Sinon**
      **Tant que** $L > M$ **faire**
        Insérer premier élément factice

nous occupons que de l'ensemble des catégories d'emojis obtenues automatiquement (Tableau 4.5), l'ensemble des catégories de références n'ayant eu qu'un intérêt d'évaluation des groupes obtenus vis-à-vis des catégories d'Ekman (EKMAN, 1999). Cette méthode permet d'obtenir une chaîne de traitements automatisée de bout en bout.

Pour obtenir une prédiction optimale, nous avons exploré plusieurs algorithmes de classification supervisée avec une attention particulière portée aux algorithmes fondés sur des arbres de décision ou des réseaux de neurones.

**Boosting d'arbre de décisions.** Le *boosting* d'arbre de décision (FRIEDMAN, 2001) est considéré afin de comparer un algorithme de classification transparent, contrairement à d'autres algorithmes dits "boîte noire" tels que les réseaux de neurones profonds, où l'interprétation reste difficile. Avant l'avènement de ces derniers, le *boosting* d'arbres de décision a longtemps été utilisé pour obtenir les meilleurs résultats dans diverses compétitions [2]. Comme lors de la prédiction d'emojis, l'usage d'un algorithme à base d'arbres de décisions présente certains avantages et inconvénients : la meilleure compréhension du modèle et l'impossible prise en compte de l'ordonnancement des mots en sont des exemples respectifs. Ce type d'algorithme a été comparé (Tableau 5.1) à partir d'un vote majoritaire entre classifieurs comprenant des forêts d'arbres de décisions, une régression logistique, des réseaux bayésiens naïfs et un algorithme de boosting d'arbre de décisions. Pour ce dernier nous avons utilisé les hyper-paramètres suivants :

— 16 feuilles par arbre
— Profondeur illimitée
— Taux d'apprentissage : 0.1
— Nombre d'arbres : 1 000

2. http ://blog.kaggle.com/2017/01/23/a-kaggle-master-explains-gradient-boosting/

— Type de boosting : boosting du gradient

Ces hyper-paramètres ont été définis empiriquement par validation croisée à 5 plis sur le corpus entier. Pour pouvoir rendre possible la comparaison entre les paramètres, la validation croisée utilisée est échantillonnée aléatoirement et stratifiée. Le choix des paramètres suit une recherche quadrillée (*grid search*) focalisée sur les paramètres listés plus haut. Il convient enfin de préciser que nous utilisons l'implémentation *Light Gradient Boosting Machine* [3] (LGBM) (KE, MENG et al., 2017) pour obtenir les résultats présentés au tableau 5.3.

**Réseaux de neurones profonds.** Les réseaux de neurones profonds sont utilisés dans le but principal de comparer le meilleur algorithme transparent obtenu par un système de vote à l'aide de la méthode précédemment présentée, avec l'apprentissage profond. Plus qu'une simple comparaison, nous cherchons à savoir si le recours à des algorithmes neuronaux s'avère judicieux et utile pour obtenir une recommandation d'emojis.

Pour mettre en place un réseau adapté à la tâche, nous partons dans un premier temps des travaux de Kim (KIM, 2014), ces derniers ayant proposé une architecture de réseaux de neurones convolutionels pour la classification de séquence textuelles. Nous l'utilisons, avec leurs paramètres, comme point de départ pour définir un modèle de classification de messages instantanés, auquel nous ajoutons une couche de réseaux récurrents. Ce modèle final, mis en place à l'aide de Keras [4] (CHOLLET et al., 2015) et de Tensorflow [5], est visible au tableau 5.1 et diffère du système proposé par Kim de par les paramètres choisis et le nombre de couches.

**Couche de plongements lexicaux.** Premièrement, nous appliquons une couche de plongements lexicaux sur les données encodées qui correspondent à des listes de listes d'indices relatifs au vocabulaire observé. Cette couche de plongements lexicaux produit des vecteurs de dimension $15 * 64$ dont la première valeur est dépendante du processus précédent d'élagage et de remplissage (Section 5.4), tandis que la seconde est définie empiriquement. En effet, en variant la taille des vecteurs issus des plongements lexicaux, le résultat final est légèrement amélioré ($+ - 1\%$ en macro f-mesure) mais influe énormément sur la mémoire utilisée lors de l'apprentissage, qui empêche alors d'attribuer de nombreuses unités aux couches cachées suivantes. Il convient également de noter que les vecteurs ne sont pas statiques et sont voués à être modifiés et ajustés durant l'entraînement.

**Couche de convolution.** La couche de convolution (LECUN, BENGIO et al., 1995)

3. https://github.com/Microsoft/LightGBM
4. https://keras.io/
5. https://www.tensorflow.org/

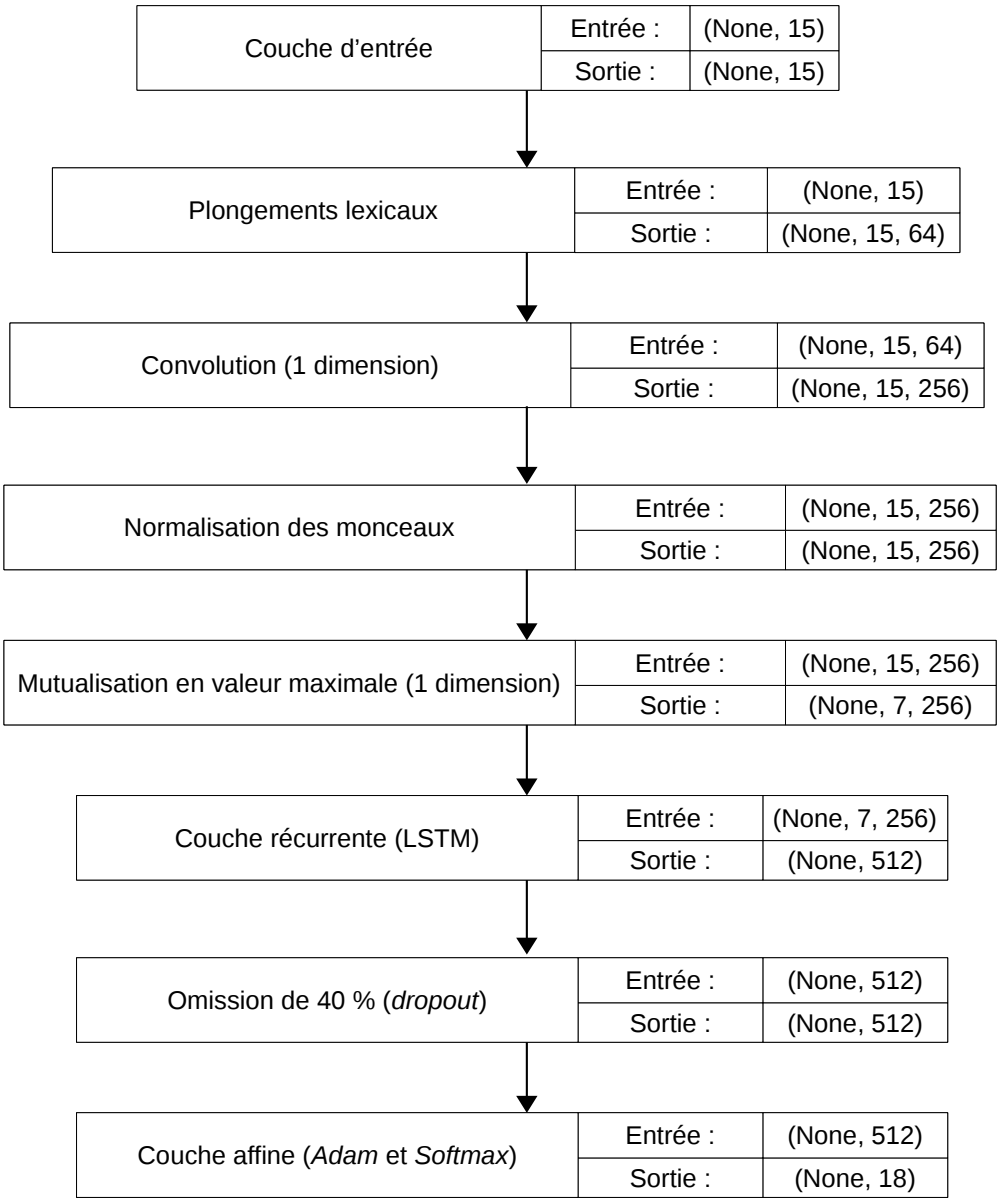


Figure 5.1. – Réseau CNN-LSTM utilisé pour prédire les catégories d'emojis.

utilisée est une couche à une dimension afin de pouvoir prendre en compte des vecteurs de deux dimensions. En effet chaque mot est représenté par un vecteur $w_i$ à une dimension de taille 64. Chaque message est une suite de vecteurs de mots $w_i$, un vecteur de message $m_i$ possédant alors deux dimensions, 15 sur 64. La couche de convolution prend alors en compte l'ensemble du texte, représenté par ce vecteur à deux dimensions. Cette couche est initialisée uniformément en suivant une fourchette de valeurs définie par $\sqrt{\frac{6}{(UET+UST)}}$ où UET signifie

"Unités d'Entrée du Tenseur" et UST signifie "Unités de Sortie du Tenseur", également appelée l'initialisation uniforme de Glorot (GLOROT et BENGIO, 2010). Cet hyper-paramètre est choisi à l'aide d'un quadrillage dont les différentes valeurs possibles sont visibles au tableau 5.2 et n'a apporté qu'un gain de 0.1% par rapport à la même architecture avec une initialisation aléatoire. L'usage de cette couche de convolution a pour objectif de récupérer des informations que nous aurions pu obtenir à l'aide de $n$-grammes sur ces vecteurs constamment ajustés durant l'entraînement. Ainsi, nous utilisons 256 fenêtres de convolution de taille 2 (*kernel size* ou window size) qui prendront donc en compte un couple de termes simultanément en glissant d'un mot à l'autre (*stride* à valeur 1), et qui est rectifiée en mettant à zéro toutes les valeurs négatives lors de l'activation $Y = max(0, X)$ (*Rectifier Linear Unit*). Puisqu'une normalisation des monceaux (*batch normalization*) est ensuite appliquée, nous choisissons un rembourrage (*padding*) de sorte que les valeurs en sortie aient la même longueur que celles en entrée, soit une longueur de 15. Ce paramétrage donne le comportement représenté en figure 5.2.

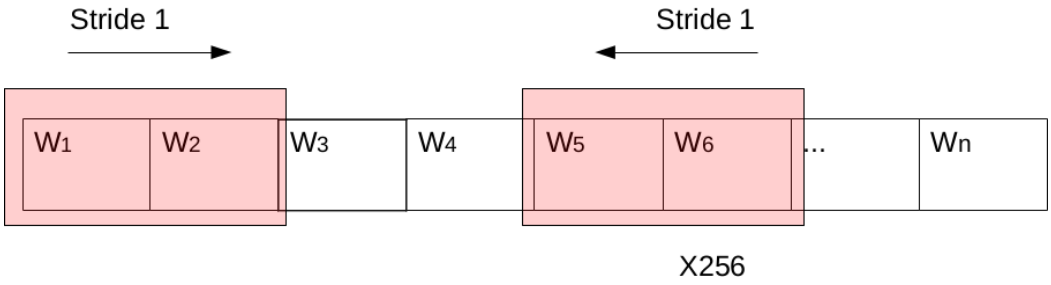


Figure 5.2. – Comportement de la convolution mise en place.

| Paramètre | Valeurs possibles |
|---|---|
| Fonctions d'optimisation | RMSProp (TIELEMAN et HINTON, 2012), Adam (KINGMA et BA, 2014) |
| Initialisations | 'glorot uniform', 'normal', 'uniform' |
| Fonctions d'activations | 'sigmoid', 'tanh', 'relu' |
| Taille des monceaux | 16, 32, 64 |

Tableau 5.2. – Quadrillage ciblé d'hyper paramètres pour le CNN-LSTM. Les résultats utilisés sont des moyennes d'une validation croisée à 5 plis.

**Normalisation des monceaux et mutualisation.** La normalisation des monceaux de données (ou *batch normalization*) est ensuite uniquement ajoutée afin d'accélérer l'entraînement, puisqu'en normalisant la sortie de la couche de convolution autour de 0, elle permet de réduire la variance et ainsi introduire un pas

d'apprentissage (*learning rate*) plus grand. En revanche, l'utilisation d'une mutualisation (*pooling*) de ces données normalisées nous permet d'introduire l'idée principale derrière l'architecture de ce réseau : les couches de convolution et de mutualisation ne servent qu'à obtenir une représentation différente des données et de leurs relations (représentées par la taille des fenêtres) pour l'utilisation de réseaux récurrents. À ce titre, la mutualisation par valeur maximale est utilisée, ainsi chaque couple de vecteurs de mot est représenté par celui ayant la valeur la plus élevée. Le résultat qui en ressort est donc un vecteur à deux dimensions de longueur 7 et de profondeur 256.

**Couche récurrente.** L'objectif étant de pouvoir prendre en compte l'ordre des mots lors de la recommandation, nous ajoutons une couche récurrente avec des *Long Short Term Memory* (HOCHREITER et SCHMIDHUBER, 1997) (LSTM). Cette couche a pour entrée la représentation simplifiée issue des couches précédentes. Cette couche récurrente prend en compte les relations entre les couples de termes et elle ne suit donc pas exactement une logique de simple ordonnancement des mots mais plus précisément de l'ordonnancement des relations entre mots. Au final, toutes les étapes et couches précédentes ne servent qu'à remplacer une sélection de caractéristiques discriminantes (*feature selection*) en changeant la représentation des données. C'est pourquoi le but est que la sortie de ce réseau mette en avant les derniers éléments, uniquement s'ils sont jugés pertinents par le système. Le nombre d'unités a été choisi empiriquement lors de la mise en place de cette architecture afin de contre-balancer la relative petite taille des plongements lexicaux issus de la première couche. Cela est principalement dû à des limitations matérielles quant à la mémoire disponible lors de l'apprentissage.

**Couches de sortie.** Les dernière couches sont l'application d'un *dropout* qui consiste à ignorer 40% des données au hasard afin de limiter le sur-apprentissage, ce dernier est ainsi évité principalement par le *dropout* et la mutualisation. Bien qu'il est commun d'effectuer un *dropout* à 0.5, enlever 50% des informations nous paraît quelque peu exagéré compte tenu de la taille limitée du corpus, justifiant ainsi une suppression moindre. Ce choix a été déterminé empiriquement entre les valeurs de *dropout* de 0,3, 0,4 et 0,5. La dernière couche affine (ou "dense") est la couche de classification et possède une optimisation par *Adam* avec les paramètres par défaut de l'article originel (KINGMA et BA, 2014), soit un pas d'apprentissage de $\alpha = 0.001$. Cette fonction d'optimisation est une extension de la descente stochastique du gradient (KIEFER, WOLFOWITZ et al., 1952) adaptant le pas d'apprentissage en combinant l'adaptation du pas d'apprentissage des paramètres améliorant les performances sur les gradients épars de l'*Adaptative Gradient Algorithm* (AdaGrad) (DUCHI, HAZAN et al., 2011), avec l'adaptation du pas d'apprentissage des paramètres changeant rapidement issue du *Root Mean Square Propagation* (RMSProp) (TIELEMAN et HINTON, 2012) pour mieux prendre en compte le bruit. L'activation est faite par *softmax* qui trans-

forme les scores en probabilités, soit la probabilité qu'un élément $y$ appartienne à une classe $k$ donnée à partir de la matrice de poids $W$ et de l'entrée $X$. Ceci nous donne :

$$P(y = k|W, X) = \frac{exp(f_k)}{\sum_{j=0}^{C-1} exp(f_j)} = softmax(f, k) \tag{5.1}$$

La fonction de perte (*loss*) utilisée est celle des *categorical cross entropy* qui est couramment utilisée pour de la classification muti-classes.

| **Classifieur** | **Précision** | **Rappel** | **F-Mesure** |
|---|---|---|---|
| Random Clf | 05.68 | 05.98 | 04.70 |
| Majority Clf | 00.97 | 05.55 | 01.66 |
| LGBM | 50.10 | 31.28 | 36.16 |
| CNN-LSTM | 55.82 | 52.03 | 53.10 |

Tableau 5.3. – Moyennes macro de la précision, du rappel et de la F-Mesure pour chaque classifieur. C'est moyennes sont issues d'une validation croisée à 5 plis.

Le tableau 5.3 montre les résultats obtenus à l'aide d'une validation croisée à 5 plis stratifiés en utilisant les deux algorithmes présentés précédemment mais également deux autres bases de comparaison. La première, *Random Clf*, est un classifieur aléatoire qui attribue une classe aléatoirement de manière égale et non pondérée à chaque message. Ce classifieur est non seulement une bonne base de comparaison pour savoir si notre approche est nécessaire mais également une indication en plus sur la difficulté de la tâche. La seconde, *Majority Clf*, est un classifieur majoritaire qui attribue tout le temps la même classe, ou le même ensemble de classes aléatoires, en prenant uniquement en compte les classes sur-représentées dans le corpus d'entraînement. Cette approche permet de déterminer si un classifieur est inutile puisqu'il suffirait d'attribuer uniquement les classes majoritaires, cas probable lorsque l'on possède une distribution déséquilibrée des données. C'est notre cas avec 3 catégories qui sont sur-représentées : les emojis (😘 😙 😘 😗 😙), la catégorie hilare (😂 😭) qui a été étiquetée comme ambiguë mais qui correspond davantage à l'hilarité, et donc à la joie, qu'aux pleurs de par leur taux d'usage et l'ambiguïté du second emoji souvent utilisé dans des cas similaires, et enfin l'emoji 😍 qui est seul dans sa catégorie. On peut observer que ces deux classifieurs obtiennent des résultats très bas, surtout une fois comparés aux *boosting* d'arbre de décision (LGBM) ou à notre réseau (CNN-LSTM). Seul le CNN-LSTM obtient un score au dessus de 50% en macro f-mesure, le LGBM ayant un rappel bas. Les LGBM donnent une précision et un rappel moindre, le score de rappel étant très éloigné des CNN-LSTM. Il apparaît clairement que le

réseau de CNN-LSTM est l'algorithme de classification le plus efficace pour notre tâche et, bien qu'il s'agisse d'un réseau simple de 5 couches, d'autres alternatives sans apprentissage profond essayées lors du vote entre les classifieurs ne fonctionnant pas aussi bien sur cette tâche.

**Ajustements.** Plusieurs ajustements d'hyper-paramètres ont été tentés avant d'être abandonnés :

— **Le pas d'apprentissage.** La gestion du pas d'apprentissage (*learning rate*) n'est jamais triviale puisqu'elle dépend grandement de la fonction d'optimisation mais aussi, dans notre cas, de l'usage ou non de normalisation des monceaux. Pour cela nous avons mis en place un déclin automatique du pas d'apprentissage afin que celui-ci puisse être élevé de $\alpha = 0.1$ au début et diminuer au fur à mesure du nombre d'itérations $e$ sur le corpus d'entraînement de la manière suivante : déclin = $\frac{\alpha}{\frac{e}{e/2}}$. Les résultats n'ont pas dépassé ceux essayés lors du quadrillage avec un macro score de 19,51% qui s'explique par la courbe de perte (*loss*) ne cessant d'augmenter.

— **Nombre d'itérations complètes.** Nous avons constaté une amélioration moins efficace au fur et à mesure des itérations complètes sur le corpus d'entraînement (*epochs*). La mise en place d'un mécanisme de détection de l'inefficacité de l'apprentissage pour l'arrêter automatiquement (*early stopping*) a donc été envisagé. Cependant, lors de l'apprentissage la courbe de perte a un effet de rebondissement comme on peut le voir dans la figure 5.3, effet qui rend difficile l'application d'un tel mécanisme.

**Déséquilibre des données.** Les données étant réparties par catégories, la distribution s'en retrouve moins déséquilibrée que lors de la prédiction directe d'emojis du chapitre 3. Toutefois, comme montré plus haut avec l'exemple des 3 catégories sur-représentées, leur répartition reste inégale et nous avons donc essayé d'y remédier. Les approches sont les mêmes que pour la prédiction directe d'emojis : sur et sous échantillonnage des données ainsi qu'une tentative de mise à l'échelle des données (*scaling*). Les résultats sont les mêmes en ce sens que ces approches de gestion de données déséquilibrées n'entraînent pas une amélioration du modèle, en plus d'amener un risque de dénaturer les données d'entrée.

## 5.6. Impact de la temporalité

Ce que nous appelons temporalité est la capacité du réseau à prendre en compte l'ordre des mots et surtout le fait que certains mots du contexte gauche sont éloignés, moins récents que les mots les plus proches du mot actuel. Bien que le modèle CNN-LSTM se révèle être le plus efficace en terme de métriques, rien n'indique l'impact de l'ajout d'une couche de LSTM sur le résultat. Pour vérifier cela, nous mettons en place le même processus sur trois architectures dif-

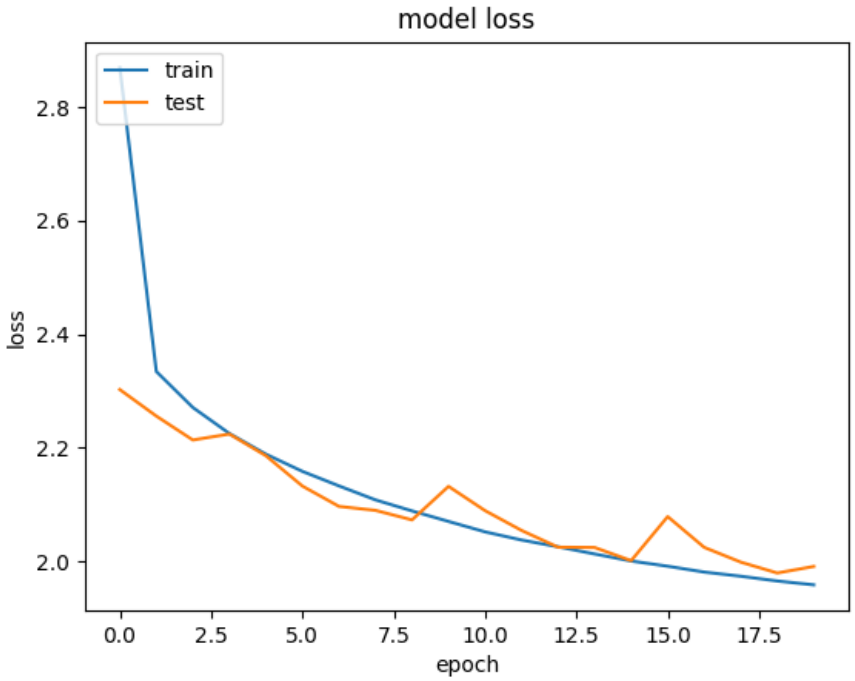


Figure 5.3. – Courbe de perte du CNN-LSTM sur 20 itérations.

férentes visibles en figure 5.4. La première est une seule couche de convolution avec les mêmes approches de normalisation et de mutualisation (CNNx1) et la seconde consiste en un remplacement de la couche de LSTM par une deuxième couche de convolution (CNNx2) afin de déterminer la différence d'impact entre une seconde couche de CNN et une de LSTM prenant en compte l'ordre des mots. La troisième et dernière architecture correspond à deux couches de Bi-LSTM (GRAVES, MOHAMED et al., 2013) entraînées en 50 itérations choisies par arrêt préventif automatique.

Le tableau 5.5 montre les résultats obtenus en utilisant ces trois architectures avec l'ajout du score de *Mean Reciprocal Rank* (MRR) (RADEV, QI et al., 2002) qui représente la capacité du classifieur à donner le bon résultat à la première position parmi les probabilités fournies pour chaque classe. À partir de ces scores obtenus nous pouvons constater que la couche de convolution seule obtient déjà de bons résultats. De plus, ajouter une seconde couche de convolution en lieu et place des LSTM utilisés dans notre architecture principale (CNN-LSTM) entraîne une baisse de précision même si les résultats de F-mesure et MRR sont améliorés. L'impact des LSTM est important en ce sens qu'ils permettent la prise en compte de la proximité et du contexte précédent éloigné là où l'empilement de couches de convolution ne donne qu'une indication de voisinage comme le font des $n$-grammes. Aussi, les Bi-LSTM sont notamment utilisés pour l'étique-

| Nom | CNNx1 | CNNx2 | Bi-LSTM |
|---|---|---|---|
| **Couche 1** | Plongements lexicaux | Plongements lexicaux | Plongements lexicaux |
| **Couche 2** | Convolution | Convolution | Bi-directionnel |
| **Couche 3** | Normalisation des monceaux | Normalisation des monceaux | Bi-directionnel |
| **Couche 4** | Mutualisation Max. | Mutualisation Max. | Omission (40 %) |
| **Couche 5** | Omission (40 %) | Convolution | Couche affine |
| **Couche 6** | Couche affine | Normalisation des monceaux | |
| **Couche 7** | | Mutualisation Max. Globale | |
| **Couche 8** | | Omission (40 %) | |
| **Couche 9** | | Couche affine | |

Tableau 5.4. – Architectures connexes pour estimer l'impact du LSTM sur la sortie du CNN.

| Classifieur | Precision | Rappel | F-mesure | MRR |
|---|---|---|---|---|
| **CNN-LSTM** | 55.82 | 52.03 | 53.10 | 68.87 |
| **CNNx1** | 50.52 | 41.94 | 44.75 | 64.57 |
| **CNNx2** | 48.68 | 46.07 | 46.32 | 66.91 |
| **Bi-LSTM** | 51.60 | 36.93 | 41.33 | 62.21 |

Tableau 5.5. – Comparaison des performances avec des architectures connexes.

tage de séquences (HUANG, W. XU et al., 2015) puisqu'ils permettent une prise en compte d'un ordonnancement à double sens de la phrase, ce qui améliore la prise en compte du contexte. Cependant, les résultats en rappel sont en deçà des autres architectures, ce qui peut s'expliquer par l'ordre à sens unique (de gauche à droite) dont fait preuve l'écriture d'un message : considérer deux directions permettrait alors d'obtenir uniquement des informations générales de séquences, tout comme le font très bien les couches de convolution. Il convient également de rappeler que les emojis sont majoritairement utilisés en fin de message ou de phrase, renforçant ainsi la nécessité d'une prise en compte du sens d'écriture dans la prédiction des catégories d'emojis. Cette approche sied donc à la prédiction du rôle de l'emoji en tant qu'ajout de méta-informations (Chapitre 1).

## 5.7. Utilisation du système prédictif pour une recommandation

L'objectif de ce système prédictif est d'utiliser les groupes d'emojis comme des sous-ensembles de choix possibles laissés à l'utilisateur. C'est dans ce sens que les métriques utilisées ont été choisies : les macro scores permettent de favoriser la bonne prédiction de toutes les catégories plutôt que des prédictions trop ciblées, tandis que le MRR permet de connaître le taux de confiance dans l'ordre décroissant des probabilités du classifieur. Ainsi, en regardant les scores détaillés par classe d'une des prédictions du CNN-LSTM nous pouvons déterminer l'efficacité de la prédiction pour chaque catégorie d'emojis.

| Catégorie | Précision | Rappel | F-Mesure | Support |
|---|---|---|---|---|
| 😴 | 0.74 | 0.69 | 0.72 | 764 |
| 🤐 | 0.48 | 0.29 | 0.36 | 213 |
| 😙 😘 😗 😚 | 0.49 | 0.55 | 0.52 | 9458 |
| 😔😢😓😥😤😖😭😣😫 | 0.62 | 0.58 | 0.60 | 3768 |
| 😄 😃 | 0.54 | 0.63 | 0.58 | 701 |
| 😞 😔 | 0.56 | 0.50 | 0.53 | 1548 |
| 😊 | 0.65 | 0.40 | 0.49 | 717 |
| 😂 😭 | 0.71 | 0.58 | 0.64 | 8043 |
| 😒 😤 👺 | 0.63 | 0.55 | 0.59 | 1046 |
| 😅 😬 😀 😝 😜 | 0.38 | 0.47 | 0.42 | 3405 |
| 🤑 | 0.48 | 0.56 | 0.52 | 618 |
| 😊 😌 🙂 😊 | 0.50 | 0.58 | 0.54 | 3325 |
| 😍 | 0.43 | 0.36 | 0.39 | 8179 |
| 😄 😆 😂 | 0.40 | 0.59 | 0.48 | 2848 |
| 😏 😎 😉 | 0.48 | 0.58 | 0.53 | 4286 |
| 😀 😕 ☹ 😟 | 0.69 | 0.47 | 0.56 | 883 |
| 🤯😱😮😳😲😰😳😦 | 0.61 | 0.44 | 0.51 | 1522 |
| 😑 😶 😐 😒 😕 | 0.66 | 0.54 | 0.59 | 2596 |
| Moyenne pondérée | 0.54 | 0.52 | 0.52 | 53920 |

Tableau 5.6. – Résultats détaillés de la prédiction du CNN-LSTM par catégorie sur une séparation aléatoire du corpus. Les scores moyens sont pondérés en fonction du nombre d'occurrences (*support*).

Le tableau 5.6 montre les scores pondérés obtenus pour chaque catégorie. Nous pouvons constater que les trois catégories sur-représentées ne sont pas forcément les mieux prédites. En effet, la catégorie 😍 obtient des résultats relativement bas. Malgré cela, nous pourrions attendre du modèle qu'il rencontre

davantage de difficulté à prédire correctement des catégories nuancées dans le jeu d'étiquettes, à savoir la joie forte 😄 😆 😁, la joie modérée 😄 😃 et la joie faible mélangée aux plaisirs sensoriels 😊 😌 🙂 😊. Toutefois leur prédiction reste relativement imprécise mais le taux de rappel compense cette imprécision. Force est de constater que plus une catégorie d'emojis est divisée en plusieurs sous-catégories nuancées, moins son taux de précision sera élevé. Les catégories de la colère 😒 😤 👺 ou encore de la tristesse 😓 😪 😔 😞 😰 😣 😭 😖 😫 en sont de bons exemples et, comme indiqué pour la classe 😍, le nombre d'occurrences ne suffit pas à justifier la qualité de la prédiction. La matrice de confusion visible en figure 5.4 confirme la difficile gestion des nuances par le classifieur, comme en témoigne le fait que la catégorie 😍 est très souvent confondue avec la catégorie 😙 😘 😗 😚, et ce, dans les deux sens entraînant ainsi une faible précision pour ces deux catégories. Le classifieur échoue donc à trouver des caractéristiques discriminantes pour ces deux classes.

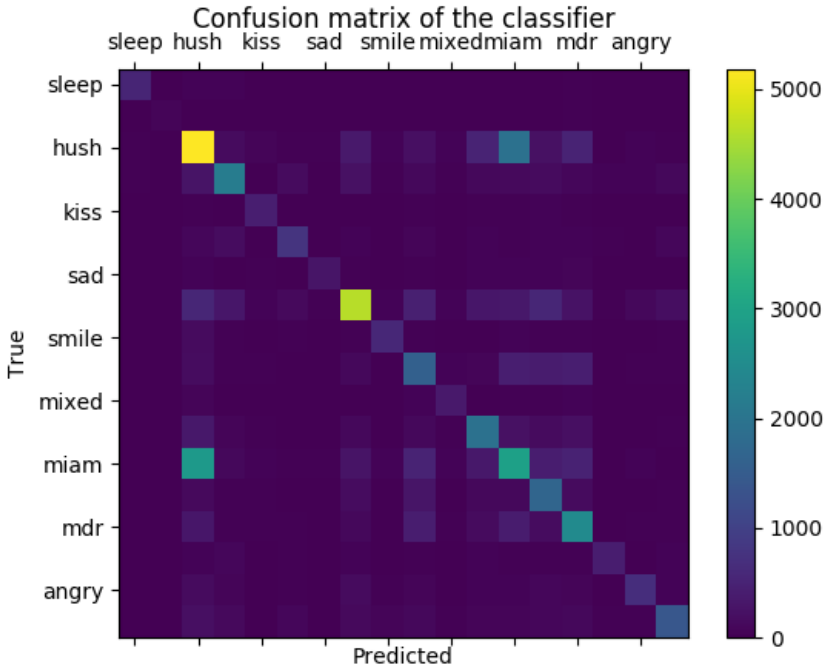


Figure 5.4. – Matrice de confusion du CNN-LSTM. L'ordre des étiquettes est le même que celui des scores détaillés 5.6, de gauche à droite et haut en bas.

Nous souhaitons utiliser ce modèle prédictif pour recommander des catégories d'emojis, chaque catégorie représentant un panel de choix restreints et orientés que l'utilisateur final pourra alors sélectionner. L'utilisation de la macro f-mesure

lors de l'entraînement, utilisée pour évaluer les monceaux du corpus de validation et ainsi ajuster le modèle, a donc pour but de construire un modèle prédictif qui n'ignore aucune catégorie possible et, compte tenu des scores détaillés du tableau 5.6, cela semble être le cas avec 14 catégories dépassant les 50% de f-mesure. Dans le chapitre précédent (Chapitre 4), nous avons proposé un modèle de classification multi-étiquettes lors de la prédiction directe d'emojis pour pouvoir prédire beaucoup d'étiquettes différentes. Cependant, nous souhaitons ici seulement proposer un ou deux panels de choix possibles à l'utilisateur, c'est-à-dire une ou deux catégories contenant plusieurs emojis. Pour ce faire les scores de prédictions de chaque catégorie sont nécessaires. De cette manière, nous pouvons proposer les deux premières catégories prédites et ainsi utiliser le système de prédiction mis en place comme un système de recommandation.

Le deuxième avantage d'un tel système est son adaptabilité : utiliser des catégories d'emojis, même définies automatiquement, se révèle être modifiable ou adaptable manuellement. En effet, comme précisé en introduction de ce manuscrit, de nombreux logiciels ou applications de messagerie possèdent leurs propres emojis propriétaires, il serait donc difficile d'introduire de tels emojis dans un système de prédiction fondé sur de l'apprentissage supervisé ne contenant pas d'exemples d'utilisation du nouvel emoji (🥲 par exemple). Ici, l'insertion manuelle d'un nouvel emoji dans une catégorie permet de pallier au démarrage à froid (*cold start*), problème récurrent aux débuts de tout système de recommandation.

## 5.8. Limites

L'usage de ce système prédictif pour la recommandation présente des limites qu'il convient d'aborder. Ces limites sont essentiellement liées au contexte logiciel et aux données utilisées.

### 5.8.1. Multiple usages des emojis dans le corpus

Étant donné que nous utilisons un système de classification supervisée, le système se base sur des étiquettes, des emojis, réellement sélectionnées par des utilisateurs. Les catégories d'emojis utilisées proviennent donc de regroupement sur des usages réels, ce qui semble être adéquat compte tenu de l'objectif visé. Toutefois, nous pouvons nous demander comment ces emojis ont été sélectionnés, ou encore à quoi ressemblerait l'interface de sélection de ces emojis. Un biais est alors inévitable puisque les conditions de sélection influencent le choix de l'emoji par l'utilisateur. Dans nos travaux présentés plus haut, nous pouvons signaler que comme ces données proviennent de messages d'une application mobile de

messagerie[6], le clavier utilisé influence la sélection, or il existe de nombreux claviers disponibles, certains proposant une suggestion d'emojis par mots-clés (Figure 5.5). Cette idée d'une disposition du clavier en tant qu'orientation première du choix des emojis fait d'ailleurs l'objet de travaux de recherche dédiés (POHL, DOMIN et al., 2017). De plus, l'application Mood Messenger possédant une barre prédictive d'emojis par mots clés, visible en figure 5.5, l'utilisation de cette barre influence très certainement l'usage des emojis obtenu dans ce corpus. Nous n'avons malheureusement pas accès à une information qui nous permettrait de toujours savoir dans quel cadre logiciel l'emoji a été utilisé. La seule donnée à notre disposition est l'indication d'un emoji en remplacement d'un mot, que nous ignorons volontairement ici.


Figure 5.5. – Interface initiale de suggestion d'emojis avec distance d'édition et lexique dans l'application source créant un biais dans l'obtention du corpus.

6. Mood Messenger

### 5.8.2. De la différence entre tweets et messages informels privés

La nature des données peut être considérée comme une seconde limite du système. Les données utilisées pour l'apprentissage du modèle prédictif sont des données privées dont l'obtention passe par un accord préalable pour une utilisation précise restreinte à l'entreprise. Bien qu'elles soient de nature plus proche aux besoins d'une recommandation d'emojis dans un cadre de conversation instantanée privée et que notre objectif est d'appliquer la recommandation d'emojis dans un tel contexte, la disponibilité de tweets en masse, plus faciles à récolter et à distribuer et sans besoin d'accord préalable, nous amène à comparer le système également sur ce type de données avec le même pré-traitement. La seule différence étant la suppression de mots dièse propres aux tweets. Pour ce faire, nous avons collecté 387 037 tweets uniquement en anglais avec la même méthode que lors de l'obtention automatique des catégories d'emojis (chapitre 4), à ceci près que ces tweets sont uniquement issus de conversations sur Twitter. En effet, les tweets pouvant être de nature différente, nous avons voulu enrichir notre corpus avec des tweets issus de conversations Twitter, des réponses à une discussion en cours, en ignorant automatiquement les tweets promotionnels ou factuels qui ont peu de chance de se retrouver dans le cadre d'une conversation instantanée privée. De tels tweets à éviter sont de type :

> Everyone ! ! 😍 Concert at @adress this evening ! ! 👍 👍 #hashtag #hashtag

Pour éviter de tels tweets nous avons uniquement regroupé les tweets de réponse à un fil existant, puis les avons mélangés avec le corpus de messages privés. Les résultats obtenus avec le réseau CNN-LSTM présenté plus haut ne sont pas satisfaisants : des macro score de 30,15 % en précision et 15,28 % en rappel pour une moyenne harmonique de 18,09 %. Au travers des scores détaillés du tableau 5.6, ces résultats s'expliquent par la sur-représentation de la catégorie 😂 😭, ce qui n'est pas surprenant compte tenu de son fort usage sur twitter[7]. Toujours est-il que l'ajout de telles données perturbe la capacité du classifieur à trouver les caractéristiques discriminantes nécessaires à une bonne classification.

Cette différence entre tweets et messages informels privés mérite d'être approfondie en mettant en place un système dédié uniquement à ce type de tweets pour la recommandation d'emojis. Les scores obtenus en mélangeant les sources ne sauraient suffire à considérer de tels tweets comme inadéquats à l'apprentissage d'un modèle de recommandation d'emojis pour une utilisation dans un contexte informel privé. Ces tweets peuvent alors servir de corpus d'entraînement et de validation, pour ensuite tester sur les données de messages privés.

7. http://www.emojitracker.com/

## 5.9. Conclusion

Dans ce chapitre nous avons mis en place une comparaison de différents systèmes de classification par apprentissage supervisé afin de prédire les catégories d'emojis précédemment obtenues automatiquement. En montrant l'impact de différentes architectures de réseaux de neurones nous avons montré l'importance de la prise en compte de la temporalité, avec la couche de LSTM, dans la prédiction des catégories d'emojis. Par cette approche nous contribuons en montrant comment un système de classification supervisée appliqué à des fins de recommandations sied à la recommandation d'emojis, qui comporte beaucoup de redondance dans leur manière d'être sélectionnés (Chapitre 1). Cette approche permet une recommandation adaptative à de nouveaux éléments et constitue ainsi une solution à un problème connu dans le domaine de la recommandation : le démarrage à froid, c'est-à-dire le lancement initial de la recommandation sans aucune utilisation de référence. Pour orienter notre système vers ce but final, et non uniquement pour obtenir la meilleure prédiction possible, nous avons utilisé des métriques adaptées, comme la macro f-mesure ou encore le taux de MRR, qui focalisent l'évaluation de l'homogénéité de la prédiction, amenant ainsi à conserver plusieurs possibilités.

Le système présente certaines limites comme l'imparfaite prise en compte des nuances entre deux catégories d'une même émotion basique au sens d'Ekman (EKMAN, 1999), mais dont l'intensité varie. La dépendance du système à la nature des données est également une limite de notre approche. Comme solutions possibles, l'amélioration de la prise en compte des nuances par le système pourrait ainsi passer par un mécanisme d'attention (K. XU, BA et al., 2015) sur l'intensité des mots porteurs de sentiment pour aider le système à distinguer les catégories fines, tandis qu'une amélioration du filtrage de tweets pourrait peut-être permettre d'obtenir des données plus corrélées à la réalité de la messagerie instantanée privée. La quantité de données utilisées étant relativement petite, des algorithmes plus interprétables pourraient alors être favorisés tels que les LGBM, répondant ainsi au besoin de comprendre et de pouvoir interpréter les raisons de chaque recommandation.

Enfin, bien que nous ayons utilisé des métriques pensées pour la recommandation, l'évaluation objective par mesures de performances du système ne peut être complètement fiable : des choix possibles et corrects non présents dans le corpus d'entraînement ne peuvent être pris en compte. Il convient alors d'évaluer le système par le biais de l'utilisateur au travers d'une évaluation subjective, ce qui est l'objet du chapitre suivant.

## 5.10. Synthèse

**Objectifs** : recommander à l'utilisateur des emojis sous forme d'un ensemble de choix en prenant en compte l'ordonnancement des mots.

**Approche méthodologique** : classification par apprentissage profond pour la recommandation appris sur des messages instantanés privés. CNN-LSTM choisi en comparaison avec d'autres algorithmes : régression logistique, boosting d'arbre de décision, *etc.*

**Résultats** : les 18 catégories sont prédites avec une macro f-mesure de 52%, la plupart des catégories étant correctement prédites malgré le déséquilibre des données. Le système montre une corrélation entre les différentes granularités : des catégories aux sens proches auront plus de chance d'être confondues.

**Limites** : les emojis sources proviennent d'un contexte mixte de suggestion et de sélection d'emojis, ce qui entraîne une évaluation objective pas entièrement représentative. L'approche méthodologique est moins adaptée sur des tweets malgré le filtre préalable.

**Contribution** : premier système de recommandation d'emojis fondé sur des groupes d'emojis. Les classes à recommander sont obtenues automatiquement, ce qui en fait un système *end-to-end*. La recommandation est directement issue de l'usage dans un contexte conversationnel privé.

# 6. Évaluation subjective par l'utilisateur

## Plan du chapitre


6.1 Introduction **127**
6.2 Évaluation du système de recommandation d'emojis pour l'enrichissement extra-linguistique **128**
6.2.1 Conception de la plateforme de prédiction d'emojis 129
6.2.2 Résultats de l'évaluation du système de recommandation par prédictions combinées d'emojis 131
6.3 Évaluation du système de recommandation automatique de catégories d'emojis **132**
6.3.1 Conception de la plateforme de conversations instantanées 132
6.3.2 Résultats 135
6.4 Limites **139**
6.5 Conclusion et perspectives **140**
6.6 Synthèse **142**


## 6.1. Introduction

L'évaluation des systèmes est un problème récurrent dans le domaine de la recommandation (MASSA et AVESANI, 2007). Trois approches se distinguent pour leur évaluation : une évaluation hors-ligne, par étude utilisateur (*user study*) ou en ligne. Dans ce manuscrit les chapitres 3 et 5 représentent l'évaluation hors-ligne dans laquelle l'utilisateur n'a pas un rôle direct. Le système de recommandation a ainsi été évalué à partir d'indices objectifs correspondant à des métriques de performance. Dans ce chapitre l'attention est portée sur l'évaluation subjective, ce qui correspond aux évaluations par étude utilisateur ou en ligne, dans laquelle l'utilisateur évalue directement la recommandation fournie. Plus précisément, nous abordons l'évaluation subjective par le biais d'une étude utilisateur.

Une évaluation en ligne nécessite d'intégrer la recommandation dans un système existant sans que l'utilisateur ne soit au courant qu'il est en train d'évaluer le système, tandis qu'une évaluation par étude utilisateur consiste à reproduire le comportement de l'utilisateur en lui demandant explicitement d'interagir avec le système de recommandation pour l'évaluer. Les deux approches sont coûteuses

comparées à une évaluation hors-ligne et, surtout, la simulation du contexte lors de l'étude utilisateur est considérée comme difficile (RICCI, ROKACH et al., 2011). Cette reproduction, ou simulation, du contexte peut par exemple passer par la présentation d'un exemple de sortie de la recommandation dans un contexte précis, accompagnée d'une question directe à l'utilisateur. Elle a l'avantage de permettre aisément de varier les sujets de la recommandation tel que le type d'emoji, mais également de permettre de varier les contextes de recommandation, en indiquant par exemple un état d'esprit préalable à l'usage de l'emoji. L'autre méthode serait de mettre en place un questionnaire de satisfaction pour plusieurs cas de recommandation du système, en générant le texte sur lequel la recommandation est fondée. Toutefois, les emojis sont des éléments à recommander très subjectifs avec un usage dont la granularité des catégories peut varier davantage que pour un film ou un roman par exemple. Il convient donc de mettre en place une évaluation par étude utilisateur dans laquelle il lui est demandé d'intéragir directement avec le système de recommandation. De cette manière, l'étude utilisateur se rapproche le plus de l'évaluation en ligne qui s'avère être l'évaluation la plus proche d'une utilisation réelle (SHANI et GUNAWARDANA, 2011).

L'objectif est de pouvoir ajouter une évaluation concrète aux systèmes proposés tout en prenant en compte les spécificités de l'objet recommandé, les emojis. En effet, là où une recommandation a traditionnellement pour but d'influencer le comportement de l'utilisateur, la recommandation d'emojis vise uniquement à lui faciliter la tâche. Notons que comme souligné précédemment, le concept de nouveauté ou d'exploration propre aux systèmes de recommandation traditionnels n'est pas pertinent ; les emojis étant des objets à usage redondant (voir chapitre 1).

Deux systèmes d'évaluation par étude utilisateur sont présentés dans ce qui suit. Le premier est dédié à l'évaluation de la prédiction directe d'emojis à fonction d'enrichissement extra-linguistique issue du chapitre 3 (section 6.2). Le second système vise à évaluer la recommandation d'emojis par prédiction de catégories automatiques obtenues décrite dans les chapitres 4 et 5 (section 6.3). Leurs limites sont ensuite analysées (section 6.4).

## 6.2. Évaluation du système de recommandation d'emojis pour l'enrichissement extra-linguistique

La prédiction d'emojis à partir de l'anglais détaillée au chapitre 3 a fait l'objet d'une démonstration (GUIBON, OCHS et al., 2017) présentée ensuite dans le fo-

rum de valorisation en sciences humaines et sociales InnovativeSHS [1] du CNRS. La plateforme était disponible sous forme de borne tactile à accès immédiat à l'utilisateur avec une explication de son fonctionnement fournie oralement à l'utilisateur. Cette plateforme a pour objectif de permettre à l'utilisateur d'utiliser une combinaison de l'approche générative et discriminante (modèles présentés en chapitre 3) pour la prédiction d'emojis. L'utilisateur évalue ensuite le système comme présenté en section 6.2.1.

### 6.2.1. Conception de la plateforme de prédiction d'emojis

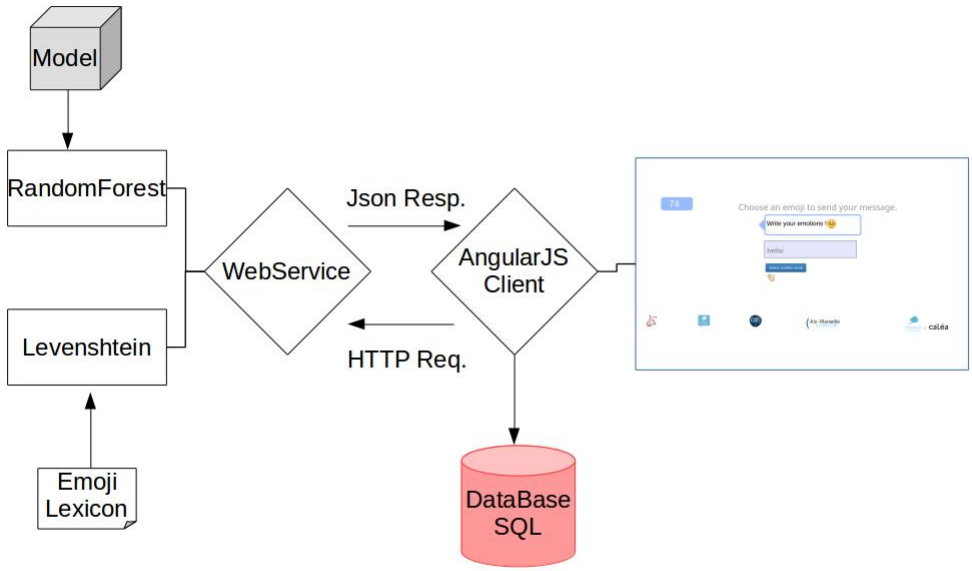


Figure 6.1. – Architecture globale de l'application

À partir de modèles génératifs et discriminants présentés au chapitre 3, une application permettant la recommandation en temps réel d'emojis a été développée. L'application est une SPA (*Single Page Application*) qui se détache en plusieurs modules distincts visibles en figure 6.1. Le premier est un service web de type REST [2] qui fournit le résultat de la prédiction combinée d'emojis. Cette

1. http://innovatives.cnrs.fr/presentation/actualites/article/innovatives-shs-salon-de-la-valorisation-en-sciences-humaine-et-sociales-17-18
2. Exemple : http://lsis-mood-emoji.lsis.org:9999/?query=hello&type=ed

prédiction fonctionne en utilisant les approches du chapitre 3 puis en fournissant dans l'ordre, le premier emoji de remplacement (Section 3.2), puis le ou les emojis issus de la classification muti-étiquettes apprise sur le corpus étendu ayant 1070 emojis (Section 3.3.3.2). Cette prédiction combinée est demandée par requête HTTP sous forme de JSON de la manière suivante :

```
1 {"status": "ok", "query": "super", "init": "", "emojis": [{"iosUrl"
    : "img/ios/_0t.gif", "eCode": "_0t"}, {"iosUrl": "img/ios/_bqn.
    gif", "eCode": "_bqn"}, ...], "type": "rf"}
```

Le second module est l'interface utilisateur qui affiche une liste d'emojis animés en fonction du texte tapé ou en cours de frappe. Un micro délai de 0,3 seconde d'inaction est respecté avant d'envoyer la requête du client au serveur de prédiction pour obtenir les emojis prédits. Ce délai est mis en place pour des raisons de fluctuation de la connexion lors de la démonstration. Enfin, le dernier module consiste en une simple base de données SQL.

L'utilisateur accède à l'application par l'interface visible en figure 6.2 dans lequel un décompte du nombre de messages envoyés est indiqué à l'utilisateur. Il n'y a volontairement pas de bouton d'envoi du message ni d'activation par la touche "entrée". Pour envoyer un message, l'utilisateur doit choisir un des emojis recommandés ou alors indiquer qu'aucun de ces emojis recommandés ne convient. Si tel est le cas, plusieurs choix lui sont alors possibles : envoyer le message sans emoji ou choisir parmi 7 emojis émotionnels présentés en figure 6.3.

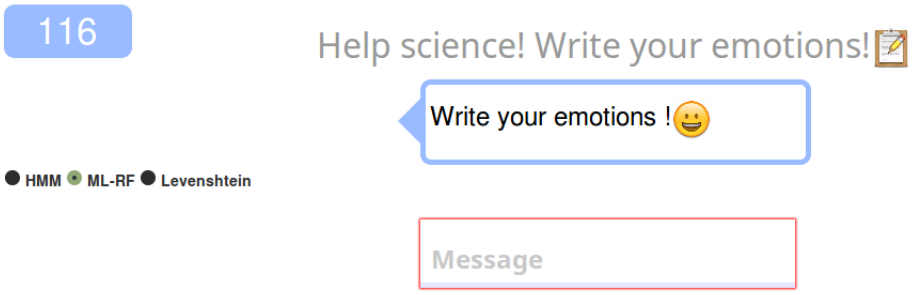


Figure 6.2. – Interface de test utilisateur. La prédiction utilisée reprend les modèles appris au chapitre 3.

Pour connaître et sauvegarder les emojis utilisés, chaque message enregistré est accompagné du texte, des emojis recommandés, de l'emoji finalement choisi si tel est le cas et de la façon dont cet emoji a été choisi. Une étiquette "validation" est ainsi donnée si un emoji recommandé est sélectionné, en partant du principe

qu'une telle sélection valide la recommandation. *A contrario*, une étiquette "tag" est donnée si l'emoji sélectionné ne provient pas de la recommandation mais de la sélection manuelle visible en figure 6.3. De cette manière, une indication de la qualité de la recommandation est obtenue. Nous présentons les résultats de cette évaluation plus précisément dans la section suivante.

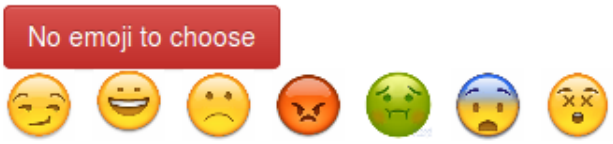


Figure 6.3. – Emojis sélectionnables manuellement et possibilité d'indiquer l'absence d'emoji adéquat.

### 6.2.2. Résultats de l'évaluation du système de recommandation par prédictions combinées d'emojis

Au total 115 messages ont été obtenus. Chaque utilisateur a envoyé un à deux messages. Les résultats montrent que globalement les emojis recommandés sont utilisés par l'utilisateur. En effet, en considérant les messages avec emoji (115-21), majoritairement les emojis recommandés on été utilisés (67 sur 94). Le protocole expérimental présente cependant certaines limites. En effet, dans cette application, l'utilisateur n'est pas impliqué dans une conversation avec un autre utilisateur. Par conséquent, lors de l'évaluation, dès que les emojis étaient affichés, nombreux furent les utilisateurs à effacer leur message en cours pour en taper un nouveau sans envoyer le message. Les instructions fournies en accueil de la borne d'évaluation n'ont pas suffit à résoudre ce problème. Il s'agit d'une limite majeure de cette approche puisque les messages envoyés sont souvent les premiers et constituent un message succinct, tandis que lorsque l'utilisateur veut mettre la prédiction au défi, l'envoi n'était plus la priorité.

Dans la section suivante, nous proposons un autre protocole expérimental d'évaluation permettant de pallier cette problématique. Nous nous concentrons plus spécifiquement sur l'évaluation du système de recommandation d'emojis par groupes proposé au chapitre 5.

## 6.3. Évaluation du système de recommandation automatique de catégories d'emojis

Dans cette section, nous présentons l'évaluation du système de recommandation automatique de catégories d'emojis à partir d'une interface d'évaluation particulière, nommée *Em😃Rec😳*. En effet, afin d'être proche d'une utilisation réelle, le contexte d'une conversation instantanée privée est reproduit. Pour ce faire, une application web de messagerie instantanée est conçue afin de permettre à différents utilisateurs de discuter entre eux lors de conversations privées à deux. La recommandation d'emojis par leurs catégories apparaît lors de la frappe et s'affiche d'une manière similaire à ce que l'on peut trouver dans des messageries instantanées telles que Mood Messenger[3] à ceci près qu'il ne s'agit plus de prédire chaque emoji mais les catégories d'emojis obtenues au chapitre 4.

### 6.3.1. Conception de la plateforme de conversations instantanées

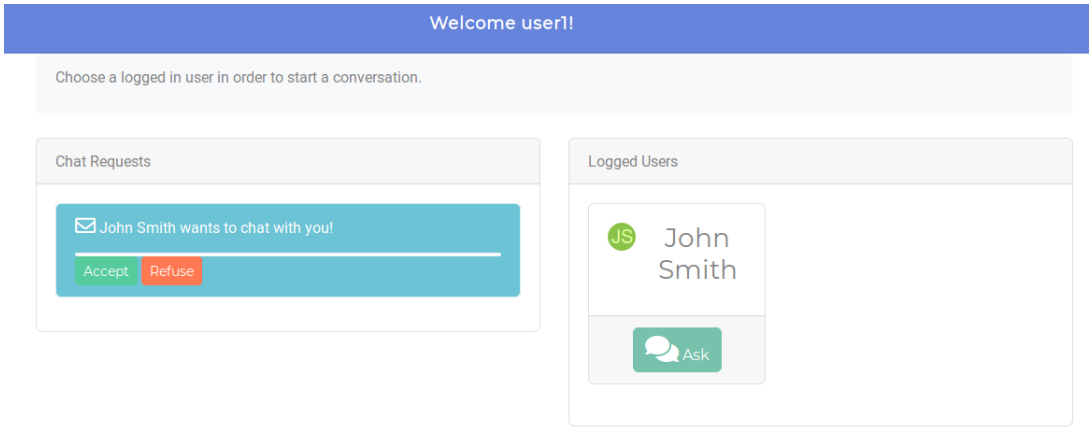


Figure 6.4. – Impression écran du salon d'accueil développé.

La plateforme d'évaluation que nous avons développée reproduit une application de chat dans laquelle un salon général permet d'envoyer ou de recevoir des requêtes de discussion privée. Ce salon est illustré en figure 6.4. Durant une conversation privée, les emojis sont recommandés sous forme de deux groupes possibles. Le réseau CNN-LSTM utilisé (présenté au chapitre 5 section 5.5) étant un modèle de classification mono-étiquette, nous utilisons ses scores de prédiction pour sélectionner les deux premiers groupes d'emojis les plus probables.

3. http://www.moodmessenger.com/

Ce choix s'explique par le score de *Mean Reciprocal Rank* (MRR) (RADEV, QI et al., 2002) de 68,87% obtenu par le réseau lors de la prédiction de catégories présentée au chapitre 5. Ce score indique la propension du modèle à placer la bonne catégorie dans les premières places. Le MRR n'est pas la seule raison pour laquelle nous décidons d'utiliser les deux groupes les plus probables. Nous émettons en effet l'hypothèse selon laquelle l'orientation émotionnelle des emojis attendus par l'utilisateur peut être double et donc parfois inverse au contenu du texte. Selon cette hypothèse, un message tel que "I'm fine" pourrait amener des emojis tristes ou joyeux. C'est pourquoi, nous prenons les deux premiers groupes d'emojis en y appliquant, de plus, un seuil minimum sur les scores de probabilité donnés en sortie du classifieur : un minimum de 10% pour la première catégorie et de 5% pour la seconde. Ces seuils ont été déduits empiriquement à partir de 20 prédictions et rendent possible la considération d'une seule catégorie dans la recommandation. L'affichage des groupes dans l'interface utilisateur se fait en deux bulles représentant les catégories et dans chacune le ou les emojis de la catégorie. L'ordre des emojis dans chaque catégorie est aléatoire afin de ne pas influencer la sélection des emojis d'une même catégorie. La figure 6.5 montre un exemple de deux catégories prédites par le modèle CNN-LSTM.

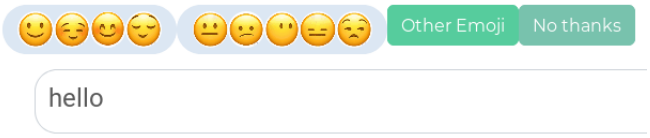


Figure 6.5. – Affichage des bulles d'emojis recommandés accompagnés du bouton de sélection manuelle.

L'affichage des catégories d'emojis recommandés est accompagné de la possibilité pour les utilisateurs d'utiliser les 67 emojis de façon neutre, c'est-à-dire sans passer par la recommandation. L'interface est pensée pour mettre légèrement en avant la recommandation, avec une seule action supplémentaire nécessaire à la sélection manuelle d'emojis : il faut appuyer sur le bouton "*Other emojis*" pour pouvoir en choisir un manuellement. Ce choix d'interface nous permet d'amener l'utilisateur à évaluer les emojis recommandés en les considérant dans un premier temps et, si aucun ne convient, d'en choisir un manuellement.

L'utilisateur peut également choisir le bouton "*No thanks*" pour explicitement indiquer qu'aucun emoji n'est nécessaire dans ce contexte. Notez que ne pas choisir un emoji et envoyer le message sans emojis ne signifie pas nécessairement que la recommandation est mauvaise. En effet, les modèles proposés ne permettent pas de prédire la présence ou l'absence d'emoji. Ils s'appliquent avec l'hypothèse qu'un emoji est présent.

#### 6.3.1.1. Protocole expérimental

Notre objectif lors de l'évaluation est de vérifier que la recommandation est efficace dans tous les contextes et, en particulier, dans tous les contextes émotionnels considérés dans nos travaux. Les sessions de conversations privées sont donc orientées par un scénario de départ choisi au hasard parmi plusieurs scénarios. Chaque scénario correspond à un contexte émotionnel parmi les suivants :

— Colère
— Tristesse
— Joie
— Peur
— Honte
— Fierté
— Mépris
— Amusement

Ces scénarios sont présentés à l'utilisateur avec comme indication de s'imaginer dans le scénario décrit. Ils sont affichés en haut de l'écran, accompagnés des informations relatives à la progression de la session d'évaluation. Chaque session est composée de 5 scénarios pour lesquels 10 messages doivent être envoyés afin de passer au scénario suivant, la progression est donc coupée afin de recentrer la discussion autour du scénario. La phase de pré-tests, réalisée auprès d'étudiants du laboratoire, a en effet démontré le glissement rapide d'un sujet à un autre au bout de plus de dix messages lors de l'échange. Les scénarios sont présentés simultanément aux deux utilisateurs, de manière symétrique ou asymétrique selon le type du scénario. Un exemple de scénario asymétrique est : "*Your dog died a few days ago. You express your sadness to the person you talk to.*" pour un utilisateur et "*You express empathy to the person you talk to who lost his/her dog a few days ago.*" pour l'autre utilisateur. Avec ce dispositif, un prétexte est donné aux utilisateurs pour démarrer la conversation. De plus, les scénarios nous permettent d'évaluer le système de recommandation dans différents contextes émotionnels.

#### 6.3.1.2. Questionnaire pré-expérience

Avant de démarrer l'évaluation, une inscription est nécessaire pendant laquelle l'utilisateur doit fournir un pseudonyme, une adresse email et d'autres informations personnelles énumérées ci-dessous :

— L'âge : 13-17, 18-25, 25-35, ...
— Le genre
— Le niveau d'anglais : de basique à bilingue (*native*)

Enfin, l'utilisateur est pris en main par des indications intégrées sous forme de guide permettant de lui indiquer l'usage de l'interface ainsi que les objectifs. Ce guide permet d'expliquer les groupes d'emojis et les boutons disponibles une fois

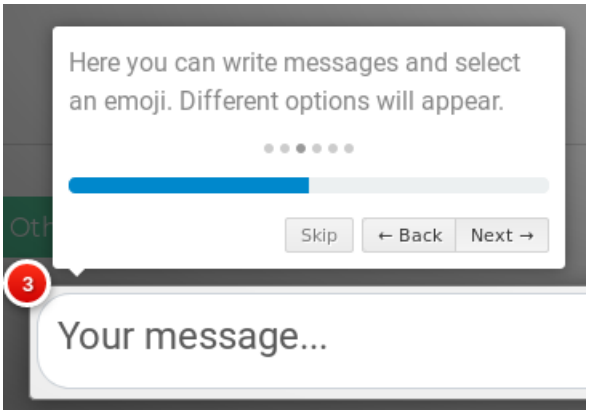


Figure 6.6. – Guide intégré pour s'assurer de l'usage attendu

la recommandation affichée.

#### 6.3.1.3. Questionnaire post-expérience pour l'évaluation subjective de la recommandation

À la fin d'une conversation, il est demandé à l'utilisateur d'évaluer à travers un questionnaire la recommandation d'emojis. Ce questionnaire est présenté en figure 6.7. Il est constitué de trois questions : la première porte sur la satisfaction vis-à-vis de la recommandation d'emojis, la seconde sur la pertinence de cette recommandation et la dernière sur l'utilité de cette recommandation d'emojis pour transmettre des sentiments. L'utilisateur indique sa réponse à travers une échelle de Likert (LIKERT, 1932) de 5 points.

Les utilisateurs peuvent également fournir un commentaire sur la recommandation d'emojis et ce qu'ils pensent du système proposé. Ces commentaires sont optionnels.

### 6.3.2. Résultats

| | |
|---|---|
| **Utilisateurs** | 19 |
| **Messages** | 266 |
| **Messages avec emojis** | 166 |
| **Emojis** | 235 |
| **Emojis recommandés sélectionnés** | 172 |

Tableau 6.1. – Résultats de l'évaluation par analyse de la sélection d'emojis.

L'évaluation a permis de collecter 266 messages de 19 utilisateurs (16 hommes

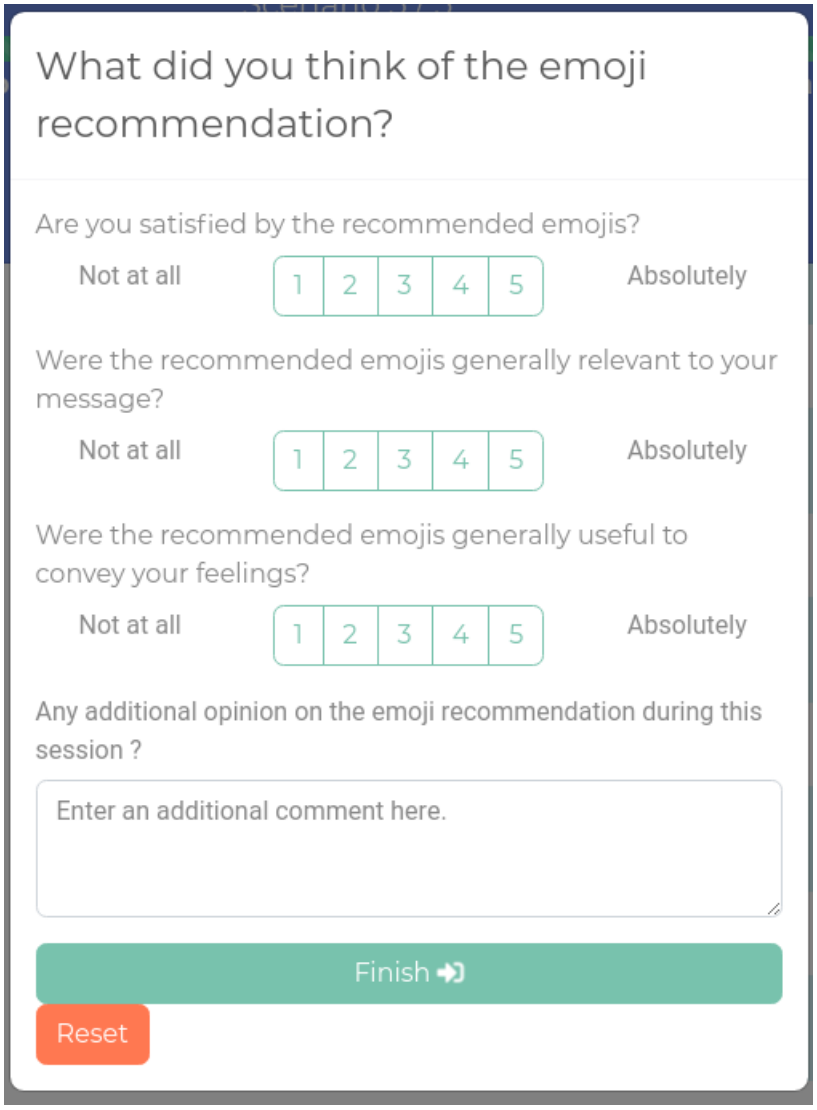


Figure 6.7. – Questionnaire d'évaluation présenté à chaque utilisateur après 5 scénarios. L'utilisateur indique sa réponse à travers une échelle de Likert de 5 points.

et 3 femmes), dont 13 âgés entre 25-35 ans et 6 entre 18-25 ans. La majorité des utilisateurs ont indiqué avoir un anglais fluide (10 *fluent*) ou basique (6), contre 2 basiques et 1 bilingue. Les utilisateurs ont été rémunérés par des bons d'achats distribués aléatoirement parmi ceux ayant complété une évaluation, c'est-à-dire avoir participé à 5 scénarios à la suite et rempli le questionnaire post-expérience. Parmi les messages envoyés 62,41% l'ont été avec au moins un emoji issu de la recommandation. Dans 22,56% des messages un emoji a été sélectionné en dehors de la recommandation. Finalement, plusieurs messages ont des emojis à la fois issus de la sélection manuelle et de la recommandation.

Parmi les emojis utilisés, 172 proviennent de la recommandation et 63 de la sélection manuelle, soit 73,19% d'emojis recommandés. C'est ce taux d'emojis recommandés par rapport au total d'emojis que nous considérons comme un score pertinent. Il valide une action explicite de confirmation ou d'infirmation du système de recommandation de la part de l'utilisateur. Ce score obtenu est supérieur aux 53,10% de macro f-mesure obtenus par le même modèle lors de l'évaluation hors-ligne (voir chapitre 5).

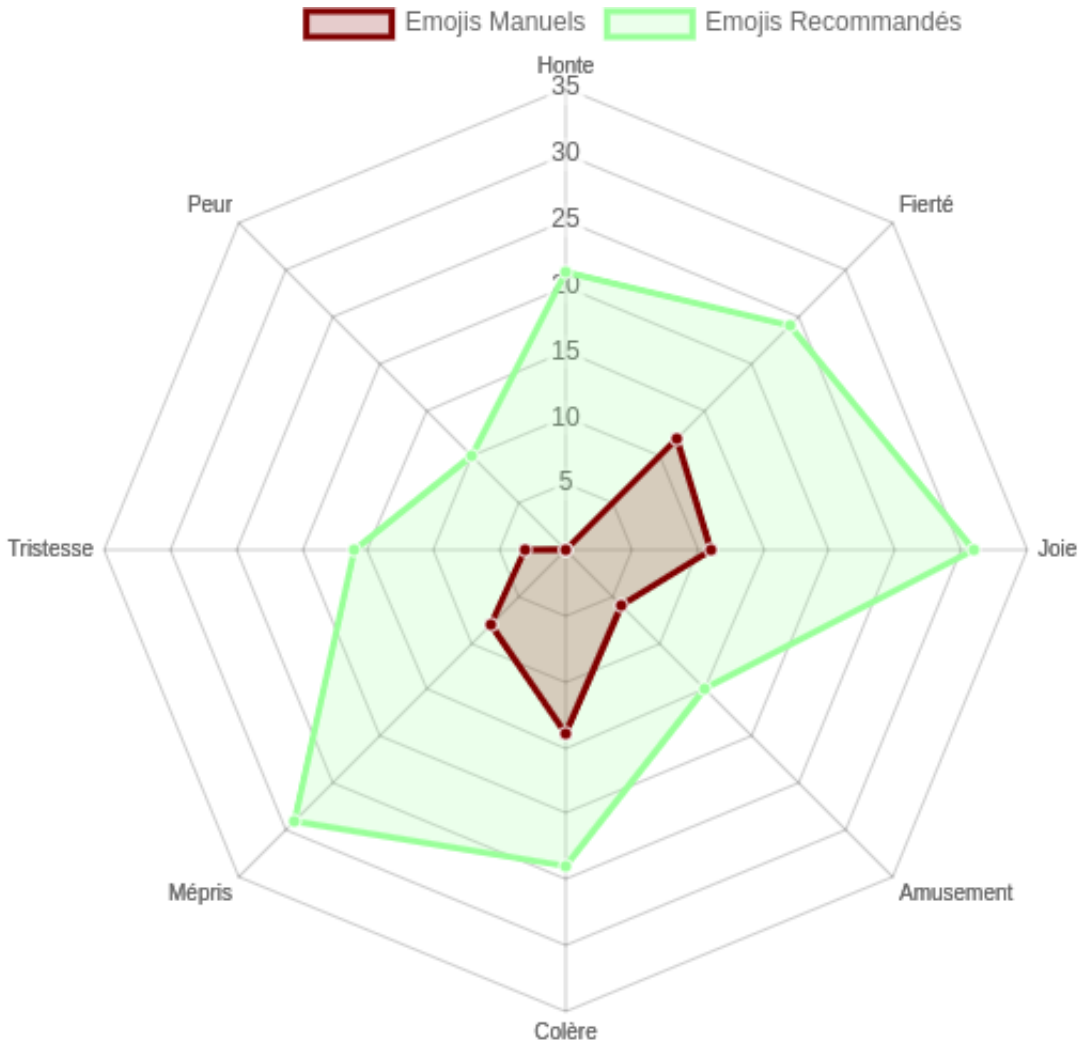


Figure 6.8. – Sélection manuelle ou par recommandation des emojis selon les contextes émotionnels.

La répartition entre la sélection par la recommandation ou la sélection manuelle des emojis est présentée en figure 6.8 pour chaque contexte émotionnel. Cette répartition permet de constater de bons taux de recommandation pour le

mépris, la fierté, la joie et la honte. Cependant, la recommandation apparaît légèrement moins efficace pour la colère en n'ayant que 63,16% d'emojis choisis par la recommandation. Globalement, la recommandation d'emojis est équilibrée et efficace pour chaque contexte émotionnel.

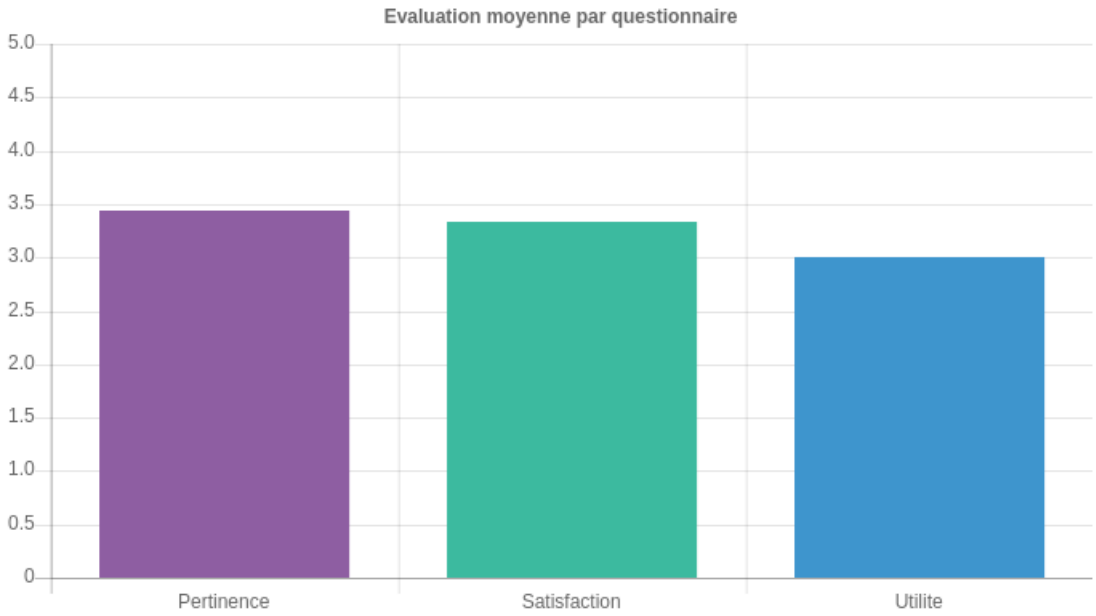


Figure 6.9. – Résultats de l'évaluation par questionnaire.

Les résultats moyens des évaluations présentés en figure 6.9 indiquent une pertinence correcte de la part du système de recommandation proposé avec un score moyen de 3/5 à la question "*Were the recommended emojis generally relevant to your message ?*". En revanche, les scores de satisfaction ("*Are you satisfied by the recommended emojis ?*") et d'utilité ("*Were the recommended emojis generally useful to convey your feelings ?*") sont moyens.

Le questionnaire possède également un champ de commentaire optionnel que certains utilisateurs ont rempli. Il en ressort la difficulté du système à conserver une bonne recommandation lorsque le message est ponctué de mots tels que "*as well*". Selon ces commentaires, la recommandation est perfectible lorsque plusieurs phrases avec des émotions inverses sont insérées. Toutefois, cette dernière remarque est directement liée à la limitation volontaire du contexte proche à 11 mots précédents lors de l'élagage et du remplissage des données d'entrée du CNN-LSTM (Section 5.4).

#### 6.3.2.1. Sarcasme, ironie

Cette plateforme d'évaluation met en avant un autre avantage de la recommandation d'emojis par prédiction de catégories obtenues automatiquement : le

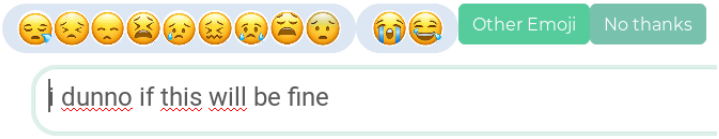


Figure 6.10. – Exemple de recommandation prenant en compte une ironie possible.

sarcasme et l'ironie sont pris en compte. Cet effet est visible en figure 6.10 où les deux catégories proposées sont totalement opposées, l'une suivant la polarité négative générale de la phrase et la seconde orientant l'émotion vers le rire, où plus précisément l'ironie. Comme ils se distinguent comme étant deux exemples de cas d'opposition entre polarité du texte et polarité de l'emoji, l'ironie et le sarcasme sont intégrés dans la recommandation car présents initialement dans le corpus d'entraînement.

Toutefois, les cas particuliers ne sont pas toujours pris en compte. La figure 6.11 montre ainsi un cas de recommandation dans laquelle l'ironie pourrait être utilisée suite à la combinaison du mot "*already*" avec la forme interrogative.

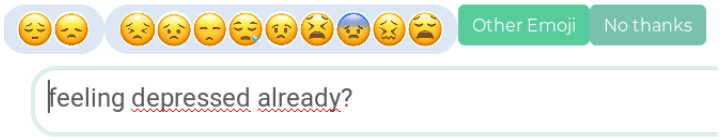


Figure 6.11. – Exemple de recommandation à sens unique

L'évaluation proposée présente certaines limites qui sont discutées ci-dessous.

## 6.4. Limites

Cette évaluation de la recommandation d'emojis possède plusieurs limites. Tout d'abord, l'imitation d'un contexte informel privé n'est pas parfait puisque les utilisateurs ne se connaissent pas et ne peuvent s'identifier qu'à l'aide du pseudo choisi. Par conséquent, toute la spontanéité propre à un contexte privé n'est pas totalement reproduite. Les scénarios ne sont quant à eux pas toujours suivis.

De plus, l'évaluation d'une recommandation se fait également par la vérification de son influence sur le comportement de l'utilisateur (SHANI et GUNAWARDANA, 2011), comment la recommandation modifie son comportement habituel. Avec le système présenté cette donnée ne peut pas être prise en compte puisque le système n'est pas intégré à une application existante avec laquelle l'utilisateur

est familier.

L'interprétation de la recommandation n'est pas évaluée ici. Le fait de savoir si l'utilisateur a toujours trouvé la recommandation compréhensible en pouvant aisément l'interpréter est une donnée qui n'est pas présente dans le formulaire. Les libellés des catégories d'emojis ne sont pas d'ailleurs pas indiqués.

Enfin, parmi les commentaires fournis, la question de la combinaison de phrases opposées est soulevée, ainsi que la difficulté du modèle à donner une recommandation pertinente lorsque certaines expressions sont ajoutées en ponctuation de fin de message, comme par exemple "*as well*" ou "*in general*".

Le nombre de participants est faible malgré la motivation des bons d'achats. Ce qui s'explique notamment par la nécessité de réunir les participants au même moment pour qu'ils puissent interagir. Cela rend la validation des résultats difficiles et il convient de prendre cela en compte en perspectives.

## 6.5. Conclusion et perspectives

Dans ce chapitre nous avons élaboré deux manières d'évaluer subjectivement la recommandation d'emojis par étude du comportement de l'utilisateur (dite *étude utilisateur*) afin de compléter les évaluations objectives obtenues (dite *hors ligne*) dans les chapitres 3 et 5. À notre connaissance, nous sommes les premiers à aborder la question de l'évaluation de la recommandation d'emojis en dehors d'une évaluation de leur prédiction.

La première évaluation a pris la forme d'une interface à état unique présentée lors d'un salon dédié à la valorisation en Sciences Humaines et Sociales du CNRS. Une combinaison de l'approche générative pour prédire un emoji de remplacement de mot et discriminante pour la classification multi-étiquettes d'emojis a été utilisée. Ce système a montré des failles au travers du retour général des participants au salon mais surtout, l'évaluation elle-même est restée assez limitée puisqu'elle ne permettait pas d'obliger les utilisateurs à envoyer leur message dans la base de données.

Une application web de discussion informelle privée a été conçue pour mener une seconde évaluation. Dernière l'interface, le système de prédiction de catégories d'emojis fut utilisé pour recommander les emojis des deux catégories les plus probables. Ce système repose sur deux principes d'évaluations : la confirmation de la recommandation et l'évaluation générale sous forme de questionnaire.

Cette évaluation subjective mise en place confirme à 73,19% la recomman-

dation effectuée contre 53,10% de macro f-mesure obtenus lors de l'évaluation subjective du modèle prédictif. Elle a également permis de savoir que la pertinence est perçue comme correcte par les utilisateurs, tandis que l'utilité et la satisfaction sont moyennes. L'approche proposée dans ce chapitre tend à confirmer l'inadéquation des métriques objectives d'évaluation avec l'évaluation réelle de la recommandation.

Ces évaluations ont montré des limites qu'il convient de corriger. Toutes deux ont mis en avant la difficulté d'obtenir des données par une évaluation dans une application autonome, une application dédiée pour une étude utilisateur. Il apparaît nécessaire d'intégrer cette évaluation dans une application existante telle que Mood Messenger, afin de mettre en place une évaluation *en ligne* dans laquelle l'utilisateur n'est pas au courant du processus d'évaluation. La mise en place d'un test A/B qui consisterait à partager la nouvelle version de la recommandation à 50% des utilisateurs, l'autre moitié conservant la version précédente. Cela permettrait à la fois de profiter de la manne d'utilisateurs actifs, ainsi que de comparer l'impact de la recommandation sur l'utilisation de l'application et la satisfaction globale procurée par la nouvelle recommandation. À défaut, une application mobile indépendante (voir annexes) permettrait *a minima* de prendre en compte le contexte logiciel (claviers, etc.). Enfin, la qualité de prise en compte de l'ironie et du sarcasme dans la recommandation mériterait une évaluation dédiée.

## 6.6. Synthèse

**Objectifs** : évaluer la qualité de la recommandation directement par les utilisateurs.

**Méthode** : une évaluation par étude utilisateur. Deux approches sont proposées.
— Évaluation de la prédiction d'emojis en combinant modèles génératifs et discriminants pour une classification multi-étiquettes. Présentée sous forme de borne dans un salon de valorisation.
— Évaluation de la recommandation d'emojis par catégories automatiques. Distribuée sur des listes locales et reproduisant le contexte conversationnel privé.

**Résultats** :
— Recommandation de catégories d'emojis considérée comme correcte par l'utilisateur dans 69,49%. Ce score est bien au dessus de celui obtenu par métriques d'évaluations.

**Limites** :
— Comparaison impossible avec un comportement habituel. Le contexte logiciel ainsi que le contexte informel privé ne peuvent pas être complètement reproduits.
— L'interprétation de la recommandation n'est pas évaluée.

**Contributions** :
— Système de recommandation d'emojis évalué à la fois par métriques objectives et par évaluation d'utilisateurs.
— Confirmation du caractère indispensable d'une évaluation réelle de la recommandation d'emojis.

# Conclusion générale

Ce manuscrit présente plusieurs approches pour effectuer de la recommandation automatique d'emojis, principalement sentimentaux (*i.e.* représentant une expression de l'émotion), utilisable en temps réel qui puisse s'adapter à de nouveaux ajouts d'emojis. En partant de l'hypothèse que la recommandation d'emojis diffère selon les fonctions que ces emojis exercent dans la conversation, dans les travaux de thèse présentés dans ce manuscrit, plusieurs approches pour la recommandation en temps réel d'emojis ont été proposées. Elles ont l'avantage de prendre en compte les différentes fonctions de l'emoji mais aussi de s'adapter à des nouveaux emojis émergeant ou de nouveaux usages.

## Contributions

Dans ce manuscrit, nous avons décrits plusieurs contributions, chaque contribution principale ayant fait l'objet d'un chapitre dédié qu'il convient de résumer.

### Prédiction d'emojis

Notre première contribution fut la mise en place de systèmes prédictifs d'emojis permettant de prendre en compte plusieurs fonctions des emojis dans la conversation parmi six possibles (JAKOBSON, 1963) : l'enrichissement méta linguistique, l'expression, la fonction référentielle ainsi que la fonction conative. La première approche se fonde sur un modèle génératif pour prendre en compte les fonctions référentielles (l'emoji sert à désigner un objet) et d'expression (l'emoji exprime une idée) de l'emoji. Le mot suivant ou en cours est alors prédit ou complété à partir du contexte proche d'un seul ou de deux mots précédents, il est ensuite comparé avec un lexique agrégé d'associations mots-emoji à l'aide d'une distance d'édition pour permettre d'obtenir l'emoji correspond au mot désiré ou au mot que le système pense voulu par l'utilisateur. Ce système part du principe que l'emoji remplace un mot en cours et a été évalué par reproduction du contexte précédant l'utilisation d'un emoji pour remplacer un mot dans un corpus de messages privés. Une telle approche s'est montrée efficace uniquement pour un nombre restreint d'emojis représentant principalement des objets tels qu'un hôpital ou un drapeau, démontrant ainsi qu'une recommandation à partir de cette approche, qui est une amélioration du système industriel présent dans Mood Messenger, ne saurait suffire à obtenir une bonne recommandation.

Afin de prendre en compte les autres fonctions méta-linguistiques (l'emoji enrichit le texte), conatives (l'emoji influence le récepteur) et toujours expres-

sives de l'emoji, une approche fondée sur un modèle discriminant a été mise en place. De tels systèmes permettant la prédiction d'emoji ont récemment émergé (BARBIERI, BALLESTEROS et SAGGION, 2017 ; HUANG, W. XU et al., 2015 ; BARBIERI, CAMACHO-COLLADOS et al., 2018). Cependant, les systèmes existants se concentrent sur un ensemble restreint d'emojis en conservant les emojis les plus fréquents. De plus, parce que nous utilisons une classification supervisée multi-étiquettes, notre approche permet de laisser un choix final à l'utilisateur contrairement à ce qui se fait actuellement dans l'état de l'art à base uniquement de prédiction d'un seul emoji par contenu textuel. À l'aide de forêts aléatoires d'arbres de décision nous nous distinguons donc en prédisant plusieurs emojis possibles par phrases, mais aussi en appliquant cette prédiction sur un corpus de messages instantanés privés, contrairement aux tweets. Ce corpus a l'avantage de posséder un indicateur d'humeur de l'utilisateur, motivant une approche par caractéristiques discriminantes orientées sur les sentiments. L'approche est finalement appliquée sur deux types de corpus, un avec uniquement 169 emojis sentimentaux extraits à l'aide d'une table de correspondance emoji-sentiment existante (KRALJ NOVAK, SMAILOVIĆ et al., 2015) et l'autre utilisant les 1 070 emojis possibles, nous distinguant également par le nombre d'emojis considérés. Les résultats obtenus ont donné un score global de 76,84% de f-mesure, ce qui est élevé comparé aux résultats actuels sur les tweets. Les résultats ont également montré un rôle majeur de l'indicateur de l'humeur (*mood*) dans la qualité de la prédiction ainsi qu'une classification plus efficace en prenant en compte les sacs de caractères. Malgré ce fort impact de l'humeur, la prédiction d'emojis sentimentaux est équivalente à celle appliquée sur les autres emojis, avec une F-mesure de 76,60%, ne rendant pas l'approche particulièrement plus efficace sur les emojis sentimentaux que sur l'ensemble des emojis.

## Catégorisation d'emojis émotionnels

La seconde contribution mise en avant dans ces travaux concerne l'obtention automatique de catégories d'emojis, le but étant dans un second temps d'utiliser ces catégories pour la recommandation d'emojis (Section suivante). L'objectif était d'obtenir automatiquement des classes d'emojis en prenant ici le cas des emojis représentant des expressions des émotions à travers le visage. Une catégorisation de 64 emojis sentimentaux répertoriés comme tels par le consortium Unicode [4] a ainsi été mise en place à l'aide de plongements lexicaux d'emojis en contexte (MIKOLOV, CHEN et al., 2013 ; MIKOLOV, SUTSKEVER et al., 2013) tout d'abord, puis à partir de partitionnement automatique. Une comparaison a été faite en utilisant un modèle de plongements lexicaux existant (POHL, DOMIN et al., 2017) avec la constitution de modèles appris sur des tweets. Nous nous dif-

4. http://unicode.org/emoji/charts/full-emoji-list.html

férencions aussi de par notre objectif, avec l'obtention de catégories à réutiliser et non une simple exploration de corpus. De cette manière l'effet néfaste d'une représentation trop précise et prenant en compte les mots rares contrairement à une représentation vectorielle plus en surface a été mise en avant, tandis que la constitution de plongements lexicaux à des fins de classification a des besoins inverses.

En partant de l'hypothèse que l'usage des emojis reflète l'usage des indices faciaux dans une conversation en face-à-face, nous avons comparé la catégorisation obtenue avec une théorie existante des catégories d'émotions basiques par expression du visage (EKMAN, 1999). Les résultats montrent une corrélation globale entre les catégories issues de la théorie et celles obtenues automatiquement, par le biais d'un score de V-mesure de 76,70%. De plus, la granularité des catégories obtenues est bien souvent plus fine que la théorie en faisant par exemple la distinction entre différentes intensités de joie.

## Recommandation d'emojis par prédiction de leur catégorie

Notre troisième contribution concerne la recommandation d'emojis par le biais de la prédiction de leur catégorie. Nous avons réutilisé les catégories d'emojis visage obtenues automatiquement pour les considérer comme un jeu d'étiquettes pour la classification de messages instantanés privés. Cette recommandation d'emojis consiste donc à prédire un panel de choix à l'utilisateur, en l'occurrence 18 catégories regroupant au moins un emoji. La tâche est abordée comme une classification mono-étiquette dans laquelle nous comparons plusieurs algorithmes avant d'utiliser l'apprentissage profond. En partant initialement du réseau de convolution de Kim (KIM, 2014), nous l'avons modifié pour notamment y ajouter des couches récurrentes pour la prise en compte de l'ordonnancement des mots dans le message. La méthode est évaluée avec des scores macro de précision, rappel et f-mesure, permettant d'obtenir de bons résultats avec 53,10% en macro F-mesure comparé aux autres algorithmes n'atteignant pas les 50%. Quelques confusions persistent entre les catégories d'emojis à granularité fine.

Cette recommandation est à notre connaissance le premier cas de recommandation d'emojis au travers de leur catégorie. Cette approche pallie également aux systèmes de prédiction d'emoji existants qui ne recommandent qu'un seul emoji par classification mono-étiquette. Le fait que notre système soit totalement automatisé avec des classes à prédire obtenues automatiquement, en fait un système bout en bout adaptable sans nécessiter d'experts pour la recommandation d'emojis.

### Évaluation par étude du comportement de l'utilisateur

Enfin, nous avons mis en place une évaluation directe auprès d'utilisateurs des différentes approches de recommandation d'emojis, étape nécessaire pour compléter les évaluations objectives (macro F-mesure, exactitude, MRR) et hors ligne (RICCI, ROKACH et al., 2011). La première a été effectuée à l'aide d'une interface de recommandation des emojis en combinant l'approche générative et discriminante par classification multi-étiquettes directe d'emojis décrite ci-dessus (Section 6.6). Plusieurs utilisateurs ont interagi avec le système intégrant cette approche dans le cadre d'un forum de valorisation de la recherche du CNRS. Les résultats de l'évaluation montrent un usage majoritaire des emojis recommandés mais également sont limités par l'absence de contexte conversationnel dans l'interface.

La seconde évaluation fut dédiée à la recommandation d'emojis par prédiction de leurs catégories. En prenant en compte les manquements de la première évaluation, l'interface est ici disponible en ligne pour plusieurs utilisateurs simultanés, avec l'intégration du système de recommandation dans un chat privé. Le contexte conversationnel privé a été reproduit en y ajoutant des scénarios prédéfinis pour chaque session de discussion. Ces scénarios ont été orientés par 8 contextes émotionnels afin d'évaluer le système dans chacun d'entre eux. Nous avons pallié à la prédiction mono-étiquette par l'utilisation des scores de régression du classifieur. En considérant le taux de MRR obtenu lors de la mise en place du modèle de recommandation (Section 6.6), les deux catégories les plus probables sont affichées moyennant un score de probabilité supérieur à 10% pour la première catégorie et 5% pour la seconde. Les résultats d'analyse de l'utilisation de l'interface ont montré une utilisation d'emojis provenant à 76,19% de la recommandation, avec une validation de la recommandation homogène dans l'ensemble des contextes conversationnels, le plus bas étant à 63,16% pour la colère. Les scores moyens des questionnaires de fin de session ont quant à eux montré des scores de pertinence, de satisfaction et d'utilité corrects, chacun ayant une moyenne au-dessus de 3 sur 5 à l'échelle de Likert (LIKERT, 1932).

## Perspectives

Pour chaque contribution, plusieurs limites ont été identifiées. Il convient donc de les prendre en compte pour améliorer le système de recommandation d'emojis ainsi que la mise en œuvre de son évaluation lors de travaux futurs aussi bien à court terme qu'à moyen et long terme.

Une grande partie des travaux présentés dans cette thèse ont porté leur at-

tention sur les emojis liés aux sentiments et à l'expression des émotions par le visage. Pour valider notre méthodologie de recommandation d'emojis par prédiction de leurs catégories, il conviendrait de l'appliquer à d'autres types d'emojis afin de tester la robustesse sur des emojis complètement différents tels que les emojis représentant des objets. Ces derniers ont déjà été pris en compte lors de la prédiction d'emojis du chapitre 3 et nécessiteraient l'usage d'une telle approche pour tenir compte de catégories telle que 🌳🌲🌴.

L'autre évolution à court terme envisagée est l'usage de ce système pour détecter des émotions nuancées en prenant en compte les relations entre ces émotions, à l'aide par exemple de catégorisations existantes (PLUTCHIK, 1960). Ne plus se concentrer uniquement sur l'expression des émotions par le visage, mais sur les émotions en elles-mêmes permettrait alors d'améliorer la recommandation effectuée.

L'évaluation proposée tend à reproduire un contexte réel d'application de messagerie instantanée privée. Il convient toutefois de l'appliquer directement dans l'application cible de l'entreprise. Ainsi, un test A/B fournissant la nouvel recommandation à un sous-ensemble d'utilisateurs permettrait de vérifier les avantages et inconvénients du système de recommandation, vis-à-vis du système existant dans l'application. Il s'agira plus d'une évaluation par étude du comportement de l'utilisateur, mais d'une évaluation *en ligne* qui permettra également d'obtenir un nombre de données bien plus important que celui utilisé. Les prémices de ces travaux sont présentées en annexes.

À moyen terme nous identifions deux axes : l'adaptation des données au contexte informel privé et la personnalisation de la recommandation.

Notre approche est en grande partie fondée sur l'utilisation de messages instantanés privés afin de correspondre à l'utilisation finale et pour des besoins industriels. Il convient de s'affranchir de ces données sans pour autant perdre en efficacité dans ce contexte. Une telle perspective a été commencée en section 5.8.2 dans laquelle nous avons essayé de mélanger données privées et publiques pour obtenir une meilleure généralisation, mais sans succès. Une autre approche serait d'utiliser de l'apprentissage par transfert (PRATT, 1993) pour prendre en compte les informations de données publiques dans un contexte informel privé.

La recommandation actuelle est apprise en amont et ne génère pas de problème de démarrage à froid. Bien qu'elle s'adapte à l'ajout de nouveaux emojis et permet de s'adapter à l'évolution générale de l'utilisation des emojis puisqu'il s'agit d'un système automatisé de bout en bout, il convient de la personnaliser. Une telle approche pourrait être envisagée par apprentissage par renforcement (SUTTON, 1988 ; WATKINS et DAYAN, 1992) ou par création d'un modèle de préfé-

rence de l'utilisateur influençant la recommandation finale. Cette approche reste une perspective à moyen terme car elle nécessiterait tout de même d'avoir un apprentissage permanent allant à l'encontre des besoins de gestion des ressources que possèdent les applications de messagerie sur mobile.

La personnalisation indiquée à moyen terme nous amène à une orientation à long terme vers des systèmes de graphes de modèles dédiés. Un graphe de catégories d'utilisateurs peut alors être considéré dans lequel chaque nœud représente un modèle pré-entraîné dédié à une recommandation d'emojis spécifique. Par exemple, l'inclusion du mot "*game*" influencerait le modèle vers un sport, un jeu de société, ou autre préférence de cette catégorie d'utilisateurs.

# Bibliographie


[1] Sho AOKI et Osamu UCHIDA. « A method for automatically generating the emotional vectors of emoticons using weblog articles ». In : *Proceedings of the 10th WSEAS international conference on Applied computer and applied computational science*. World Scientific, Engineering Academy et Society (WSEAS). 2011, p. 132–136 (cf. p. 47).

[2] John Langshaw AUSTIN. *How to do things with words*. Oxford university press, 1975 (cf. p. 36).

[3] Jason BALDRIDGE. « The opennlp project ». In : *URL : http ://opennlp. apache. org/index. html,(accessed 2 February 2012)* (2005) (cf. p. 68).

[4] Francesco BARBIERI, Miguel BALLESTEROS, Francesco RONZANO et Horacio SAGGION. « Multimodal emoji prediction ». In : *arXiv preprint arXiv :1803.02392* (2018) (cf. p. 52).

[5] Francesco BARBIERI, Miguel BALLESTEROS et Horacio SAGGION. « Are emojis predictable ? » In : *arXiv preprint arXiv :1702.07285* (2017) (cf. p. 51, 83, 106, 144).

[6] Francesco BARBIERI, Jose CAMACHO-COLLADOS, Francesco RONZANO, Luis Espinosa ANKE, Miguel BALLESTEROS, Valerio BASILE, Viviana PATTI et Horacio SAGGION. « Semeval 2018 task 2 : Multilingual emoji prediction ». In : *Proceedings of The 12th International Workshop on Semantic Evaluation*. 2018, p. 24–33 (cf. p. 51, 144).

[7] Francesco BARBIERI, Luis MARUJO, Pradeep KARUTURI, William BRENDEL et Horacio SAGGION. « Exploring emoji usage and prediction through a temporal variation lens ». In : *arXiv preprint arXiv :1805.00731* (2018) (cf. p. 52).

[8] Francesco BARBIERI, Francesco RONZANO et Horacio SAGGION. « What does this Emoji Mean ? A Vector Space Skip-Gram Model for Twitter Emojis. » In : *LREC*. 2016 (cf. p. 45).

[9] Nikhil BOJJA, Satheeshkumar KARUPPUSAMY, Pidong WANG, Shivasankari KANNAN et Arun NEDUNCHEZHIAN. *Systems and methods for suggesting emoji*. US Patent App. 15/384,950. Juin 2017 (cf. p. 66).

[10] John S BRIDLE. « Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition ». In : *Neurocomputing*. Springer, 1990, p. 227–236 (cf. p. 51).

[11] Rui CAI, Chao ZHANG, Chong WANG, Lei ZHANG et Wei-Ying MA. « Musicsense : contextual music recommendation using emotional allocation modeling ». In : *Proceedings of the 15th ACM international conference on Multimedia*. ACM. 2007, p. 553–556 (cf. p. 57).

[12] Spencer CAPPALLO, Thomas MENSINK et Cees G.M. SNOEK. « Image2Emoji : Zero-shot Emoji Prediction for Visual Media ». In : ACM Press, 2015, p. 1311–1314. ISBN : 978-1-4503-3459-4. DOI : 10.1145/2733373.2806335. URL : http://dl.acm.org/citation.cfm?doid=2733373.2806335 (visité le 16/09/2016) (cf. p. 52).

[13] Spencer CAPPALLO, Stacey SVETLICHNAYA, Pierre GARRIGUES, Thomas MENSINK et Cees GM SNOEK. « New Modality : Emoji Challenges in Prediction, Anticipation, and Retrieval ». In : *IEEE Transactions on Multimedia* 21.2 (2019), p. 402–415 (cf. p. 52).

[14] François CHOLLET et al. « Keras : Deep learning library for theano and tensorflow ». In : *URL : https ://keras. io/k* 7.8 (2015), T1 (cf. p. 111).

[15] Ferdinand DE SAUSSURE. *Cours de linguistique générale*. critique. Grande Bibliothèque Payot, 1916. 269 p. (cf. p. 31–33).

[16] Bhuwan DHINGRA, Zhong ZHOU, Dylan FITZPATRICK, Michael MUEHL et William W COHEN. « Tweet2vec : Character-based distributed representations for social media ». In : *arXiv preprint arXiv :1605.03481* (2016) (cf. p. 46, 50).

[17] John DUCHI, Elad HAZAN et Yoram SINGER. « Adaptive subgradient methods for online learning and stochastic optimization ». In : *Journal of Machine Learning Research* 12.Jul (2011), p. 2121–2159 (cf. p. 114).

[18] Ben EISNER, Tim ROCKTÄSCHEL, Isabelle AUGENSTEIN, Matko BOŠNJAK et Sebastian RIEDEL. « emoji2vec : Learning Emoji Representations from their Description ». In : *arXiv preprint arXiv :1609.08359* (2016). URL : http://arxiv.org/abs/1609.08359 (visité le 05/10/2016) (cf. p. 41, 48).

[19] P EKMAN. *Basic Emotions In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion Pp. 45-60*. 1999 (cf. p. 86, 97, 98, 101, 110, 124, 145).

[20] Bjarke FELBO, Alan MISLOVE, Anders SØGAARD, Iyad RAHWAN et Sune LEHMANN. « Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm ». In : *arXiv preprint arXiv :1708.00524* (2017) (cf. p. 51, 106).

[21] Joseph L FLEISS, Jacob COHEN et Brian S EVERITT. « Large sample standard errors of kappa and weighted kappa. » In : *Psychological Bulletin* 72.5 (1969), p. 323 (cf. p. 48).

[22] G David FORNEY. « The viterbi algorithm ». In : *Proceedings of the IEEE* 61.3 (1973), p. 268–278 (cf. p. 61).

[23] Jerome H FRIEDMAN. « Greedy function approximation : a gradient boosting machine ». In : *Annals of statistics* (2001), p. 1189–1232 (cf. p. 110).

[24] Xavier GLOROT et Yoshua BENGIO. « Understanding the difficulty of training deep feedforward neural networks ». In : *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010, p. 249–256 (cf. p. 113).

[25] Alex GRAVES, Abdel-rahman MOHAMED et Geoffrey HINTON. « Speech recognition with deep recurrent neural networks ». In : *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE. 2013, p. 6645–6649 (cf. p. 117).

[26] Gaël GUIBON, Magalie OCHS et Patrice BELLOT. « From emojis to sentiment analysis ». In : *WACAI 2016*. 2016 (cf. p. 36).

[27] Gaël GUIBON, Magalie OCHS et Patrice BELLOT. « Une plateforme de recommandation automatique d'emojis ». In : *Traitement Automatique du Langage Naturel*. 2017 (cf. p. 128).

[28] David GUTHRIE, Ben ALLISON, Wei LIU, Louise GUTHRIE et Yorick WILKS. « A closer look at skip-gram modelling. » In : *LREC*. 2006, p. 1222–1225 (cf. p. 96).

[29] Hussam HAMDAN, Patrice BELLOT et Frederic BECHET. « Sentiment lexicon-based features for sentiment analysis in short text ». In : *In Proceeding of the 16th International Conference on Intelligent Text Processing and Computational Linguistics*. 2015. URL : https://www.researchgate.net/profile/Hussam_Hamdan/publication/274249633_Sentiment_Lexicon-Based_Features_for_Sentiment_Analysis_in_Short_Text/links/5530bdce0cf2f2a588ab2b65.pdf (visité le 01/10/2016) (cf. p. 69, 71).

[30] Kaiming HE, Xiangyu ZHANG, Shaoqing REN et Jian SUN. « Deep residual learning for image recognition ». In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 770–778 (cf. p. 52).

[31] Sepp HOCHREITER et Jürgen SCHMIDHUBER. « LSTM can solve hard long time lag problems ». In : *Advances in neural information processing systems*. 1997, p. 473–479 (cf. p. 46, 51, 114).

[32] Alexander HOGENBOOM, Daniella BAL, Flavius FRASINCAR, Malissa BAL, Franciska de JONG et Uzay KAYMAK. « Exploiting emoticons in sentiment analysis ». In : *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. ACM, 2013, p. 703–710. URL : http://dl.acm.org/citation.cfm?id=2480498 (visité le 07/03/2016) (cf. p. 50).

[33] Zhiheng HUANG, Wei XU et Kai YU. « Bidirectional LSTM-CRF models for sequence tagging ». In : *arXiv preprint arXiv :1508.01991* (2015) (cf. p. 51, 118, 144).

[34] Paul JACCARD. « The distribution of the flora in the alpine zone. 1 ». In : *New phytologist* 11.2 (1912), p. 37–50 (cf. p. 46).

[35] Roman JAKOBSON. « Essais de linguistique générale ». In : (1963) (cf. p. 33, 143).

[36] Tanimu Ahmed JIBRIL et Mardziah Hayati ABDULLAH. « Relevance of Emoticons in Computer-Mediated Communication Contexts : An Overview ». In : *Asian Social Science* 9.4 (28 mar. 2013). ISSN : 1911-2025, 1911-2017. DOI : 10.5539/ass.v9n4p201. URL : http://www.ccsenet.org/journal/index.php/ass/article/view/26102 (visité le 05/04/2016) (cf. p. 35).

[37] Armand JOULIN, Edouard GRAVE, Piotr BOJANOWSKI et Tomas MIKOLOV. « Bag of tricks for efficient text classification ». In : *arXiv preprint arXiv :1607.01759* (2016) (cf. p. 52).

[38] Guolin KE, Qi MENG, Thomas FINLEY, Taifeng WANG, Wei CHEN, Weidong MA, Qiwei YE et Tie-Yan LIU. « Lightgbm : A highly efficient gradient boosting decision tree ». In : *Advances in Neural Information Processing Systems*. 2017, p. 3146–3154 (cf. p. 111).

[39] Caroline KELLY. « Do you know what I mean > :( : A linguistic study of the understanding ofemoticons and emojis in text messages ». In : (2015). URL : http://www.diva-portal.org/smash/record.jsf?pid=diva2:783789 (visité le 07/03/2016) (cf. p. 36).

[40] Ryan KELLY et Leon WATTS. « Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships ». In : *Experiences of Technology Appropriation : Unanticipated Users, Usage, Circumstances, and Design* (2015). URL : https://projects.hci.sbg.ac.at/ecscw2015/wp-content/uploads/sites/31/2015/08/Kelly_Watts.pdf (visité le 25/05/2016) (cf. p. 40).

[41] Jack KIEFER, Jacob WOLFOWITZ et al. « Stochastic estimation of the maximum of a regression function ». In : *The Annals of Mathematical Statistics* 23.3 (1952), p. 462–466 (cf. p. 114).

[42] Yoon KIM. « Convolutional neural networks for sentence classification ». In : *arXiv preprint arXiv :1408.5882* (2014) (cf. p. 51, 111, 145).

[43] Diederik P KINGMA et Jimmy BA. « Adam : A method for stochastic optimization ». In : *arXiv preprint arXiv :1412.6980* (2014) (cf. p. 113, 114).

[44] Petra KRALJ NOVAK, Jasmina SMAILOVIĆ, Borut SLUBAN et Igor MOZETIČ. « Sentiment of Emojis ». In : *PLOS ONE* 10.12 (7 déc. 2015). Sous la dir. de Matjaz PERC, e0144296. ISSN : 1932-6203. DOI : 10.1371/journal.pone.0144296. URL : http://dx.plos.org/10.1371/journal.pone.0144296 (visité le 07/03/2016) (cf. p. 47, 69, 80, 144).

[45] Yann LECUN, Yoshua BENGIO et al. « Convolutional networks for images, speech, and time series ». In : *The handbook of brain theory and neural networks* 3361.10 (1995), p. 1995 (cf. p. 46, 111).

[46] Geoffrey N LEECH. *Principles of pragmatics*. Routledge, 2016 (cf. p. 36).

[47] Vladimir I LEVENSHTEIN. « Binary codes capable of correcting deletions, insertions, and reversals ». In : *Soviet physics doklady*. T. 10. 8. 1966, p. 707–710 (cf. p. 62, 63).

[48] Jiwei LI, Minh-Thang LUONG et Dan JURAFSKY. « A hierarchical neural autoencoder for paragraphs and documents ». In : *arXiv preprint arXiv :1506.01057* (2015) (cf. p. 51).

[49] Xiang LI, Rui YAN et Ming ZHANG. « Joint emoji classification and embedding learning ». In : *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data*. Springer. 2017, p. 48–63 (cf. p. 51, 83).

[50] Rensis LIKERT. « A technique for the measurement of attitudes. » In : *Archives of psychology* (1932) (cf. p. 135, 146).

[51] Laurens van der MAATEN et Geoffrey HINTON. « Visualizing data using t-SNE ». In : *Journal of machine learning research* 9.Nov (2008), p. 2579–2605 (cf. p. 88).

[52] James MACQUEEN et al. « Some methods for classification and analysis of multivariate observations ». In : *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. T. 1. 14. Oakland, CA, USA. 1967, p. 281–297 (cf. p. 89, 94).

[53] Paolo MASSA et Paolo AVESANI. « Trust-aware recommender systems ». In : *Proceedings of the 2007 ACM conference on Recommender systems*. ACM. 2007, p. 17–24 (cf. p. 127).

[54] Tomas MIKOLOV, Kai CHEN, Greg CORRADO et Jeffrey DEAN. « Efficient estimation of word representations in vector space ». In : *arXiv preprint arXiv :1301.3781* (2013) (cf. p. 45, 46, 48, 93, 108, 144).

[55] Tomas MIKOLOV, Ilya SUTSKEVER, Kai CHEN, Greg S CORRADO et Jeff DEAN. « Distributed representations of words and phrases and their compositionality ». In : *Advances in neural information processing systems*. 2013, p. 3111–3119 (cf. p. 45, 48, 93, 109, 144).

[56] George A MILLER. « WordNet : a lexical database for English ». In : *Communications of the ACM* 38.11 (1995), p. 39–41 (cf. p. 92).

[57] Hannah MILLER, Jacob THEBAULT-SPIEKER, Shuo CHANG, Isaac JOHNSON, Loren TERVEEN et Brent HECHT. « “Blissfully happy” or “ready to fight” : Varying Interpretations of Emoji ». In : (2016). URL : http://www-users.cs.umn.edu/~bhecht/publications/ICWSM2016_emoji.pdf (visité le 10/04/2016) (cf. p. 36).

[58] Roberto NAVIGLI et Simone Paolo PONZETTO. « BabelNet : Building a very large multilingual semantic network ». In : *Proceedings of the 48th annual meeting of the association for computational linguistics*. Association for Computational Linguistics. 2010, p. 216–225 (cf. p. 48).

[59] Andrew Y NG, Michael I JORDAN et Yair WEISS. « On spectral clustering : Analysis and an algorithm ». In : *Advances in neural information processing systems*. 2002, p. 849–856 (cf. p. 99).

[60] Thien Huu NGUYEN et Ralph GRISHMAN. « Relation extraction : Perspective from convolutional neural networks ». In : *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*. 2015, p. 39–48 (cf. p. 109).

[61] John O'DONOVAN et Barry SMYTH. « Trust in recommender systems ». In : *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM. 2005, p. 167–174 (cf. p. 105).

[62] Umashanthi PAVALANATHAN et Jacob EISENSTEIN. « Emoticons vs. Emojis on Twitter : A Causal Inference Approach ». In : *arXiv preprint arXiv :1510.08480* (2015). URL : http://arxiv.org/abs/1510.08480 (visité le 07/03/2016) (cf. p. 38, 79).

[63] Michael J PAZZANI et Daniel BILLSUS. « Content-based recommendation systems ». In : *The adaptive web*. Springer, 2007, p. 325–341 (cf. p. 105).

[64] Charles Sanders PEIRCE. « Logic as semiotic : The theory of signs ». In : (1902). URL : http://philpapers.org/rec/peilas (visité le 19/04/2016) (cf. p. 31).

[65] Robert PLUTCHIK. « The multifactor-analytic theory of emotion ». In : *the Journal of Psychology* 50.1 (1960), p. 153–171 (cf. p. 47, 147).

[66] Robert PLUTCHIK. « The nature of emotions : Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice ». In : *American scientist* 89.4 (2001), p. 344–350 (cf. p. 98).

[67] Henning POHL, Christian DOMIN et Michael ROHS. « Beyond Just Text : Semantic Emoji Similarity Modeling to Support Expressive Communication ». In : *ACM Transactions on Computer-Human Interaction (TOCHI)* 24.1 (2017), p. 6 (cf. p. 46, 87–89, 93, 122, 144).

[68] Lorien Y PRATT. « Discriminability-based transfer between neural networks ». In : *Advances in neural information processing systems*. 1993, p. 204–211 (cf. p. 147).

[69] Dragomir R RADEV, Hong QI, Harris WU et Weiguo FAN. « Evaluating Web-based Question Answering Systems. » In : *LREC*. 2002 (cf. p. 51, 117, 133).

[70] Yanghui RAO, Qing LI, Xudong MAO et Liu WENYIN. « Sentiment topic models for social emotion mining ». In : *Information Sciences* 266 (mai 2014), p. 90–100. ISSN : 00200255. DOI : 10.1016/j.ins.2013.12.059. URL : http://linkinghub.elsevier.com/retrieve/pii/S002002551400019X (visité le 21/11/2016) (cf. p. 50).

[71] Jonathon READ. « Using emoticons to reduce dependency in machine learning techniques for sentiment classification ». In : *Proceedings of the ACL student research workshop*. Association for Computational Linguistics, 2005, p. 43–48. URL : http://dl.acm.org/citation.cfm?id=1628969 (visité le 07/03/2016) (cf. p. 50).

[72] Radim REHUREK et Petr SOJKA. « Software framework for topic modelling with large corpora ». In : *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Citeseer. 2010 (cf. p. 93, 109).

[73] Francesco RICCI, Lior ROKACH et Bracha SHAPIRA. « Introduction to recommender systems handbook ». In : *Recommender systems handbook*. Springer, 2011, p. 1–35 (cf. p. 128, 146).

[74] Andrew ROSENBERG et Julia HIRSCHBERG. « V-measure : A conditional entropy-based external cluster evaluation measure ». In : *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. 2007 (cf. p. 99).

[75] John Rogers SEARLE. *Speech acts : An essay in the philosophy of language*. T. 626. Cambridge university press, 1969 (cf. p. 36).

[76] Guy SHANI et Asela GUNAWARDANA. « Evaluating recommendation systems ». In : *Recommender systems handbook*. Springer, 2011, p. 257–297 (cf. p. 128, 139).

[77] Karianne SKOVHOLT, Anette GRØNNING et Anne KANKAANRANTA. « The Communicative Functions of Emoticons in Workplace E-Mails : :-) ». In : *Journal of Computer-Mediated Communication* 19.4 (juil. 2014), p. 780–797. ISSN : 10836101. DOI : 10.1111/jcc4.12063. URL : http://doi.wiley.com/10.1111/jcc4.12063 (visité le 23/03/2016) (cf. p. 35).

[78] Richard SOCHER, Alex PERELYGIN, Jean Y. WU, Jason CHUANG, Christopher D. MANNING, Andrew Y. NG et Christopher POTTS. « Recursive deep models for semantic compositionality over a sentiment treebank ». In : *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*. T. 1631. Citeseer, 2013, p. 1642. URL : http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.383.1327&rep=rep1&type=pdf (visité le 01/10/2016) (cf. p. 71).

[79] Richard S SUTTON. « Learning to predict by the methods of temporal differences ». In : *Machine learning* 3.1 (1988), p. 9–44 (cf. p. 147).

[80] *SwiftKey Emoji Report*. Avr. 2015, p. 18. URL : http://www.aargauerzeitung.ch/asset_document/i/129067827/download (cf. p. 38, 39).

[81] Duyu TANG, Furu WEI, Bing QIN, Ming ZHOU et Ting LIU. « Building large-scale twitter-specific sentiment lexicon : A representation learning approach ». In : *Proceedings of coling 2014, the 25th international conference on computational linguistics : Technical papers*. 2014, p. 172–182 (cf. p. 47).

[82] Emoji Research TEAM. *Emoji_Report*. Sept. 2015, p. 40. URL : http://emogi.com/documents/Emoji_Report_2015.pdf (cf. p. 35, 38).

[83] Mike THELWALL, Kevan BUCKLEY, Georgios PALTOGLOU, Di CAI et Arvid KAPPAS. « Sentiment strength detection in short informal text ». In : *Journal of the American Society for Information Science and Technology* 61.12 (2010), p. 2544–2558. URL : http://onlinelibrary.wiley.com/doi/10.1002/asi.21416/full (visité le 18/05/2016) (cf. p. 69–71, 75, 78).

[84] Tian TIAN, Marco DINARELLI, Isabelle TELLIER et Pedro CARDOSO. « Etiquetage morpho-syntaxique de tweets avec des CRF ». In : *TALN 2015*. 2015 (cf. p. 91).

[85] Tijmen TIELEMAN et Geoffrey HINTON. « Lecture 6.5-rmsprop : Divide the gradient by a running average of its recent magnitude ». In : *COURSERA : Neural networks for machine learning* 4.2 (2012), p. 26–31 (cf. p. 113, 114).

[86] Chad C. TOSSELL, Philip KORTUM, Clayton SHEPARD, Laura H. BARG-WALKOW, Ahmad RAHMATI et Lin ZHONG. « A longitudinal study of emoticon use in text messaging from smartphones ». In : *Computers in Human Behavior* 28.2 (mar. 2012), p. 659–663. ISSN : 07475632. DOI : 10.1016/j.chb.2011.11.012. URL : http://linkinghub.elsevier.com/retrieve/pii/S0747563211002561 (visité le 01/03/2016) (cf. p. 37).

[87] Nguyen Xuan VINH, Julien EPPS et James BAILEY. « Information theoretic measures for clusterings comparison : Variants, properties, normalization and correction for chance ». In : *Journal of Machine Learning Research* 11.Oct (2010), p. 2837–2854 (cf. p. 99).

[88] Ulrike VON LUXBURG. « A tutorial on spectral clustering ». In : *Statistics and computing* 17.4 (2007), p. 395–416 (cf. p. 100).

[89] Soroush VOSOUGHI, Prashanth VIJAYARAGHAVAN et Deb ROY. « Tweet2vec : Learning tweet embeddings using character-level cnn-lstm encoder-decoder ». In : *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM. 2016, p. 1041–1044 (cf. p. 46, 50).

[90] Christopher JCH WATKINS et Peter DAYAN. « Q-learning ». In : *Machine learning* 8.3-4 (1992), p. 279–292 (cf. p. 147).

[91] Sanjaya WIJERATNE, Lakshika BALASURIYA, Amit SHETH et Derek DORAN. « A semantics-based measure of emoji similarity ». In : *Proceedings of the International Conference on Web Intelligence*. ACM. 2017, p. 646–653 (cf. p. 48).

[92] Ruobing XIE, Zhiyuan LIU, Rui YAN et Maosong SUN. « Neural emoji recommendation in dialogue systems ». In : *arXiv preprint arXiv :1612.04609* (2016) (cf. p. 51, 53, 83, 106).

[93] Ruobing XIE, Zhiyuan LIU, Rui YAN et Maosong SUN. « Neural emoji recommendation in dialogue systems ». In : *arXiv preprint arXiv :1612.04609* (2016) (cf. p. 73).

[94] Kelvin XU, Jimmy BA, Ryan KIROS, Kyunghyun CHO, Aaron COURVILLE, Ruslan SALAKHUDINOV, Rich ZEMEL et Yoshua BENGIO. « Show, attend and tell : Neural image caption generation with visual attention ». In : *International conference on machine learning*. 2015, p. 2048–2057 (cf. p. 124).

[95] Mao YE, Xingjie LIU et Wang-Chien LEE. « Exploring social influence for recommendation : a generative model approach ». In : *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM. 2012, p. 671–680 (cf. p. 57).

[96] Luda ZHAO et Connie ZENG. *Using neural networks to predict emoji usage from twitter data*. 2017 (cf. p. 51, 106).

# ANNEXES

# A. EmoReco : plateforme d'évaluation

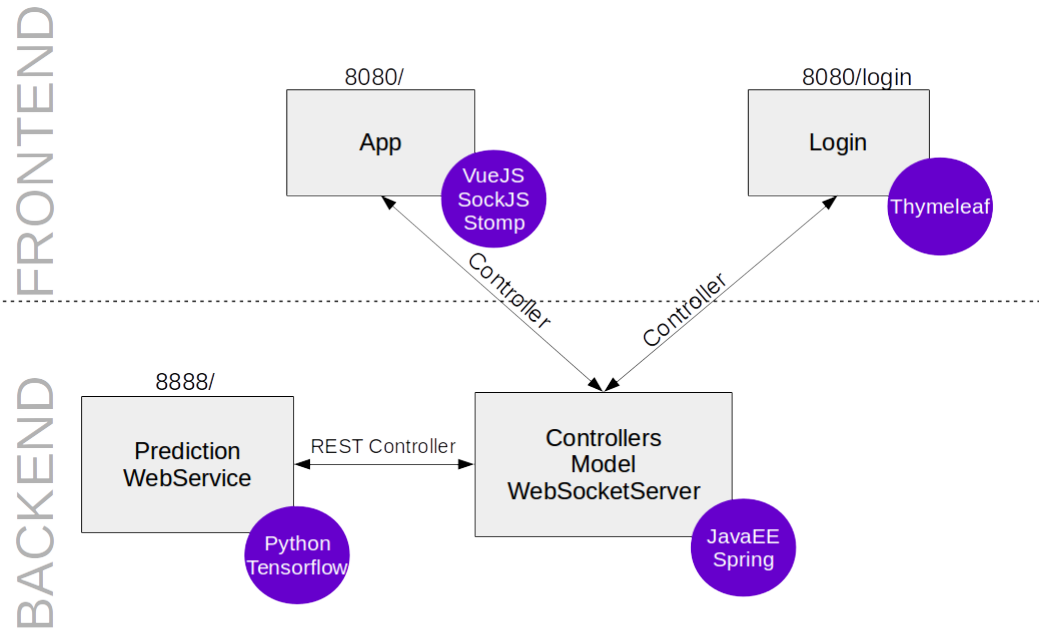


Figure A.1. – Organisation principale de l'application

La plateforme d'évaluation du système de recommandation a été détaillée au travers des choix effectués et de son orientation lors du chapitre 6. Dans ce chapitre d'annexe, l'accent est mis sur l'aspect technique de sa conception.

L'architecture globale de l'application est visible en figure A.1 et résulte de plusieurs choix techniques de conception. Ainsi, il s'agit avant tout d'une application en page unique (*Single Page Application* - SPA) à laquelle nous avons greffé un système de connexion par une page d'accès externe. En arrière-plan, l'application se divise en deux composantes principales : les modèles, contrôleurs et services d'un côté, et le service web de prédiction d'emojis de l'autre. Nous détaillons ces composantes dans les sections suivantes.

## A.1. Vue et contrôleurs

La vue de l'application est réactive et adapte l'affichage pour un accès mobile ou de bureau. Elle a pour particularité d'être divisée en deux composantes dis-

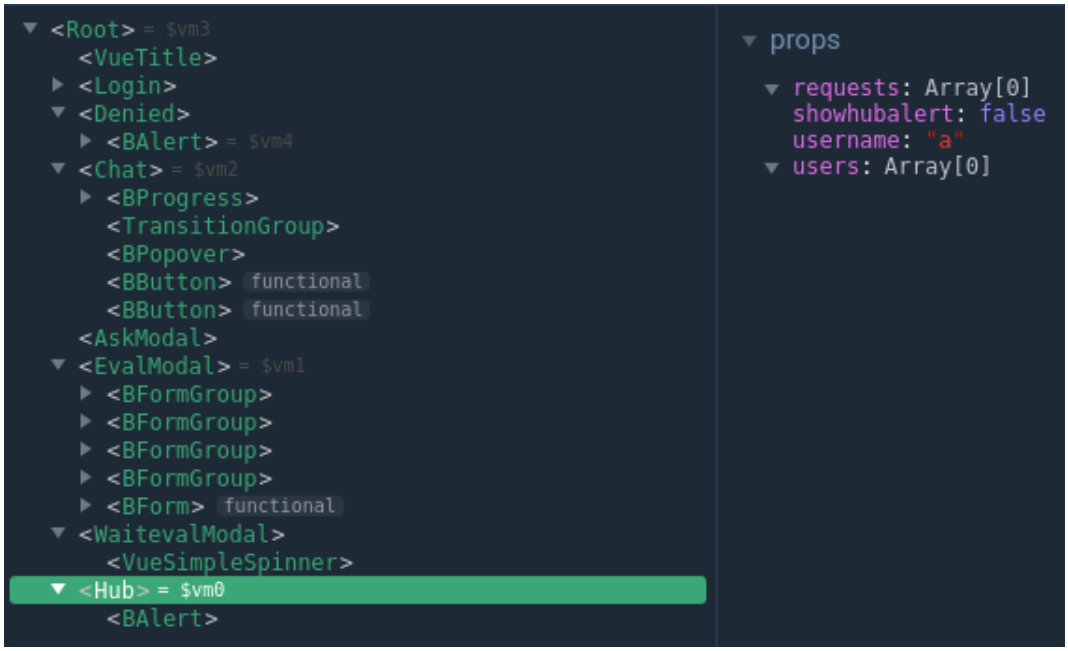


Figure A.2. – Composants VueJS et modèle de vue.

tinctes : la page de connexion et l'application.

L'interaction entre ces deux composantes est effectuée par le biais de contrôleurs dédiés. Ainsi, lorsque l'utilisateur accède à l'application, son état de connexion le redirigera vers l'application ou vers la page de connexion gérée par le contrôleur dédié à l'URL 8080/login. Cette page de connexion a été mise en place en combinant le gestionnaire de *templates* Thymeleaf avec le module Spring Security permettant de sécuriser l'accès et la gestion des inscriptions et connexions.

L'application en elle-même est mise en place en Javascript avec VueJS permettant d'en faire une SPA une fois la connexion effectuée. De cette manière, la réactivité de l'application en est améliorée en passant du modèle MVC au modèle MVVM permettant ainsi à la vue d'avoir sa propre réplique de données (visible en figure A.2), utile pour la gestion d'utilisateurs connectés par exemple.

Parce qu'il s'agit d'une application de messagerie instantanée, l'utilisation de l'AJAX pour gérer les événements s'avère trop lourde : il est nécessaire de réagir instantanément à plusieurs changements tels que l'arrivée d'un nouvel utilisateur, l'envoi et la réception de messages et autres. Pour répondre à ce besoin, un serveur de websockets est mis en place et communique avec la vue par le biais de contrôleurs et de gestionnaires de sockets (SockJS) sur le courtier (*broker*) en attente sur l'URL /ws.

## A.2. Modèles

En arrière-plan, l'application gère les informations à deux niveaux selon le type d'information :

— Information mineure et temporaire : gérée au niveau du serveur de websocket Spring-websocket
— Information majeure à conserver : gérée au niveau de la modélisation des relations objets (ORM)

Chaque type d'information possède ainsi son modèle mais seules les informations à conserver ont leur classe de modèle Java annotée pour l'ORM d'Hibernate. Cet ORM est utilisé pour faire le lien avec une base de données Java (HSQLDB) permettant de sauvegarder les messages, les conversations, les utilisateurs, les scénarios et les évaluations. Ce sont ces informations qui sont étudiées au chapitre 6 et gérées par un répertoire CRUD.

## A.3. Services et prédiction

Les services permettent de fournir les différentes données avec une logique ajoutée telle que l'autorisation ou non de connexion. Toutefois, un de ses services a pour particularité de retourner un scénario aléatoire à l'aide d'un contrôleur REST retournant le résultat sous forme d'objet JSON.

C'est également le cas du service web de prédiction qui retourne la prédiction d'emojis sous format JSON selon le texte fourni. Ce service est externe à l'application web puisqu'il s'agit directement de l'utilisation du modèle Keras/Tensorflow initialisé et en attente sur le port 8888 qu'il est possible d'interroger en requête GET. Contrairement à la première application web de la section 6.2, l'accès à cette URL est restreint, c'est pourquoi il est nécessaire de passer par l'application web JavaEE et son contrôleur REST adéquat pour obtenir une prédiction.

Il convient de noter que le choix de l'utilisation d'un service webs python dépend de la taille de l'application. En effet, sous un nombre élevé de requêtes, les ressources utilisées pour prédire ne peuvent suffire et un système de file d'attente devrait alors être mis en place, rendant l'application peu réactive pour une utilisation à grande échelle. Pour pallier à cela il conviendrait d'utiliser TensorflowJS pour utiliser les ressources du navigateur de l'utilisateur, avec l'inconvénient pour l'accès mobile.

# B. Déploiement Android

La présente thèse s'intégrant dans un contexte industriel, plusieurs éléments d'ingénierie ont été effectués. Bien qu'ils n'aient pas leur place dans le corps principal de cette thèse, il convient d'en présenter les principaux éléments.

## B.1. Motivation

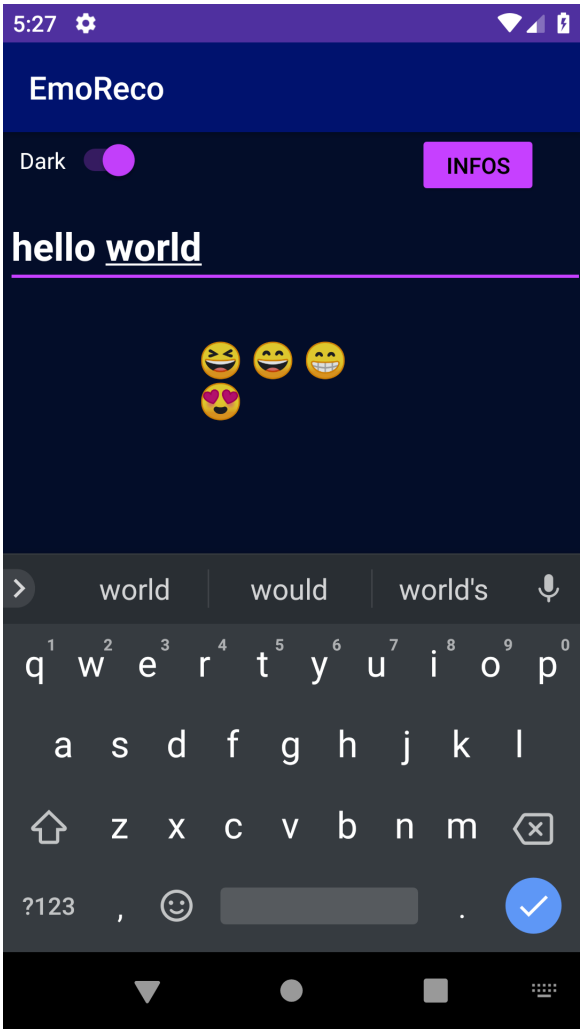


Figure B.1. – Système déployé sous Android

L'objectif industriel est de mettre en place un système de recommandation d'emojis performant qui puisse s'intégrer dans l'application de messagerie SMS de l'entreprise, à savoir Mood Messenger[5]. Cette application étant disponible

5. https://play.google.com/store/apps/details?id=com.calea.echo&hl=fr

uniquement sous android, plusieurs choix de départ ont été faits en gardant en tête le déploiement final visible en figure B.1, tout particulièrement pour la recommandation par catégories d'emojis (voir Chapitre 5) et les trois points suivants :

1. Système en Java : l'application Mood Messenger étant en Java, l'intégration du système de recommandation doit se faire par ce langage de programmation ;
2. Ré-utilisabilité du modèle : le modèle de prédiction appris doit être utilisable directement dans l'application finale ;
3. Optimisations d'espace disque et de ressources de calcul : l'utilisation de la batterie doit être limitée et l'augmentation de la taille de l'application finale doit être moindre.

Ces trois points résultent d'obstacles réels provenant de la nature même de l'application finale. En effet, une messagerie SMS n'est pas attendue comme étant une application lourde, contrairement à une application de streaming (Netflix) ou de partage d'images (Snapchat).

## B.2. Déploiement

Le déploiement de l'application est pensé en amont, avant les expérimentations, afin de répondre aux besoins du langage et de la ré-utilisation du modèle. En effet, le modèle de prédiction appris en Python doit être réutilisable en Java, c'est pourquoi l'interface de programmation (API) Keras [6] a été choisie : les modèles appris sous Keras fonctionnent directement sous DeepLearning4J [7], l'interface de programmation *open source* dédiées à l'apprentissage profond en Java. Un modèle appris en python sera alors immédiatement utilisable sous Java, alliant la rapidité de prototypage du Python avec la nécessité du Java pour Android.

La choix de DeepLearning4J n'est toutefois pas final et a été remplacé par l'utilisation directe de TensorFlow [8]. L'utilisation de TensorFlow au lieu de DeepLearning4J permet une réduction des dépendances en passant du nombre élevé de dépendances DL4J au seul import de la classe d'interface Java pour les inférences de TensorFlow. Cette dernière nécessite l'utilisation du Native Development Kit (NDK) pour lire la version native en C++ de Tensorflow.

Lors de chaque requête entrante, le pré-traitement effectué précédemment à l'aide de librairies Python comme NLTK ou Scikit-Learn doit être reproduit en

6. https://keras.io/
7. https://deeplearning4j.org/
8. https://www.tensorflow.org/

Java. Pour ce faire, nous créons une classe *InputUtils* contenant toutes les fonctions nécessaires au pré-traitement de chaque entrée textuelle.

1. Nettoyage des données : suppression des liens hypertextes, limitation de la répétition typographique, tokenisation du texte et prise en compte optionnelle d'un lexique de contractions dépendant de la langue.

```
public static String[] cureData( String inputText, JSONObject
    enContractions ) {
    inputText = inputText.toLowerCase();
    inputText = inputText.replaceAll(URLPATTERN, "") ;
    inputText = inputText.replaceAll("(.)\\1+", "$1$1");
    List<String> simpletokens = tokenize(inputText);
    String tokensStr = TextUtils.join(" ", simpletokens);
    simpletokens = new ArrayList<String>();
    for (String token : tokensStr.split(" ")) simpletokens.add(
       fixedTokenizeEN(token, enContractions));
    tokensStr = TextUtils.join(" ", simpletokens);
    Log.i(TAG, tokensStr);
    return tokensStr.split(" ");
}
```

2. Vectorisation à l'aide du vocabulaire issu du CountVectorizer de Scikit-Learn. Ce vocabulaire est extrait sous forme de liste et fourni en tant que ressource interne à l'application.

```
public static List<Integer> vectorize(HashMap<String, Integer>
     vectorizer, String[] tokens) {
    List<Integer> vectorizedTxt = new ArrayList<Integer>();
    for (int i = 0; i<tokens.length;i++ ){
        if(vectorizer.containsKey(tokens[i]) ){
            vectorizedTxt.add( vectorizer.get(tokens[i]) );
        }else{
            vectorizedTxt.add( UNK_CHAR );
        }
    }
    return vectorizedTxt;
}
```

3. Élagage et remplissage : cette étape est effectuée directement sur les données vectorisées. La valeur de `SEQ_SIZE` dépend du modèle appris et est ici de 15.

```
public static List<Integer> padTrim(List<Integer> indices) {
    if(indices.size() > SEQ_SIZE){ return leftTrim(indices); }
```

```
    else if(indices.size() < SEQ_SIZE){ return leftPad(indices)
        ; }
    else { return indices; }
}
```

4. Envoi des données à Tensorflow : les valeurs numérisées sont ensuite transformées en FLOAT avant d'être envoyées à Tensorflow à l'aide de la méthode feed() de la classe TensorflowInferenceInterface, initialisée en amont avec les atouts (*assets*) et le modèle.

```
FLOAT_VALUES = InputUtils.convertToFloatArray(vectorizedTxt);
tf.feed(INPUT_NAME,FLOAT_VALUES,1,15);
```

5. Prédiction et obtention des scores les plus élevés avec une fonction argmax $\{x|f(x) \geq f(y), \forall(y) \in D\}$ modifiée pour prendre en compte les deux éléments les plus élevés de l'ensemble $D$.

```
public static Object[] argmax2(float[] array){
    int best = -1;
    int second_best = -2;
    float best_confidence = 0.0f;
    float second_best_confidence = 0.0f;
    for(int i = 0;i < array.length;i++){
        float value = array[i];
        if (value > best_confidence){
            second_best_confidence = best_confidence;
            second_best = best;
            best_confidence = value;
            best = i;
        }
    }
    return new Object[]{best,best_confidence, second_best,
        second_best_confidence};
}
```

Ces 5 étapes sont effectuées à chaque nouvelle entrée, bénéficiant de l'utilisation du NDK pour la rapidité de la prédiction ainsi que de l'allègement de la tokenisation. Toutefois, le modèle importé a une taille élevée pour une application mobile. Il convient donc de faire un point sur les performances et l'optimisation de l'espace disque nécessaire.

## B.3. Optimisation de taille

### B.3.1. Optimisation de la taille du modèle

La taille du modèle Keras est de 35,8 Mo, ce qui est trop élevé pour s'ajouter à une application de messagerie. Il convient alors d'optimiser le modèle non plus uniquement vis-à-vis des performances comme ce fut le cas dans le chapitre 5, mais également vis-à-vis de sa taille finale, de son efficience.

L'optimisation de la taille du modèle peut se faire de plusieurs manières. Elle peut se faire en amont tout d'abord, lors de la limitation du vocabulaire pris en compte avec, par exemple, un nombre minimum d'occurrences par mot plus élevé ou une limite de taille de vocabulaire définie arbitrairement en prenant uniquement en compte les éléments les plus fréquents du corpus d'entraînement. Cette approche est risquée puisqu'elle modifie directement la capacité du modèle à inférer. Pour cette raison et parce que la taille du corpus d'apprentissage ne nous permet pas de le limiter, nous n'avons pas appliqué cette méthode.

La seconde méthode pour optimiser la taille du modèle est celle recommandée par Google qui consiste à quantifier le modèle (*quantize*) précédemment figé. Il s'agit dans un premier temps de figer le modèle Keras en un graphe Tensorflow uniquement, permettant de passer de 35,8 Mo à 11,7 Mo. Ensuite, il convient de quantifier ce modèle, c'est-à-dire limiter le nombre de décimales prises en compte après la virgule pour toutes les valeurs des vecteurs du modèle, permettant de réduire la taille de plus de deux tiers. La quantification a été utilisée sur les matrices de poids du modèle permettant de passer de 11,7 Mo à 3,0 Mo.

### B.3.2. Optimisation de la taille de la librairie

Le modèle faisant désormais 3,0 Mo, il convient d'optimiser la librairie Tensorflow afin de limiter la taille embarquée dans l'application android. Pour cela, il convient de compiler notre propre version C++ de Tensorflow en y sélectionnant uniquement les opérations utiles pour notre modèle. Chaque opération est un fichier à importer et bon nombre de ces opérations ne sont pas nécessaires lors de la prédiction mais uniquement lors de l'entraînement. Nous avons donc sélectionné les opérations nécessaires afin de recompiler la librairie d'inférence pour les architectures ARM. C'est ensuite cette librairie native d'inférence qui sera utilisée par le NDK d'android et manipulée à l'aide de la classe Java *org.tensorflow.contrib.android.TensorFlowInferenceInterface*. Au final, la librairie native recompilée a une taille de 14,6 Mo. Cette taille pourrait encore être optimisée en filtrant davantage les opérations inutilisées même pendant la prédiction, et ainsi obtenir une version de Tensorflow totalement dépendante du modèle déployé. La figure B.2 permet de voir les détails d'utilisation disque de

l'application finale compressée au format APK.

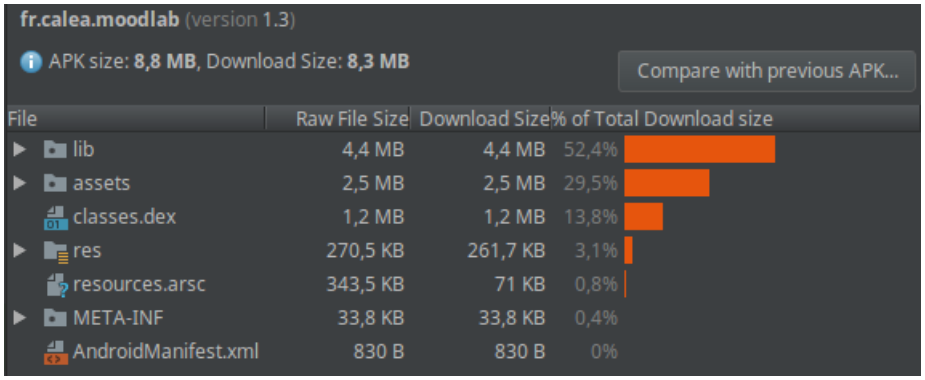


Figure B.2. – Utilisation de l'espace disque dans l'application finale

La taille de la librairie d'inférence est un sujet actuel ayant motivé une version de Tensorflow dédiée aux systèmes embarqués, nommée Tensorflow-lite. Cette version n'était pas stable lors de la mise en place du déploiement android de notre modèle, mais pourrait désormais remplacer l'utilisation directe de Tensorflow.

## B.4. Performances

### B.4.1. Performances du modèle

L'objectif lors du déploiement n'est pas d'optimiser les performances du modèles comme ce fut le cas au chapitre 5. Toutefois, la réduction de la taille du modèle via sa quantification altère légèrement les performances finales puisqu'elle supprime des informations mineures dans les matrices de poids du modèle. De plus, la performance finale est possiblement altérée par la différence de pré-traitement. En effet, même si le pré-traitement appliqué est le même, des différences mineures peuvent venir du fait qu'il s'agissait d'utilisation de librairies Python, ré-implémentées en Java afin de se passer de toute dépendance supplémentaire. La tokenisation est actuellement moins approfondie lors du passage au Java.

Les différences de performance ne sont pas chiffrées puisque les modèles finaux sont appris sur l'ensemble du corpus disponible. En revanche, leurs différences sont observables en insérant le même texte dans l'application EmoReco Android, et dans l'application EmoReco Web [9].

9. http://lsis-mood-emoji.lsis.org:8080/

### B.4.2. Utilisation des ressources

L'utilisation des ressources par le système prédictif est un point sensible pour l'inclusion du système en production, tout particulièrement dans le cas d'une application de messagerie. Deux points sont à vérifier : l'utilisation de la mémoire et la consommation énergétique.

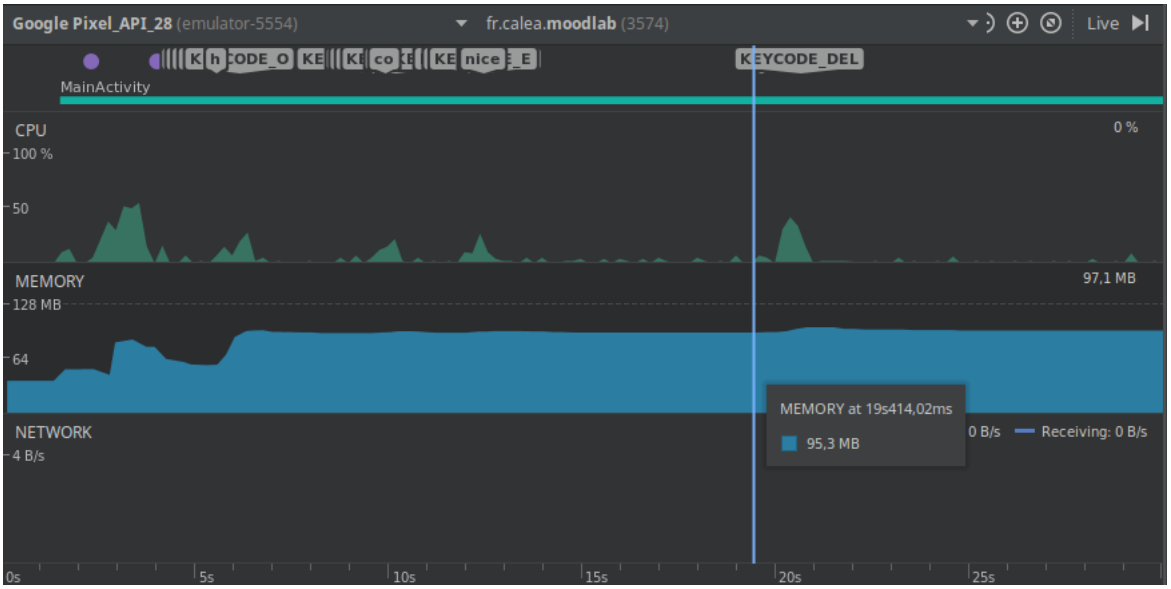


Figure B.3. – Analyse des consommations globales

Nous partons du principe que la consommation énergétique est directement liée à l'utilisation du ou des processeurs. Cette utilisation apparaît dans la figure B.3 sous forme de pics localisés principalement lors du chargement de l'*activity* principale et varie lors de la frappe. La consommation excessive lors du lancement de l'*activity* principale est due à l'initialisation de la classe d'inférence avec l'intégration du modèle. Initialiser cette classe lors de l'accès à l'application, quelle qu'en soit l'*activity* déclenchée peut être une bonne idée afin de réduire cette consommation au lancement initial de l'application, et non à chaque accès comme c'est actuellement le cas. Pour une application de messagerie à laquelle on accède plusieurs fois par jour, ce détail est important.

En analysant plus en détails l'utilisation des processeurs, comme visible en figure B.4, nous constatons un pic d'usage de 60 % maximum au chargement de la classe d'inférence, puis d'un usage régulier oscillant entre 17 et 20 % maximum lors de la prédiction. L'usage de la batterie reste ainsi modeste si l'application n'est pas sans cesse relancée et peut être davantage optimisé en ajoutant un délai d'attente de fin de frappe de 0.4 seconde entre chaque insertion de lettre, comme c'est le cas dans l'application web.

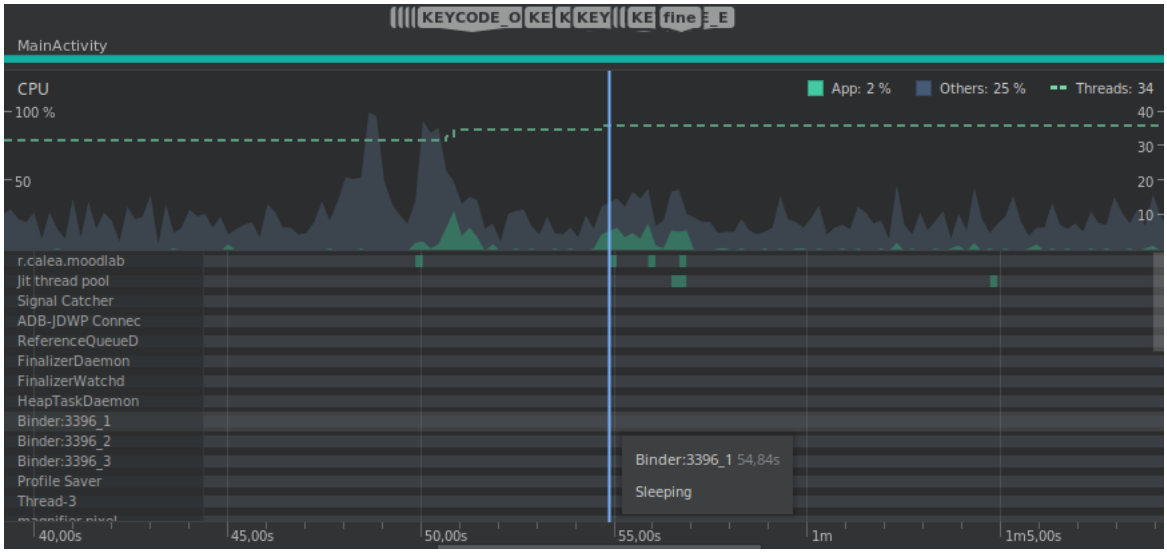


Figure B.4. – Analyse de la consommation du processeur

Concernant la mémoire utilisée, le constat est plus mitigé. La figure B.5 permet d'observer un usage correct de la mémoire issue du code de l'application avec seulement 19 Mo de mémoire constante utilisée, que ce soit pendant ou en dehors de la prédiction. Cependant, la librairie native consomme initialement un peu plus de 32 Mo de mémoire avant de doubler à 63,7 Mo suite aux prédictions. dans le code actuel, la représentation vectorielle du texte en flottants n'est pas vidée à la fin de chaque prédiction. Le minimum incompressible est donc supérieur à 32 Mo de mémoire utilisée avec des pics à 63,7 Mo. Cela peut paraître peu en soi, mais une fois le système intégré dans une application finale possédant d'autres fonctionnalités et images générées, cet ajout d'utilisation de mémoire supplémentaire n'est pas anodin.

## B.5. Conclusion

Le déploiement du système de recommandation d'emojis dans une application android est l'objectif industriel final vis-à-vis de l'entreprise. Ce processus doit être pris en compte dès la constitution du modèle afin d'optimiser l'utilisation de toutes les ressources de l'environnement mobile, qu'il s'agisse de l'espace disque, de la mémoire ou du processeur et de la consommation de la batterie.

Notre système déployé en une application android est le résultat d'un léger compromis entre intégration logicielle et performance de prédiction pure.

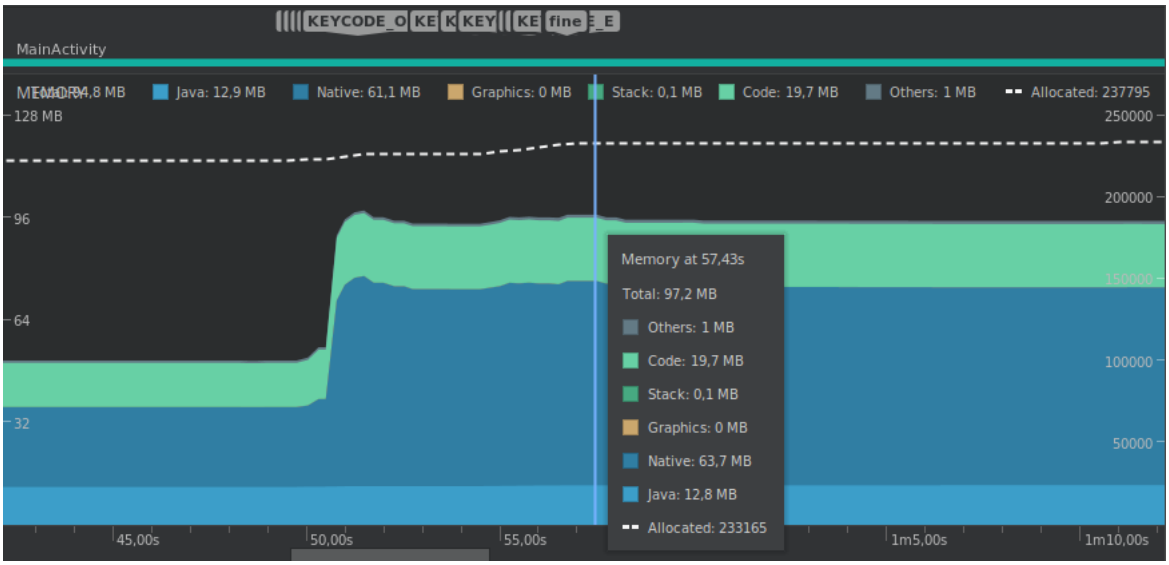


Figure B.5. – Analyse de la consommation en mémoire